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

SUBCENTRALIDADES E PRÊMIO SALARIAL INTRA-URBANO NA REGIÃO
METROPOLITANA DE SÃO PAULO

Rodger Barros Antunes Campos

Orientador: Prof. Dr. Carlos Roberto Azzoni

SÃO PAULO

2018

Prof. Dr. Vahan Agopyan
Reitor da Universidade de São Paulo

Prof. Dr. Fabio Frezatti
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. José Carlos de Souza Santos
Chefe do Departamento de Economia

Prof. Dr. Ariaster Baumgratz Chimeli
Coordenador do Programa de Pós-Graduação em Economia

RODGER BARROS ANTUNES CAMPOS

**SUBCENTRALIDADES E PRÊMIO SALARIAL INTRA-URBANO NA REGIÃO
METROPOLITANA DE SÃO PAULO**

Tese apresentada ao Programa de Pós-Graduação
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

2018

FICHA CATALOGRÁFICA

Elaborada por Rafael Mielli Rodrigues – CRB-8/7286
Seção de Processamento Técnico do SBD/FEA/USP

Campos, Rodger Barros Antunes

Subcentralidades e prêmio salarial intra-urbano na região metropo-
litana de São Paulo / Rodger Barros Antunes Campos. – São Paulo,
2018.

230 p.

Tese (Doutorado) – Universidade de São Paulo, 2018.

Orientador: Carlos Roberto Azzoni.

1. Economia 2. Economia do trabalho 3. Aglomeração urbana
4. Econometria I. Universidade de São Paulo. Faculdade de Economia,
Administração e Contabilidade. II. Título.

CDD – 330

À minha mãe, Ivone Barros

AGRADECIMENTOS

Primeiramente, agradeço a Deus pela saúde e graça dispensadas para a conclusão dessa longa etapa.

Não raro, os trabalhos que se desenvolvem não demandam somente ajudas técnicas, mas suportes que orbitam outras esferas da nossa vida. Por esse motivo, agradeço a minha mãe, Ivone Barros, por todo suporte psicológico e financeiro ao longo desses anos. Agradeço também ao meu padrasto, Paulo César, pela confiança, cooperação e suporte financeiro. Agradeço a Camila Assis pelo suporte, cuidado e por fazer com que os dias complicados se tornassem mais leves. Aos amigos Diego Morales, Luiz Guilherme, Hildebrando Santana, Ravi Lima e Stephanie Gouvêa, meu mais sincero agradecimento, pois fizeram a estrada menos íngreme.

Pelas ajudas e horas despendidas às discussões técnicas e estímulos a prosseguir em frente, agradeço ao professor Carlos R. Azzoni. Além de um excelente profissional, foi capaz de tornar o ambiente agradável tanto ao aprendizado como à produção acadêmica. Os agradecimentos também se estendem aos professores do NEREUS, por contribuições e ideias relevantes, não apenas nessa tese, mas em outros artigos, seja durante os Encontros Semanais no NEREUS seja em seus escritórios. Especialmente, agradeço aos professores André S. Chagas e Danilo C. Iglori por terem composto todas as comissões de avaliação de progresso, contribuindo positivamente com ideias e direcionamento dessa pesquisa.

Aos colegas das turmas de doutoramento do segundo semestre de 2014 e de mestrado, do primeiro semestre de 2014, minha gratidão por terem sido cooperativos. Dentre esses, meu sincero agradecimento a Edivaldo Constantino Jr pela recorrente disponibilidade em ajudar e discutir questões inerentes a esse trabalho e a Diana Silva, por sua capacidade cooperativa e por compartilhar seu conhecimento, dispondo sempre do seu tempo. Da mesma forma, agradeço a disposição de Keyi Ussami em sempre cooperar com seu conhecimento. Ainda, aos amigos Ana Maria Barufi, Deborah Seabra, Jaqueline Visentin, Juliana Scipitore e Gustavo P. Serra, muito obrigado pelo apoio materializado em sugestões, incentivos e apontamentos técnicos.

Não posso me furtar de agradecer a Fundação Instituto de Pesquisa Econômicas (Fipe) pelo suporte financeiro ao georreferenciamento da base de dados da RAIS. Agradeço em especial ao amigo Diego Bogado Tomasiello pelo suporte na programação do georreferenciamento no ArcGIS. Agradeço também pelo suporte financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Não posso deixar de agradecer ao pessoal da secretária da pós-graduação por toda atenção prestada nesses últimos quatro anos.

“[...] os pensamentos, como todas as coisas que devem sua existência à memória, podem ser transformados em objetos tangíveis que, como a página escrita ou o livro impresso, se tornam parte do artifício humano. ”

Hannah Arendt
A Condição Humana

RESUMO

As economias de aglomeração apresentam externalidades que impactam sobre os vetores de preços da economia. As interações que ocorrem nas aglomerações são relevantes para a compreensão dos benefícios gerados pela proximidade. Tais benefícios impactam diretamente o salário do trabalho e os preços dos imóveis, entre outros. Muitos trabalhos focaram em estudar economias de aglomeração no contexto espacial agregado, mas poucos estiveram focados no espaço intra-urbano. Em face dessa lacuna na literatura de Economia de Aglomeração, a análise nesta tese debruça-se sobre a Região Metropolitana de São Paulo (RMSP), que foi dividida em áreas homogêneas de 1 km². Dois são os objetivos centrais desse trabalho. Primeiro, buscou-se identificar as áreas mais relevantes quanto à aglomeração de trabalhadores na RMSP, ou seja, os subcentros de emprego (SBD). Para tanto, utilizaram-se os microdados geocodificados do Ministério do Trabalho (RAIS-MTE) para os anos de 2002, 2008 e 2014. Uma nova abordagem empírica foi desenvolvida, utilizando Regressões Ponderadas Geograficamente e uma regra objetiva de valor de corte. Os resultados apontam para três SBD nos anos de 2002 e 2008. No ano de 2014, a abordagem metodológica desenvolvida nessa tese com valor de corte mais restrito identifica apenas dois SBD. Nos dois períodos iniciais, os SBD se localizam em Barueri, São Paulo e São Caetano do Sul. No último ano avaliado, localizam-se em Barueri e São Paulo. O padrão observado é que o emprego localizado nas áreas centrais de São Paulo apresenta nível de crescimento relativamente maior do que nas demais áreas. Os demais SBD perdem não apenas em termos de emprego, mas também em termos de área, isto é, as áreas dos SBD nesses municípios se reduzem. Em 2002, os SBD de Barueri e São Caetano do Sul ocupavam áreas de 5 e 7 km², respectivamente. Em 2014, o primeiro ocupava 1 km² e o segundo deixou de existir. Em São Paulo, a área se elevou de 79 km² para 90 km². Os resultados sugerem elevada concentração espacial do emprego na RMSP. No segundo ensaio, objetiva-se identificar o impacto da aglomeração sobre os salários dos trabalhadores e, adicionalmente, testa-se a hipótese de atenuação espacial da aglomeração. Para tanto, construiu-se uma base de dados longitudinal e considerou-se uma especificação com múltiplos efeitos fixos e defasagem espacial da aglomeração de emprego. Mesmo em face de uma especificação mais restritiva, os resultados sugerem efeito positivo da aglomeração, o qual é atenuado com o distanciamento espacial. Em outras palavras, a aglomeração impacta positivamente o salário. Em estimações que desconsideram a questão da endogeneidade, estima-se um efeito direto (na própria área) de 0,039%, efeito indireto de primeira ordem (em áreas contíguas) de -0,11% e efeito indireto de segunda ordem (anel externo às áreas contíguas) de -0,23%, normalizado por 100,000. Ao tratar o problema da endogeneidade, através do uso de variáveis instrumentais, obtém-se efeito direto de 1,78%, indireto de primeira ordem de -2,12% e o efeito de segunda ordem não é estatisticamente significativo.

Palavras-chave: Centros de Negócios, Subcentros de Negócio, Economias de aglomeração, Mercado de Trabalho, Prêmio Salarial

ABSTRACT

Agglomeration economies are externalities that impact on prices in the economy. The interactions that occur in agglomerations are relevant for the understanding of the benefits generated by proximity. These benefits directly affect workers' wage, real estate prices, etc. Many papers have focused on non-market interaction in aggregated labor markets, but intra-urban labor markets have received less attention. Seeking to fulfill such lack, 1 km² areas of the Metropolitan Area of São Paulo (MASP) are taken as the scope of analysis of this study. The central objectives of this dissertation are twofold. First, in chapter 3, I identified and characterized the most relevant areas in terms of job agglomeration in the MASP, named subcentral business districts (SBD). For this purpose, the geocoded matched employer-employee database of the Ministry of Labor (RAIS-TEM) was used. I developed a new empirical approach to identify the SBD, using Geographically Weight Regression and cut-off rules of identification. The results identified three SBD in the years of 2002 and 2008, and only two in 2014. Considering the two initial periods, the SBD are located in the municipalities of Barueri (SBD-BAR), São Paulo (SBD-SAO) and São Caetano do Sul (SBD-SCS). In the last year, the SBD are located in Barueri and São Paulo municipality only. The employment located in the central areas of São Paulo shows a relatively higher amount of employee than in the other SBD areas. SBD-BAR and SBD-SCS lost not only in terms of employment, but also in terms of area. In 2002, these last two SBD occupied areas of 5 and 7 km² respectively. In 2014, the SBD-BAR occupied 1 km², while the SBD-SCS is not ranked at all. In São Paulo, the area ranges from 79 km² to 90 km². The results stemmed from the first paper suggest a high spatial concentration of employment in the MASP. In the second paper (chapter 4), the objectives are to identify the impact of agglomeration on workers' wages and test the agglomeration spatial attenuation hypothesis (SAH). For that, I use employer-employee RAIS database and consider a specification with multiple fixed effects and spatial lags of the employment agglomeration as a strategy to SAH identification. Even in the face of a more restrictive specification, the results suggest a positive effect of agglomeration, which is attenuated as the spatial distance increases. In other words, agglomeration affects positively workers' wage in the workplace area. Estimates without solving for endogeneity between wage and agglomeration indicate a direct effect of 0.039% (in the area itself), a first-order indirect effect of -0.11% (in the contiguous areas), and a second-order effect of -0.23% (in the ring around the contiguous areas), normalized by 100,000. When considering instrumental variables, the estimated direct effect on wages becomes 1.78%, the first-order effect changes to -2.12%, and the second order effect is not statistically significant.

Key words: Central Business District, Subcenter Business District, Agglomeration Economies, Labor Market, Intra-urban wage premium.

LISTA DE TABELAS

Tabela 2.1 – Divisão política e Produto Interno Bruto Total e per capita por município (em mil reais)	9
Tabela 2.2 – Produto Interno Bruto Total e per capita a preços correntes (R\$ - milhões).....	10
Tabela 2.3 – Participação relativa do PIB por macrossetor para regiões selecionadas	11
Tabela 2.4 – Posição na ocupação por regiões selecionadas	13
Tabela 2.5 – População recenseada para regiões selecionadas (em mil).....	16
Tabela 3.1 – Estatísticas descritivas	62
Tabela 3. 2 – Critério de informação AIC e tamanho das bandas	73
Tabela 3.3 – Regressões ponderadas geograficamente	74
Tabela 3.4 – Valores de corte e número de SBD identificados.....	75
Tabela 3. 5 – Análise de persistência dos SBD identificados	81
Tabela 3.6 – Total de empregos, firmas, massa salarial e área por quartil.....	87
Tabela 3.7 – Diferença do total de empregos, firmas, massa salarial e área por quantil.....	88
Tabela 3.8 – Critério de Informação AIC e Tamanho das Bandas.....	98
Tabela 3.9 – Teste de sensibilidade para regressões ponderadas geograficamente.....	99
Tabela 3.10 – Valores de Corte e total de células identificadas como SBD	99
Tabela 3.11 – Matriz de aderência dos SBD por diferentes tamanho de área das células	100
Tabela 4.1 – Estatística descritiva das variáveis.....	144
Tabela 4.2 – Matriz de correlação do salário real e densidade de emprego	144
Tabela 4.3 – Renda e distribuição dos trabalhadores do mercado formal por características	146
Tabela 4.4 – Mobilidade da firma e trabalhador	150
Tabela 4.5 – Estatística de Descritiva da Área Total Construída (m ²)	150
Tabela 4.6 – Estatística descritiva das Variáveis Instrumentais	151

Tabela 4.7 – Função kernel, critério de informação AIC e tamanho das bandas.....	154
Tabela 4.8 – Efeitos Estimados da Aglomeração sobre a Produtividade do Trabalhador	155
Tabela 4.9 – Efeitos Estimados da Aglomeração sobre a Produtividade do Trabalhador (2SLS)	160
Tabela 4.10 – Efeitos estimados da aglomeração sobre a produtividade do trabalhador com diferentes variáveis instrumentais (2SLS)	162
Tabela 4.11 – Estatística descritiva para diferentes tamanhos de células	165
Tabela 4.12 – Efeitos estimados da aglomeração sobre a produtividade do trabalhador com diferentes tamanhos de áreas das células	166
Tabela 4.13 – Efeitos da aglomeração sobre a produtividade do trabalhador por ciclo educacional	170
Tabela 4.14 – Prêmio Salarial Intra-Urbano em Diferentes SBD (2º Estágio).....	173

LISTA DE FIGURAS

Figura 2.1 – Estado de São Paulo, Municípios e Região Metropolitana de São Paulo	7
Figura 2.2 – Mapa político-administrativo da RMSP e suas sub-regiões	8
Figura 2.3 – Desmembramento dos municípios da Região Metropolitana de São Paulo	12
Figura 3. 1 – Grid da Região Metropolitana de São Paulo.....	55
Figura 3.2 – Distribuição espacial do nível de emprego (ano 2002).....	67
Figura 3.3 – Distribuição espacial do nível de emprego (ano 2008).....	67
Figura 3.4 – Distribuição Espacial do Nível de Emprego (ano 2014).....	68
Figura 3. 5 – Distribuição espacial das firmas (ano 2002)	69
Figura 3. 6 – Distribuição espacial das firmas (ano 2008)	69
Figura 3. 7 – Distribuição Espacial das Firmas (ano 2014).....	70
Figura 3. 8 – Distribuição espacial massa salarial (ano 2002)	71
Figura 3. 9 – Distribuição espacial da massa salarial (ano 2008).....	71
Figura 3. 10 – Distribuição Espacial Massa Salarial (ano 2014).....	72
Figura 3. 11 – Subcentralidades da RMSP (ano-base 2002 – quantil 99%).....	82
Figura 3. 12 – Subcentralidades da RMSP ano-base 2002 – quantil 95%	82
Figura 3. 13 – Subcentralidades da RMSP (ano-base 2002 – quantil 90%).....	82
Figura 3. 14 – Distribuição Espacial dos SBD por quantil.....	82
Figura 3. 15 – Subcentralidades da RMSP (ano-base 2008 – quantil 99%).....	83
Figura 3. 16 – Subcentralidades da RMSP (ano-base 2008 – quantil 95%).....	83
Figura 3. 17 – Subcentralidades da RMSP (ano-base 2008 – quantil 90%).....	83
Figura 3. 18 – Distribuição espacial dos SBD por quantil	83
Figura 3. 19 – Subcentralidades da RMSP ano-base 2014 – quantil 99%	84
Figura 3. 20 – Subcentralidades da RMSP ano-base 2014 – quantil 95%	84

Figura 3. 21 – Subcentralidades da RMSP ano-base 2014 – quantil 90%	84
Figura 3. 22 – Distribuição Espacial dos SBD por quantil	84
Figura 3. 23 – Distribuição espacial dos SBD por quantil – grid 2 km ²	102
Figura 3. 24 – Distribuição espacial dos SBD por quantil – grid 2 km ²	102
Figura 3. 25 – Distribuição espacial dos SBD por quantil – Grid 4 km ²	102
Figura 3.26 – Distribuição espacial dos SBD por quantil – Grid 4 km ²	102
Figura 3.27 – Distribuição espacial dos SBD por quantil – Grid 2 km ²	103
Figura 3.28 – Distribuição espacial dos SBD por quantil – Grid 4 km ²	103
Figura 4. 1 – Células do grid e vizinhanças de aglomeração	135
Figura 4. 2 – Salário médio real e densidade de emprego nas células (2002-2014).....	148
Figura 4. 3 – Salário médio real e densidade da defasagem de emprego nas células (2002-2014)	148
Figura 4. 4 – Regressão linear do salário-hora e densidade de emprego	149
Figura 4. 5 – Distribuição espacial da área construída na RMSP	152

LISTA DE GRÁFICOS

Gráfico 3.1 – Parcela de trabalhadores por anos de experiência nos SBD (ano 2002), por distrito	89
Gráfico 3.2 - Parcela de trabalhadores por anos de experiência nos SBD (ano 2014), por distrito	89
Gráfico 3.3 – Parcela de trabalhadores por tipo de contrato salarial nos SBD (base 2002), por distrito.....	91
Gráfico 3.4 – Parcela de trabalhadores por tipo de contrato salarial nos SBD (base 2014), por distrito.....	91
Gráfico 3.5 – Parcela de trabalhadores por salário mínimo nos SBD (ano 2002), por distrito	92
Gráfico 3.6 – Parcela de trabalhadores por salário mínimo nos SBD (ano 2014), por distrito	92
Gráfico 3.7 – Parcela de trabalhadores por grau de instrução nos SBD (ano 2002)	93
Gráfico 3.8 – Parcela de trabalhadores por grau de instrução nos SBD (ano 2014)	93
Gráfico 3.9 – Parcela de trabalhadores por setor nos SBD (top 5).....	94
Gráfico 3.10 – Parcela de trabalhadores por setor nos SBD (top 5).....	94

SUMÁRIO

1. INTRODUÇÃO	1
2. A REGIÃO METROPOLITANA DE SÃO PAULO	7
3. ESPRAIAMENTO DO EMPREGO NA REGIÃO METROPOLITANA DE SÃO PAULO: IDENTIFICANDO E CARACTERIZANDO SUB-CENTRALIDADES	19
3.1. ECONOMIAS DE AGLOMERAÇÃO E CENTRALIDADES DE EMPREGO NO ESPAÇO INTRA-URBANO.....	20
3.1.1. Evolução dos Modelos Teóricos sobre Centralidade do Emprego	20
3.1.2. Abordagens Empíricas de Identificação de Subcentralidades	32
3.2. METODOLOGIA	42
3.2.1. Base de Dados	42
3.2.2. Padronização e Geocodificação da RAIS	44
3.2.3. Modelo Econométrico, Método De Estimação e Estratégia de Identificação ..	53
3.3. RESULTADOS	60
3.3.1. Características da Amostra	60
3.3.2. Identificação dos Subcentros da RMSP	72
3.3.3. Caracterização dos Subcentros da RMSP	85
3.3.4. Teste de Sensibilidade	97
3.4. CONCLUSÃO	104
4. EXTENSÃO ESPACIAL DA ECONOMIA DE AGLOMERAÇÃO E ABSORÇÃO DO PRÊMIO SALARIAL INTRA-URBANO	107
4.1. ECONOMIAS DE AGLOMERAÇÃO E O PRÊMIO SALARIAL URBANO: TEORIA E EVIDÊNCIAS EMPÍRICAS	107
4.2. METODOLOGIA	122
4.2.1. Modelo de Diferencial Salarial: <i>sorting</i> e <i>match</i> importam	122

4.2.2.	Especificação Econométrica e Estratégia de Identificação	124
4.2.3.	Base de Dados	131
4.2.3.1.	Dados em Painei.....	131
4.2.3.2.	Variáveis Instrumentais.....	136
4.2.4.	Estratégia de Identificação e Métodos de Estimação.....	138
4.3.	RESULTADOS	143
4.3.1.	Características do Painei e das Variáveis Instrumentais	143
4.3.2.	Testando a Hipótese de Atenuação Espacial	153
4.3.3.	Sensibilidade dos Resultados à Extensão das Áreas	164
4.3.4.	Economias de Aglomeração e Externalidades do Capital Humano	168
4.3.5.	Heterogeneidade das Economias de Aglomeração nos Subcentros.....	172
4.4.	CONCLUSÃO.....	174
5.	CONSIDERAÇÕES FINAIS.....	177
	REFERÊNCIAS.....	181
	APÊNDICE.....	197

1. INTRODUÇÃO

A alteração da paisagem intra-urbana é dependente das alterações das atividades econômicas, infraestrutura de transporte e comunicação dentro da cidade. O surgimento de regiões de negócios sub-centrais (SBD – *Subcentral Business Districts*) é um dos fatores que altera o padrão espacial da cidade. A transformação do espaço intra-urbano é relevante devido ao papel desempenhado pelas cidades. A cidade é o locus das economias de aglomeração e é a ponte entre densidade urbana e produtividade (GLAESER; RESENGER, 2010), atraindo trabalhadores mais hábeis, firmas mais produtivas e acelerando tanto o estoque de capital humano quanto as taxas de crescimento dos salários (GLAESER; MARÉ, 2001).

Economias de escala internas e externas, custos de transporte e *commuting* são elementos principais à aglomeração de firmas e trabalhadores. Deseconomias de aglomeração mais elevadas que economias de aglomeração contribuem e dão início ao processo de espraiamento das firmas e, conseqüentemente, dos empregos. A descentralização dos empregos no espaço urbano não somente altera a paisagem intra-urbana como também impacta diretamente sobre os salários médios, preços da terra urbana, demanda por transporte e deslocamento pendular (FUJITA; OGAWA, 1982, HELSLEY; SULLIVAN, 1991, ANAS; KIM, 1996).

Esse comportamento dinâmico das economias urbanas traz consigo importantes questões àqueles que estudam economias de aglomeração e sua conexão com produtividade. Primeiramente, qual o caminho do espraiamento do emprego? A descentralização é concentrada ou as firmas estão pulverizadas no espaço? A nova disposição das firmas atua como catalizador de produtividade do trabalho? E, comparativamente ao centro histórico, quão relevante, em termo de remuneração do trabalho, é ofertar sua mão de obra na nova localização central da cidade?

As perguntas acima não são questionamentos recém colocados. O reconhecimento da relevância das economias de aglomeração data do século XIX, com o trabalho de Marshall (1890). Muitos economistas urbanos, mais tarde, tanto os pioneiros da Escola Alemã (THÜNEN, 1966; CHRISTALLER, 1966) quanto os pioneiros da Nova Economia Urbana (NUE – *New Urban Economics*) (ALONSO, 1964; MUTH, 1969; MILLS, 1967, 1972) tomaram como dado, a partir das observações das cidades reais, que as atividades econômicas estavam localizadas num único centro de negócios (isto é, no CBD – *Central Business District*), internalizando às produções das

firmas as externalidades positivas derivadas das economias de aglomeração. Todavia, evidências empíricas em grandes cidades revelam um padrão policêntrico¹, e modelos teóricos relativamente recentes passaram a considerar mais do que uma centralidade de negócios, sublinhando os efeitos do espraiamento concentrado do emprego sobre os vetores de preços (salários, preço da terra, etc.). Assim, a abordagem teórica monocêntrica é limitada quando se tem como escopo de análise grandes cidades e a abordagem teórica que considera cidades policêntricas ganha espaço.

Na literatura brasileira, o foco da descentralização dos postos de trabalho avaliou essencialmente a realocação das indústrias na escala regional, isto é, entre municípios (AZZONI, 1982; DINIZ, 1993; LEMOS, DINIZ; GUERRA, 2003). Poucos trabalhos têm-se voltado à questão da suburbanização do emprego no espaço intra-urbano, cuja avaliação foque na identificação das subcentralidades, sua expansão e caracterização (KNEIB, 2008; KNEIB, 2016; SIQUEIRA, 2014; CAMPOS; CHAGAS, 2017). Todavia, nenhum trabalho, até onde se tem conhecimento, avaliou os impactos dessa alteração no retorno salarial ou identificou o prêmio salarial em trabalhar nos SBD vis-à-vis no CBD, nem avaliou a extensão espacial do efeito da aglomeração sobre a produtividade do trabalhador.

Assim, o dinamismo intra-urbano ilustra a importância das cidades para o entendimento de Economia. Como discutido acima, tal dinamismo tem efeito sobre a desigualdade espacial dos rendimentos na escala intra-urbana. A não isotropia dos salários entre cidades pode ser relacionada a três hipóteses distintas (GLAESER; MARÉ, 1994; YANKOW, 2006). A primeira das hipóteses vincula os diferenciais salariais às dotações de amenidades locais capazes de reter trabalhadores mais qualificados nos grandes centros, comparativamente aos locais menores. A segunda hipótese relaciona o prêmio salarial urbano ao custo de vida mais elevado em habitar nos grandes centros, derivando, portanto, a desigualdade da compensação do custo de vida. A terceira hipótese sugere que trabalhadores de grandes centros seriam mais produtivos em relação aos de lugares menores,

¹ Atlanta, Buffalo, Denver, Fort Worth, Baltimore, Los Angeles, San Francisco, Cleveland, Indianapolis, Portland, Saint Louis, Houston, Chicago, Harris County, Milwaukee, Dallas, New Orleans (GREENE, 1980; GUILIANO; SMALL, 1991; SMALL; SONG, 1994; BORGART; FERRY, 1999; ANDERSON; BORGART, 2001; PAN; MA, 2005; MCDONALDS, 1987; GORDON, RICHARDSON; WONG, 1986; MCMILLEN; MCDONALD, 1990; SHUKLA; WADDELL, 1991; MCDONALDS; PRATHER, 1994; CRAIG; NG, 2001; MCMILLEN, 2001A, 2001B, 2003; READFEARN, 2007; MODARRES, 2003; KANE; HIPPI; KIM, 2016), Metropolitan Area of Barcelona (MUÑIS ET AL., 2003; LOPEZ, 2007), Dijon, Ile-de-France (BOUMONT; ERTUR; GALLO, 2004; GUILLAN; GALLO, BOITEUX-ORAIN, 2006), Manaus, São Paulo, Brasília, Goiânia (KNEIB, 2008; BARROS et al., 2014; SIQUEIRA, 2014; KNEIB, 2016; CAMPOS; CHAGAS, 2017)

sendo o ganho da produtividade derivado das economias de aglomeração. Essas mesmas hipóteses podem ser aplicadas no contexto intra-urbano.

Vários estudos empíricos têm demonstrado correlação e causalidade entre economia de aglomeração e prêmio salarial entre regiões – cidades, áreas de mercado de trabalho e/ou países (COMBES; GOBBILON, 2015; MELO; GRAHAM; NOLAND, 2009). As estimações do diferencial salarial para as regiões brasileiras mostram que o prêmio salarial urbano não é negligenciável (ROCHA; SILVEIRA NETO; GOMES, 2011; CRUZ; NATICCHIONI, 2012; SILVA; SANTOS; FREGUGLIA, 2016; BARUFI; HADDAD; NEJKAMP, 2016; SILVA, 2017). Azzoni e Servo (2001) e Menezes e Azzoni (2006) mostram que a desigualdade salarial entre regiões metropolitanas é persistente mesmo após controlar pelo custo de vida regional e outras covariadas observáveis. Os autores mostram também que a diferença da desigualdade antes e após controle do custo de vida é de pequena monta, sugerindo que os maiores retornos salariais nas RM brasileiras não estão relacionados ao custo de vida.

A maioria dessas estimações circunscrevem-se ao efeito da aglomeração em escala regional (agregada), o que impossibilita entender qual a extensão do efeito da aglomeração sobre a produtividade do trabalhador no espaço intra-urbano (ROSENTHAL; STRANGE, 2003, 2004; OVERMAN, 2004). Rosenthal e Strange (2004) apontam para efeitos maiores da economia de aglomeração no espaço intra-urbano quando comparado às avaliações agregadas. Tal resultado vai ao encontro dos resultados encontrados nos estudos relacionados ao emprego (HENDERSON, 2003) e transbordamento de conhecimento que se dão via inovações e patentes (JAFTE; TRAJTENBEG; HENDERSON, 1993; AUDRETSCH; FELDMAN, 1996; ANSELIN; VARGAS; ACS, 1997). Poucos trabalhos na literatura internacional (ROSENTHAL; STRANGE, 2003, 2004; VAN SOEST; GERKING; VAN OORT, 2006; ROSENTHAL; STRANGE, 2008; ANDERSSON; KLAESSON, LARSSON, 2016) e nacional (HADDAD; BARUFI, 2017) tiveram como objetivo avaliar a relação aglomeração-produtividade do trabalhador na escala intra-urbana.

Os resultados encontrados pelos trabalhos que estimam o efeito da aglomeração no espaço intra-urbano apontam para a sua limitada extensão espacial. As estimações nos trabalhos de Rosenthal e Strange (2003, 2004); van Soest; Gerking; van Oort. (2006); Rosenthal e Strange (2008) e Andersson, Klaesson e Larsson (2016) apontam que o escopo geográfico das externalidades da

aglomeração é menor do que o tamanho de uma cidade (região metropolitana). Tais achados abrem espaço para discutir os resultados encontrados para os efeitos da aglomeração numa escala agregada, dado que tal identificação não permite capturar a heterogeneidade espacial intra-urbana, estimando um efeito homogêneo intra-urbano em economias desiguais, como é o caso da Região Metropolitana de São Paulo. Essa lacuna motiva o estudo sobre a extensão e escopo geográfico dos prêmio salarial intra-urbano na RMSP.

Estimativas para a Região Metropolitana de São Paulo mostram que o custo do distanciamento do centro histórico (gradientes de emprego) reduziu-se, saindo de 0,212 em 1987 e alcançando as marcas de 0,149 e 0,125 em 1997 e 2007, respectivamente (NADALIN, 2010)². Os resultados do gradiente da população também apontam para a queda dos valores ao longo do tempo, isto é, 0,144, 0,114 e 0,064 para os anos de 1987, 1997 e 2007, respectivamente. Esses resultados sugerem que o emprego se espalhou mais rapidamente do que a população, e os custos do distanciamento reduziram-se ao longo do tempo. No entanto, quando as densidades são consideradas, isto é, normalizadas pelas áreas de ponderação da Pesquisa Origem-Destino, a densidade média da população é maior do que a densidade média de emprego. Segundo Nadalin (2010), esse padrão de crescimento da densidade de emprego é comum em cidades onde o emprego é mais concentrado do que a população. A mudança da distribuição espacial do emprego na RMSP abre espaço para discutir o efeito sobre a remuneração do trabalhador e economias de aglomeração no espaço intra-urbano intertemporalmente.

Dado que o mercado de trabalho é espacialmente limitado (GLAESER; RESSEGER; TOBIO, 2009) e segue dinâmicas próprias, é essencial estudar as dinâmicas intra-urbanas para ter melhor compreensão do efeito da aglomeração sobre o salário do trabalhador. Neste trabalho propõe-se uma abordagem que permita identificar tal efeito numa escala mais fina possível. Para tanto, criou-se um grid da RMSP cujas células têm 1 km² cada.

Essa abordagem empírica permite identificar, de alguma forma, o mecanismo de ativação das economias de aglomeração colocado por Jacobs (1969). Segundo a autora, a importância da diversidade urbana como fonte de geração e difusão de ideias se dá via contato *face-to-face*. A

² No trabalho, Nadalin (2010) considera o marco zero localizado na Praça da Sé, no município de São Paulo, como âncora do modelo econométrico. A autora estima duas equações diferentes para calcular o custo desse distanciamento, isto é, os gradientes do emprego e da população.

partir dessas relações justifica-se a aglomeração e essa, por sua vez, tem efeito direto sobre o salário do trabalhador. Tal hipótese rebate o proposto por Marshall (1890), para quem a fonte da aglomeração deriva da especialização das cidades e concentração de uma indústria. Nesse contexto, o retorno sobre a produtividade não é da diversificação, mas sim da concentração. Essas hipóteses antagônicas deram origem ao debate Marshall-Jacobs, sendo a primeira hipótese conhecida como economia de urbanização e a segunda como economia de especialização. Assim, a estratégia empírica proposta nesse trabalho permite, de alguma forma, discutir o efeito *face-to-face* da economia de aglomeração, ao aninhar trabalhadores em células com 1 km² de área.

Dois pontos têm relevância ao avaliar o dinamismo do mercado de trabalho intra-urbano e norteiam o esforço dessa pesquisa. Primeiro, faz-se necessário compreender o processo de suburbanização, identificando suas novas localizações espaciais e sua posição relativa às demais regiões da RMSP, dado que tanto a suburbanização quanto o aprofundamento da concentração de emprego afetam os salários do trabalhador, bem como outros preços da economia. Nesse contexto, o segundo ponto relevante é: qual a importância das economias de aglomeração sobre o salário no espaço? Questionando de outra forma: qual a extensão do efeito da economia de aglomeração no espaço intra-urbano em um contexto de suburbanização do emprego? Essas questões motivam as páginas que seguem e justificam o esforço de pesquisa.

2. A REGIÃO METROPOLITANA DE SÃO PAULO

A Região Metropolitana de São Paulo (RMSP), também conhecida como Grande São Paulo, é a maior região metropolitana do Brasil. Localizada no estado de São Paulo, sudeste do Brasil, possuía mais de 19.683 milhões de habitantes em 2010 (Censo Demográfico do IBGE, 2010), representando 47,7% da população do estado e 10,3% da população do país. Apenas o município de São Paulo é responsável por 57% do total da população da RMSP.

Figura 2.1 – Estado de São Paulo, Municípios e Região Metropolitana de São Paulo

Fonte: Elaboração própria a partir da cartografia do IBGE

A RMSP é composta por 39 municípios e ocupa uma área de 7.946 km². Em termos geográficos, os municípios que compõem atualmente a RMSP derivaram essencialmente de dois grandes municípios criados em 1532 (São Paulo) e em 1611 (Mogi das Cruzes). A última criação de município foi em 1991 (São Lourenço da Serra), derivado da subdivisão do município de Itapequerica da Serra. A Tabela 2.1 apresenta informações sobre a divisão política da RMSP e os

anos de fundação de cada município e a Figura 2.3 exibe a série histórica do desmembramento dos municípios que compõem a atual configuração da Grande São Paulo. A Figura 2.2 expõe o mapa político da RM e suas sub-regiões, definidas pela lei complementar estadual 1.139, de junho de 2011. O Município de São Paulo é dividido em 8 macro-zonas: Centro, Zona Leste 1, Zona Leste 2, Zona Norte 1, Zona Norte 2, Zona Oeste, Zona Sul 1 e Zona Sul 2, que são subdivididos em 96 distritos

Figura 2.2 – Mapa político-administrativo da RMSP e suas sub-regiões

Fonte: Elaboração própria a partir da cartografia do IBGE

Nota: A relação entre acrônimos e municípios pode ser encontrados na Tabela 2.1

Tabela 2.1 – Divisão política, Ano de surgimento do município e Produto Interno Bruto Total e per capita por município (em mil reais)

Municípios	Sigla	Sub-regiões	Ano	Área (Km ²)	PIB ⁽¹⁾ (R\$)	PIB per capita ⁽²⁾
São Paulo	SAO	Capital	1558	1521,1	628.064.882	54.549
Arujá	ARU	Leste	1959	96,1	3.759.210	46.501
Biritiba-Mirim	BBM	Leste	1964	317,4	704.945	23.450
Ferraz de Vasconcelos	FVA	Leste	1953	29,6	2.747.746	15.423
Guararema	GMA	Leste	1898	270,8	1.555.846	57.099
Guarulhos	GRU	Leste	1880	318,7	51.389.524	40.320
Itaquaquecetuba	IQQ	Leste	1953	82,6	5.742.348	16.852
Mogi das Cruzes	MJC	Leste	1611	712,7	13.367.335	32.928
Poá	POA	Leste	1948	17,3	3.940.620	35.823
Salesópolis	SPS	Leste	1857	425,0	182.389	11.326
Santa Isabel	SIS	Leste	1832	363,3	1.450.356	27.695
Suzano	SNO	Leste	1948	206,2	10.130.268	36.991
Caieiras	CAI	Norte	1959	96,1	3.801.890	41.261
Cajamar	CJM	Norte	1959	131,3	9.911.054	142.505
Francisco Morato	FMO	Norte	1964	49,1	1.268.410	7.827
Franco da Rocha	FRO	Norte	1944	134,2	2.563.525	18.313
Mairiporã	MRP	Norte	1889	320,7	1.500.945	17.027
Barueri	BAR	Oeste	1948	65,7	46.151.952	184.256
Carapicuíba	CBA	Oeste	1964	34,5	4.719.835	12.407
Itapevi	IVI	Oeste	1959	82,7	8.867.669	41.238
Jandira	JAN	Oeste	1964	17,5	2.941.506	25.706
Osasco	OSA	Oeste	1959	65,0	58.566.199	87.193
Pirapora do Bom Jesus	PBJ	Oeste	1959	108,5	360.672	21.319
Santana de Parnaíba	STP	Oeste	1625	179,9	8.065.197	66.784
Diadema	DIA	Sudeste	1959	30,8	13.910.517	35.294
Mauá	MAU	Sudeste	1953	61,9	11.329.503	26.035
Ribeirão Pires	RPI	Sudeste	1953	99,1	2.606.481	22.532
Rio Grande da Serra	RGS	Sudeste	1964	36,3	519.828	11.221
Santo André	SAD	Sudeste	1889	175,8	28.119.591	41.128
São Caetano do Sul	SCS	Sudeste	1948	15,3	16.153.419	107.461
São Bernardo do Campo	SBC	Sudeste	1944	409,5	47.551.620	60.492
Cotia	COT	Sudoeste	1856	324,0	10.118.348	46.016
Embu das Artes	EMB	Sudoeste	1959	70,4	7.412.777	29.331
Embu-Guaçu	EGU	Sudoeste	1964	155,6	948.095	14.613
Itapeçerica da Serra	ITS	Sudoeste	1877	150,9	2.943.707	18.461
Juquitiba	JBA	Sudoeste	1964	522,2	422.949	14.411
São Lourenço da Serra	SLS	Sudoeste	1991	186,3	215.394	14.758
Taboão da Serra	TBS	Sudoeste	1959	20,4	7.326.979	28.143
Vargem Grande Paulista	VGP	Sudoeste	1981	42,5	1.532.991	32.672

Fonte: Instituto Brasileiro de Geografia e Estatística (IBGE); Fundação Seade. (1) O PIB do Município é estimado somando os impostos ao VA total e faz referência ao ano de 2014. (2) O PIB per capita referente às regiões paulistas foi estimado usando a população estimada pela Fundação Seade.

A RMSP representa o maior pólo de concentração de produção do país, em termos monetários. O Produto Interno Bruto (PIB) da RM em 2014 foi de R\$ 1,022 trilhões, representando 55% do PIB do estado de São Paulo e 17,7% do PIB nacional, sendo que o produto per capita foi de R\$ 40.189

(Tabela 2.2). Quando comparado ao ano de 2002, nota-se a perda de participação relativa do PIB da RMSP e do Município de São Paulo (MSP) no produto nacional.

Tabela 2.2 – Produto Interno Bruto Total e per capita a preços correntes (em mil reais)

Unidades Territoriais	2002			2014		
	PIB ⁽¹⁾	PIB per Capita ⁽²⁾	% PIB	PIB ⁽¹⁾	PIB per Capita ⁽²⁾	% PIB
Brasil	1.488.787.276	8,53	-	5.778.952.780	28,50	-
Estado de São Paulo	518.878.815	13,688	35%	1.858.196.056	43,545	32,2%
RMSP	292.547.475	12,575	20%	1.022.866.523	40,189	17,7%
Município de São Paulo	188.706.119	17,779	13%	628.064.882	54,549	10,9%

Fonte: Instituto Brasileiro de Geografia e Estatística (IBGE); Fundação Seade. (1) O PIB do Município é estimado somando os impostos ao VA total. (2) O PIB per capita referente às regiões paulistas foi estimado usando a população estimada pela Fundação Seade, e referente ao Brasil foi calculado usando a estimativa do IBGE.

O principal macro setor da produção na Grande São Paulo é o setor de serviços, cuja participação no PIB da região, no ano de 2014, foi de 60% (excluindo o setor público), seguido pela indústria (14%). Essa alta participação relativa do setor de serviços no PIB deriva da presença de muitas sedes de complexos industriais, comerciais e financeiros que atuam tanto na região metropolitana quanto no restante do Brasil. O município de São Paulo, principalmente, tem um papel importante como centro financeiro e atua como ponte regional entre o Brasil e o restante do mundo. Ao mesmo tempo, a municipalidade conecta o restante da RM às relações internacionais, reforçando o papel regional de centro de negócios internacionais e recebendo demanda por serviços e fluxos e controle de investimentos estrangeiros e comércio internacional (DINIZ; CAMPOLINA, 2007).

A comparação da participação relativa do PIB por macrossetor da RMSP com as respectivas participações no Brasil revela elevada importância do setor de serviços da RMSP (puxada pelo Município de São Paulo, como destacado acima), isto é, 25,2% da produção do setor de serviços nacional. Em relação à indústria, a RMSP produz 16,5% do total da produção desse macrossetor na esfera nacional, sendo que o MSP contribui com quase 50% dessa parcela. Essas relações foram reduzidas após 12 anos, para todos os macrossetores.

A Tabela 2.3 demonstra a redução das participações relativas, não somente nas parcelas dos PIBs setoriais do município de São Paulo em relação ao Brasil, mas da RMSP e do Estado de São Paulo. Dentre essas três regiões, a capital do estado apresentou a maior perda relativa no macrossetor de serviços, excluindo o setor público. Normalizando pela produção do Estado de São Paulo, o MSP apresentava produção relativa de 45,6% em 2002, enquanto em 2014 essa mesma relação reduziu-se para 39,8%. Embora a participação relativa da produção do macrossetor da indústria caia quando

normalizado pela respectiva produção nacional, pode-se notar que o município de São Paulo é a região que apresentou maior produção relativa, quando normalizado pela respectiva produção do Estado de São Paulo. Resumidamente, o Estado de São Paulo tem perdido importância relativa em termos de produção, o que significa que outros estados estão crescendo a taxas maiores comparativamente. O mesmo tem ocorrido com a RMSP e, especificamente, com o MSP numa análise intra-estado.

Tabela 2.3 – Participação relativa do PIB por macrossetor para regiões selecionadas

Região	Agro	Ind.	Serviços		Agro	Ind.	Serviços	
			Adm. Púb.	Total (exclusive Adm. Púb.)			Adm. Púb.	Total (exclusive Adm. Púb.)
2002								
Brasil	100,0%	100,0%	100,0%	100,0%	-	-	-	-
Estado de São Paulo	17,3%	35,3%	20,1%	39,8%	100,0%	100,0%	100,0%	100,0%
RMSP	0,4%	16,5%	9,3%	25,2%	2,6%	46,7%	46,3%	63,4%
São Paulo	0,01%	8,1%	5,3%	18,1%	0,1%	10,1%	26,2%	45,6%
2014								
Brasil	100,0%	100,0%	100,0%	100,0%	-	-	-	-
Estado de São Paulo	10,9%	28,9%	18,5%	37,9%	100,0%	100,0%	100,0%	100,0%
RMSP	0,6%	12,1%	8,4%	22,9%	5,1%	42,0%	45,3%	60,3%
São Paulo	0,02%	5,6%	4,6%	15,1%	0,1%	19,6%	25,1%	39,8%

Fonte: Elaboração própria a partir dos dados do Instituto Brasileiro de Geografia e Estatística (IBGE). Nota: Agro: setor agrícola; Ind: indústria; Adm. Púb.: administração pública.

Figura 2.3 – Desmembramento dos municípios da Região Metropolitana de São Paulo

Fonte: Elaboração própria a partir dos dados da Fundação Seade

Tabela 2.4 – Posição na ocupação por regiões selecionadas

Posição na Ocupação	Ocupados			Ocupados (%)		
	2002	2008	2014	2002	2008	2014
Formal						
Brasil	30.048.673	40.482.412	47.835.655	37,7%	43,3%	48,2%
Estado São Paulo	8.945.305	11.945.365	13.523.969	51,6%	57,4%	63,4%
RMSP	4.223.435	5.712.496	7.627.818	50,1%	57,5%	67,5%
Informal						
Brasil	19.349.719	20.926.430	19.002.628	24,3%	22,4%	19,2%
Estado São Paulo	3.861.166	3.931.675	2.601.109	22,3%	18,9%	12,2%
RMSP	1.962.996	1.979.189	1.725.512	23,3%	19,9%	15,3%
Conta Própria						
Brasil	17.746.858	18.911.737	21.171.473	22,3%	20,2%	21,4%
Estado São Paulo	3.250.038	3.345.447	3.922.012	18,7%	16,1%	18,4%
RMSP	1.611.461	1.616.853	1.879.314	19,1%	16,3%	16,6%
Empregador						
Brasil	3.380.150	4.189.651	3.728.845	4,2%	4,5%	3,8%
Estado São Paulo	867.643	1.067.840	979.422	5,0%	5,1%	4,6%
RMSP	397.162	461.249	415.760	4,7%	4,6%	3,6%
Outros						
Brasil	9.183.122	8.910.132	7.404.892	11,5%	9,5%	7,5%
Estado São Paulo	426.059	502.630	306.349	2,5%	2,4%	1,4%
RMSP	229.005	169.557	30.974	2,7%	1,7%	0,3%
Total						
Brasil	79.708.522	93.420.362	99.143.493	100,0%	100,0%	100,0%
São Paulo	17.350.211	20.792.957	21.332.861	100,0%	100,0%	100,0%
RMSP	8.424.059	9.939.344	11.305.194	100,0%	100,0%	100,0%

Fonte: Elaboração própria a partir dos dados da PNAD (2002; 2008; 2014). Nota: As posições na ocupação foram criadas a partir das seguintes agregações: a) formal: empregados e trabalhadores com carteira assinada; b) informal: trabalhadores sem carteira assinada; c) as demais categorias foram criadas pelas suas respectivas denominações, como consta no questionário da PNAD.

Mesmo em um cenário de perda de participação relativa do PIB setorial da RMSP, a taxa de crescimento de trabalhadores ocupados em empregos formais cresceu, relativamente ao total de trabalhadores no Brasil, como se observa nos dados da Tabela 2.4. Ainda que tal crescimento tenha sido observado nos dados agregados do Brasil e do Estado de São Paulo, a RMSP foi a que apresentou maiores taxas de crescimento da participação relativa nos anos de 2002, 2008 e 2014. Nota-se também a redução do trabalho informal no Brasil. Em 2002 a participação dos trabalhadores informais no total de trabalhadores do país era de 24,3%, sendo reduzida para 22,4% em 2008 e 19,2% em 2014. Avaliando a mesma posição na ocupação para a RMSP, as reduções são mais expressivas. Isto é, em 2002, a região obtinha 23,3% dos trabalhadores em ocupações informais (em relação ao Brasil), taxa que cai para 19,9% em 2008 e para 15,3% em 2014. Em relação à posição conta própria, a taxa de declínio é suave ao longo do período analisado e a participação relativa da RMSP nessa categoria segue a mesma tendência nacional. O mesmo ocorre

em relação à categoria de empregadores, em que a RMSP apresenta suave queda na participação relativa, saindo de 4,7% do total de ocupados em 2002 para 3,6% em 2014. Enquanto, em relação ao Brasil, a relação era de 4,2%, em 2002, caindo para 3,8% em 2014.

Quanto à pujança econômica, vários estudos tiveram como foco estudar os fatores que transformaram a Região Metropolitana de São Paulo no maior pólo de riqueza e renda nacional. A história do desenvolvimento econômico da RMSP está totalmente vinculada ao ciclo do café (1850-1930), principal produto exportado neste período, resultando na atração de mão de obra de diferentes lugares do país e de imigrantes para trabalhar nas plantações e em outras atividades impulsionadas pela economia cafeeira, tais como comércio e serviços. Segundo Fausto (1994), imigração, urbanização e industrialização são processos complementares que se deram concomitantemente e foram impulsionados pela economia do café.

No contexto da abolição da escravidão, a demanda por mão de obra cresceu nas plantações de café, e os imigrantes que vinham ao país procurando por uma nova posição acabavam se mudando para as recém-criadas áreas urbanas. O processo de urbanização esteve, em geral, relacionado à distribuição e ao fluxo de café direcionados à exportação, o que incentivou a construção de linhas de trem conectando com o porto. De acordo com Fausto (1994), o imigrante contribuiu com o surgimento do emprego urbano remunerado e abriu novas oportunidades para diferentes fontes de renda e empregados liberais.

A combinação da nova formatação do mercado de trabalho trazida pelos imigrantes com a estrutura urbana desenvolvida pela economia do café forneceu um ambiente apropriado para o fluxo de produtos manufaturados e para o desenvolvimento de indústrias nas áreas urbanas. Inicialmente, tais indústrias eram dependentes da indústria do café e tiveram como foco a oferta de produtos para abastecer a demanda nacional. Neste contexto, muitas máquinas foram importadas e muitos produtores de café eram os demandantes (Fausto, 1994).

Segundo Anas, Arnott e Small (1999), a estrutura espacial das cidades modernas depende em grande escala do sistema de transporte e comunicação. Assim, a proximidade da RMSP do porto de Santos, a construção do modal ferroviário (linhas de trem) e estradas (SP 150, SP 160, BR 116, SP 070, BR 381, SP 330, SP 348, SP 280, SP 270 e BR 116) criaram a infraestrutura que facilitou o transporte de bens, a comunicação entre empresários e políticos e o acesso à mão de obra. Diniz

e Campolina (2007) apontam tais conexões como essenciais para a transformação da região como uma área central no contexto nacional.

As estruturas urbanas desenvolvidas, associadas às políticas públicas de incentivo à industrialização na era Vargas (1930-1945) e nos anos 1950 (Processo de Substituição de Importações), contribuíram para a concentração de investimentos na região da RMSP e/ou nas proximidades, tais como Santos, Vale do Paraíba e Campinas. Segundo Singer (2017), tal concentração espacial das indústrias promoveu a expansão do setor de serviços, tais como atividades comerciais e financeiras, educação formal (escolas e cursos profissionais), hotéis, salões de beleza, etc. Na década seguinte ao fim do ciclo do café, a taxa de crescimento de emprego no setor de serviços na RM cresceu por volta de 50% entre 1940-50, e 60% entre as décadas 1950-1960 (SINGER, 2017).

Todavia, segundo Diniz e Campolina (2007), a partir dos anos 1970, a RMSP passa a perder sua capacidade de atrair novas firmas. Embora a taxa de industrialização tenha crescido na RMSP, essa não cresceu na mesma magnitude que o país. Apesar do crescimento com taxas declinantes, a RMSP continuou atraindo imigrantes, contribuindo para a elevada taxa de crescimento demográfico. Durante os anos 1980 teve início o processo de redução do crescimento da ocupação da indústria na produção total, em termos absolutos, bem como na taxa de crescimento da população.

Como descrito acima, o adensamento populacional na RMSP teve início no contexto do ciclo cafeeiro e foi mantido no período de industrialização do país sob o fomento de políticas públicas, tais como as leis trabalhistas desenvolvidas na era Vargas, que atraíram trabalhadores da zona rural para a área urbana (Singer, 2017). Durante os anos 1940, a população de 1.326.261 habitantes não ocupava todo o território da capital do estado. Todavia, após 30 anos, em 1970, a área urbana se espalhava por 37 municípios (dado que os municípios de São Lourenço da Serra e Vargem Grande Paulista foram criados em 1981 e 1991, respectivamente) e a população total era de 8.139.730 habitantes (Censo Demográfico do IBGE, 1970). Assim, a população cresceu cerca de 6,1 vezes em 30 anos, uma taxa média de crescimento anual de 5,5%.

A partir da Tabela 2.5, nota-se que entre os anos de 1980 e 2010 a população da RMSP cresceu de 12,6 para 19,7 milhões de habitantes. Embora durante os anos 1970 a concentração industrial tenha

ocorrido fora da RM, reduzindo a atração de pessoas por emprego na região, a Grande São Paulo permaneceu atraindo pessoas. Diniz e Campolina (2007) justificam esse comportamento devido a um atraso temporal no comportamento dos imigrantes para ajustar o processo de polarização econômica reversa e a polarização demográfica. Todavia, a taxa de crescimento populacional caiu significativamente na RM, a taxa de crescimento foi de 22,7% na comparação 1980-1991, alcançando a marca de 10,1% na comparação 2000-2010. A capital do estado teve sua taxa de crescimento reduzida a cada recenseamento, como exposto na Tabela 2.5; isto é, 13,6% (1991), 8,2% (2000) e 7,9% (2010). Comparativamente ao Brasil e ao estado de São Paulo, nota-se que tanto a região metropolitana quanto o município de São Paulo cresceram a taxas menores.

Tabela 2.5 – População recenseada para regiões selecionadas (em mil)

Unidades Territoriais	1980	1991	2000	2010	1991/1980	2000/1991	2010/2000
Brasil	119.003	146.825	169.799	190.756	23,4%	15,6%	12,3%
Estado de São Paulo	25.041	31.589	37.032	41.263	26,2%	17,2%	11,4%
Região Metropolitana de São Paulo	12.589	15.445	17.879	19.684	22,7%	15,8%	10,1%
Município de São Paulo	8.493	9.646	10.434	11.254	13,6%	8,2%	7,9%

Fonte: Elaboração própria a partir dos dados da Fundação Seade.

De acordo com o padrão de crescimento demográfico estilizados em alguns modelos teóricos (ALONSO, 1964; MUTH, 1969; MILLS, 1967, 1972), a população tende a se concentrar nas regiões centrais. Todavia, com a ampliação da densidade demográfica, os custos de aglomeração tendem a sobrepor aos retornos positivos da aglomeração, atraindo pessoas para municípios (regiões) mais próximos do município central (Leme, 1981), mas não exatamente na área central, como observado por Nadalin (2010) na RMSP. Segundo Nadalin (2010), existe uma tendência de elevação da concentração de emprego e habitação em relação ao centro histórico localizado na Praça da Sé (centro histórico da capital paulista), distrito Sé do Município de São Paulo. Os custos da aglomeração não afetam apenas as famílias, mas também as firmas, as quais buscam por localização com menores preços da terra e menor custo de transporte. Assim, a troca do município de São Paulo pelos demais municípios da RMSP deu-se devido à expansão e melhoramento do sistema de transporte, bem como em função do custo da aglomeração. Azzoni (1982), Diniz (1993), Lemos, Diniz e Guerra (2003) e Diniz e Campolina (2007) apontam para a descentralização dos postos de trabalho em função da realocação das indústrias, dando-se essencialmente na escala regional, isto é, movendo-se da capital em direção às demais municipalidades da Grande São Paulo.

Meyer, Grostein e Biderman (2004) destacam o mesmo processo, mas agora na escala intra-municipal, i.e., no município de São Paulo. Os autores apontam para o espraiamento dos postos de

emprego, saindo do centro histórico da cidade na direção das regiões da Zona Sul 1 e Oeste. Segundo os autores, o processo se inicia nos anos 1960 em direção aos distritos do Jardim Paulista, Cerqueira Cezar e Itaim Bibi. Este último foi essencialmente estimulado pela criação da Avenida Faria Lima nos anos 1970 e entre os anos 1970-1990 muitas firmas foram alocadas no distrito da Vila Olímpia. Recentemente, entre 2000-2010 os empregos estão se aglomerando em direção ao distrito de Santo Amaro, ao sul do município, e ao distrito da Lapa, zona oeste da cidade. Tal padrão pode ser visualizado nos mapas apresentados no próximo capítulo.

Portanto, essa alteração intertemporal da ampliação da formalidade no mercado de trabalho, associada com o espraiamento das firmas ao longo da RMSP, tem impacto direto sobre a economia local, sobretudo no que diz respeito à Economia Urbana. Nesse sentido, compreender esse espraiamento e seu grau de concentração ganha relevância devido à elevada influência que a concentração de empregos tem sobre o vetor de preços em um contexto urbano, tais como sobre salários, preços da terra, custos de transporte, entre outros. O capítulo seguinte explora tal relevância de forma mais focalizada e detalhada, bem como a comparatividade intertemporal da concentração dos postos de emprego.

3. ESPRAIAMENTO DO EMPREGO NA REGIÃO METROPOLITANA DE SÃO PAULO: IDENTIFICANDO E CARACTERIZANDO SUB-CENTRALIDADES

Nesse capítulo objetiva-se identificar regiões centrais de emprego na Região Metropolitana de São Paulo e apontar o comportamento intertemporal dessas regiões e caracterizar essas áreas em função da remuneração, gênero, tipo de contrato empregatício, grau de instrução (anos de educação) e experiência (anos de trabalho na mesma firma). Para alcançar tais objetivos, propõe-se uma metodologia de identificação objetiva de subcentralidades, superando as abordagens empíricas que demandam conhecimento prévio da região estudada. Em função das críticas concernentes aos vieses existentes quanto às áreas geográficas utilizadas, os microdados da RAIS foram georreferenciados e agregou-se o total de trabalhadores formais por células com áreas de 1, 2 e 4 km². Os resultados apontam para a existência de três SBD na RMSP, quando utilizados critérios mais restritos de avaliação. Ao relaxar esses critérios, nota-se a ampliação do total de subcentros, a tal ponto de criar uma grande e única mancha de emprego na região. Nesse sentido, os resultados sugerem um movimento centrífugo do emprego, mas de forma concentrada, sugerindo que a aglomeração de trabalhadores é um fator importante para a produtividade.

Este capítulo está dividido em quatro subseções. Na primeira são discutidas as definições e surgimento de centro (CBD) e subcentros (SBD) de negócios, bem como a escolha de localização espacial ótima das firmas em áreas urbanas. Inicialmente, discutem-se tais definições considerando os principais modelos desenvolvidos pelos pioneiros da Economia de Localização Espacial e da Nova Economia Urbana. Reserva-se ainda à discussão sobre as abordagens empíricas recorrentes na literatura, sublinhando as limitações e avanços dos métodos e as variáveis comumente utilizadas para identificar CBD e SBD. A segunda subseção está destinada à discussão sobre a base de dados, tais como procedimentos utilizados para limpeza e geocodificação. Adicionalmente, abordam-se a especificação econométrica, identificação e estimador utilizados nesse trabalho. Na terceira seção, discutem-se as principais localizações identificadas como subcentros de negócios, que são caracterizadas considerando características do mercado de trabalho formal da RMSP. Testes de sensibilidade são executados visando avaliar o quanto as regiões identificadas são sensíveis às alterações do tamanho das células. A última subseção traz as principais discussões desenvolvidas, bem como aponta para limitações e possíveis extensões quanto à identificação de centralidades.

3.1. ECONOMIAS DE AGLOMERAÇÃO E CENTRALIDADES DE EMPREGO NO ESPAÇO INTRA-URBANO

3.1.1. Evolução dos Modelos Teóricos sobre Centralidade do Emprego

Uma extensa literatura discute o papel dos CBD e SBD na economia e como essas áreas impactam salários, preço da terra e sistema de transporte. No entanto, as definições de SBD não são apresentadas de forma consensual na literatura, sobretudo no que se refere aos trabalhos empíricos que buscam identificá-los.

A definição de CBD é menos complexa relativamente à conceitualização de SBD. É menos complexa porque a maioria das cidades apresenta apenas um centro de negócios e tal condição garante a essa região concentradora de trabalhadores a denominação de CBD. Todavia, a questão ganha maior relevância quando se trata de cidades com múltiplas concentrações de emprego. O processo de suburbanização do emprego dá lugar a uma cidade com múltiplos centros. Todavia, o conceito de CBD está intrinsicamente relacionado ao fator temporal, isto é, ao surgimento da primeira região central. Assim, o CBD está associado à área histórica da cidade. Resultados derivados de modelos teóricos monocêntricos associam CBD à área com os mais elevados preços da terra e salários quando comparado a quaisquer outras regiões intra-urbanas.

Os autores partem das características que consideram principais dos SBD para conceituar a área. Para um grupo de pesquisadores, o subcentro é um lugar onde as atividades de produção, compras e acessibilidade do emprego contam com melhor sistema de transporte do que em qualquer outro lugar, proporcionando facilidade de saída e entrada de insumos e produtos e deslocamento pendular, fora do centro histórico (OGAWA; FUGITA, 1980; FUJITA; OGAWA, 1982; HELSLEY E SULLIVAN, 1991; MCMILLEN, 2001B). Para esse grupo, a centralização das atividades no SBD está relacionada a um bom sistema de transporte, que oferece acessibilidade eficiente. Alguns trabalhos empíricos utilizam informações sobre o fluxo de passageiros para identificar subcentralidades, considerando tal peculiaridade inerente às áreas centrais (BOURNE, 1989; BURNS; BOIX, ROCA, 2001; CLARK; KUIJPERS-LINDE, 1994; GORDON; RICHARSON, 1996; AGUILERA; MIGNOT, 2004)

No segundo grupo de pesquisadores, SBD é a região cuja concentração de firmas apresenta fortes encadeamentos produtivos e de comunicação (CASTELLS; HALL; HALL, 1994; ANAS; KIM,

1996; SASAKI; MUN, 1996; GRAHAM; MARVIN, 1996). Nessa área tendem a se aglomerar atividades de alta ordem e da administração pública e privada, bem como o sistema de transporte central. Essas atividades exigem contatos face a face e criam algumas ligações para trás e para frente, dada a concentração da atividade do setor de serviços.

Partindo dos modelos teóricos desenvolvidos por autores do primeiro grupo, os SBD são áreas que impactam os preços relativos e concentram os mais elevados aluguéis e salários do que qualquer outro lugar, exceto no CBD (PAPAGEORGIOU, 1971; WHITE, 1976, 1988 E 1990; HARTWICK E HARTWICK, 1974; ROMANOS, 1979; SULLIVAN, 1986; WIEAND, 1987; SIVITANIDOU; WHEATON, 1992; HOTCHKISS; WHITE, 1993; YINGER, 1992; ROSS; YIENGER, 1995; WREDE, 2015). Com base nessa característica, alguns autores usam o preço da terra como variável de resultado para identificar subcentralidades (HEIKKILA; GORDON; KIM, 1989; CAMPOS; CHAGAS, 2017). Assim, devido à disponibilidade de infra-estrutura, compras e proximidade de locais de trabalho, os preços da terra nessas localidades são relativamente maiores, comparados a outros espaços intra-urbanos.

O processo denominado de suburbanização implica em redução na fração da população ou emprego localizado na área central de negócios (CBD), enquanto que as atividades econômicas e a população aumentam nos locais circundantes. Essas novas centralidades (ou subcentralidades) surgem como subsidiárias do centro histórico da cidade (ANAS; ARNOTT; SMALL, 1998). Contudo, não há qualquer fator que garanta que sempre serão subsidiárias, isto é, nada impede que não possa haver uma reversão de importância relativa entre CBD e SBD, tornando o CBD subsidiário de uma nova área concentradora de emprego. Especialmente se o conceito de CBD estiver intrinsicamente relacionado ao vetor de preços espaciais, pois a suburbanização das atividades pode alterar a âncora de preços espacial.

Portanto, neste trabalho, o conceito de CBD está intrinsicamente relacionado ao processo histórico, isto é, o CBD da Região Metropolitana de São Paulo está localizado no distrito da Sé (SBD-SE), em consonância com os trabalhos de Biderman (2001), Maciel e Biderman (2013) e Campos (2017). Entende-se por SBD *clusters* de áreas contínuas que apresentam concentração de emprego mais elevada quando comparada às demais áreas da RMSP.

Economias de aglomeração derivam da concentração de atividades econômicas e de pessoas, sustentada por uma causação circular que ocorre em vários níveis geográficos, desde (sub) centros – CBD e SBD – rodeados por residências até regiões economicamente especializadas, como o Vale do Silício, nos USA. A circularidade lógica deriva de consumidores potenciais que se dirigem às regiões aglomeradas devido à diversidade de produtos ofertados, e lojas/firmas que se localizam em tais aglomerações por considerarem o acesso aos consumidores potenciais (FUJITA, KRUGMAN; VENABLES, 1999). O conceito de economias de aglomeração é abrangente, pois inclui qualquer efeito que eleva a renda do trabalhador e/ou o lucro da firma à medida que o tamanho das economias locais cresce (COMBES; GOBILLON, 2014).

As discussões sobre economias de aglomeração datam de meados do século XVIII, sendo explicitamente considerada nos modelos de desenvolvimento econômico propostos por Pety, Cantillon, Smith e Marshall. Entretanto, desde então, as formulações teóricas seguintes colocaram, gradualmente, a questão espacial à parte (SMOLKA, 1983). Nesse ínterim, as abordagens econômicas propostas pelos economistas marginalistas tornaram-se hegemônicas e dificultaram uma modelagem econômica considerando a dimensão espacial.

A questão espacial voltou ao debate com os trabalhos desenvolvidos pelos economistas da Escola Histórica Alemã, no século XIX, que atingiu seu ápice na segunda metade do século XIX, mas sem se tornar *mainstream* em Economia. Embasada no materialismo histórico, contestando a validade universal da Economia Clássica e preconizando a validade relativa das leis econômicas, no tempo e no espaço (FERREIRA, 1989), tais avaliações tornaram-se conhecidas como a Teoria da Localização. Muitos pesquisadores (VON THÜNEN, 1826; CHRISTALLER, 1933 *apud* CHRISTALLER, 1966; LÖSH, 1954) abordavam a existência de leis naturais na evolução espacial das estruturas econômicas, construindo listagens exaustivas dos fatores locacionais como indicativos de vantagens e desvantagens de uma região para abrigar atividades industriais, agropecuárias, etc. Segundo Azzoni (1982), a síntese dos modelos de Weber, von Thünen e Lösch, considerando não apenas a localização industrial, mas também o conjunto das atividades econômicas e a formalização de um modelo geral de minimização de custos (transporte, transferência e produção) proposta por Isard (1956), foram absorvidos pelos instrumentos das teorias neoclássicas de produção e preços.

É a partir dos anos 70 do século XX, com a conciliação da Teoria Econômica aos modelos da Teoria da Localização, que a abordagem ganha maior espaço, originando uma vertente conhecida como Nova Economia Urbana (NUE – *New Urban Economics*). Todavia, a relevância da economia de aglomeração e a relevância do espaço se consolidam com o modelo centro-periferia de Krugman (1991), constituindo a Nova Geografia Econômica (NEG – *New Economic Geography*). A NUE formulou uma teoria em que a paisagem urbana deriva do processo decisório de localização das famílias (onde morar e onde trabalhar) e das firmas (minimização da função custo). Já os modelos da Nova Geografia Econômica maximizam o acesso ao mercado potencial das firmas, reduzindo custos de transporte (FUJIGA; KRUGMAN; VENABLES, 1999). Ambas vertentes têm nas economias de aglomeração o efeito principal da escolha locacional das firmas. Todavia, a NEG foca a análise de aglomeração entre cidades, e os múltiplos centros derivam da relação entre setores rurais e manufaturados (FUJIGA; KRUGMAN; VENABLES, 1999; COMBES; DURANTON; GOBILLON, 2003), enquanto a NUE foca em modelos de equilíbrio geral, em que as famílias também escolhem onde morar, e permite múltiplos SBD no espaço intra-cidade (ALONSO, 1964; BECKMAN, 1974; MUTH, 1969, 1975; MILLS, 1967 E 1972).

Marshall (1980) aponta diferentes determinantes da aglomeração, tais como transbordamentos tecnológicos, agrupamento da mão de obra e encadeamento de insumos intermediários. Recentemente, Duranton e Puga (2004) distinguem os três microfundamentos da economia de aglomeração que influenciam a produtividade da economia local, a saber *sharing*, *matching* e *learning*. O efeito de *sharing* deriva dos ganhos gerados pela variedade de insumos e especialização industrial; efeito de *matching* corresponde tanto ao melhoramento quanto à qualidade ou quantidade dos *matches* entre firmas e trabalhadores; o efeito de *learning* diz respeito à geração, difusão e acumulação de conhecimento, os quais decorrem do transbordamento de conhecimento entre os trabalhadores quando as firmas se aglomeram. Assim, o espaço urbano é o local espacial que permite o efeito da aglomeração acontecer e ser distribuído entre os agentes econômicos. Complementarmente, Rosenthal e Strange (2004) consideram elementos adicionais que geram aglomeração, tais como vantagens locais, efeitos no mercado residencial, externalidades de consumo e *rent seeking*.

Segundo Gleaser (2016), a cidade é a maior invenção da história humana e sua existência perdura devido à sua capacidade de promover a mais eficiente organização da atividade econômica. Tal criação deriva de forças aglomerativas, não homogeneidade espacial, por exemplo, como é o caso de cidades criadas a partir da exploração de bens naturais (BELIANT; KONISHI, 2000). Assim, a aglomeração se origina de fatores que tornam a hipótese de distribuição uniforme das atividades no espaço instáveis. Segundo Papageourgiou e Smith (1983), não homogeneidade espacial, economias de escala na produção e economias de escala externas às firmas (incluindo contato espacial e competição imperfeita), sob baixo custo de transporte e comunicação, tornam o processo aglomerativo factível no espaço.

Inicialmente, nos modelos canônicos desenvolvidos pela Nova Economia Urbana (ALONSO, 1964; BECKMAN, 1974; MUTH, 1969 E 1975; MILLS, 1967, 1972; SOLOW, 1973; WHEATON, 1974), as economias de aglomeração derivadas da concentração de emprego aconteciam essencialmente em um único centro, isto é, no centro histórico da cidade (CBD – *Central Business District*). A cidade monocêntrica surge da interação entre firmas (consomem terra para produção e produzem bens homogêneos) e famílias (consomem terra para moradia e se deparam com custo de *commuting*). A firma pode produzir em qualquer lugar da cidade, todavia a produção deve ser conduzida ao CBD e de lá exportada, ou ali mesmo consumida. Tais deslocamentos implicam em custo de transporte (constante por unidade de distância). Mills (1972) mostra que existem duas soluções eficientes: a) toda área ao redor do CBD é utilizada para produção e b) uso misto do solo entre firmas e famílias. A solução segregada é válida quando o custo de transporte de bens supera o custo de transporte casa-trabalho dos trabalhadores (*commuting* ou deslocamento pendular), uma vez que a produção no CBD minimiza o custo da produção. A solução integrada vale quando o custo de *commuting* é relativamente mais elevado que o custo de transporte. Resumidamente, o pressuposto de cidade monocêntrica está embasado no contexto histórico de que o centro da cidade tendia a se desenvolver próximo ao nó de transporte ou de consumo, por exemplo. Tal estilização de empregos concentrados numa área central está fundamentada no prêmio gerado pelas economias de aglomeração (ANAS; ARNOTT; SMALL, 1999).

Esta hipótese é a parte central da avaliação do gradiente de preço da terra e salários. Quanto mais próximo do CBD, mais elevados são os preços das moradias, maior o salário ofertado e maior a

densidade populacional. Assim, o distanciamento do CBD pode ser visto, de forma dual, como a ampliação do custo de transporte e de *commuting*. O ponto relevante é que o preço da terra seria compensado pelo distanciamento do lugar central da cidade, buscando equilibrar todos os agentes econômicos desta economia. Ou seja, o preço da terra decresce em função do distanciamento do CBD, enquanto o custo de transporte se amplia. Nesse caso, os padrões espaciais para as outras variáveis de interesse (densidade populacional, tamanho dos lotes, etc.) são todos estimados como função da âncora desse modelo, o CBD.

Romanos (1979) aponta para a distorção dos resultados derivados de modelos monocêntricos. Segundo ele, os resultados sobre o vetor de preços (terra e salários) são subestimados em áreas onde existem subcentralidades, isto é, existência de pelo menos um centro adicional. Empiricamente, muitas cidades têm mais de um centro de negócios e/ou múltiplos subcentros. A relevância dos modelos com múltiplos centros é justificada por maior aderência às grandes cidades, onde a atividade de trabalho tende a ser descentralizada, com múltiplas concentrações, caracterizando uso misto do solo, como proposto por Wheaton (2004). Ademais, modelos com múltiplos centros podem justificar gradientes de preços não decrescentes monotonicamente a partir da área central em direção às franjas da cidade.

Essa nova distribuição espacial dos empregos altera a estrutura econômica das cidades, tais como a demanda por transporte, oferta de salários e preço da terra no espaço intra-urbano. Entretanto, de acordo com Garreau (1991), mesmo que os novos centros de emprego (SBD) alterem a paisagem da cidade, isso não significa que os SBD se apresentam como réplicas das características dos CBD e que sejam independentes das atividades centrais do polo intra-urbano. Todavia, a importância relativa CBD e SBD pode ser revertida em algum momento no tempo, dado que nada garante taxas de crescimento superiores no CBD relativamente às taxas dos SBD.

Concomitante aos fatores que geram surgimento de novas centralidades, o retorno derivado das economias de aglomeração pode ser sobrepujado por externalidades negativas (custo de congestionamento, alto preço da terra, etc.) no CBD. Elevação nos custos é o fator inicial do processo de descentralização das firmas/empregos. Segundo Giuliano, *et al.* (2005), as firmas cujo lucro derivam das economias de aglomeração tenderão a formar novos *clusters* de emprego, enquanto as firmas cujos lucros dependem pouco de tal fator tenderão a se localizar de forma mais

dispersa. Assim, a dispersão ou concentração fora do CBD depende do papel que a aglomeração desempenha na função de produção da firma. Em outras palavras, forças de aglomeração fracas implicam em uma descentralização dispersa, enquanto forças de aglomeração fortes tendem a concentrar atividades, caracterizando subcentros (SBD).

Giuliano *et al.* (2005) dividem em dois corpos teóricos os trabalhos que apontam para os fatores que levam à aglomeração ou desaglomeração do emprego no espaço intra-urbano. O primeiro grupo relaciona essas forças como determinantes para a criação de subcentralidades no espaço intra-urbano, embasado em microfundamentos. Esse primeiro grupo é conhecido na literatura como Nova Economia Urbana. O segundo grupo de trabalhos aponta outras forças favoráveis e desfavoráveis à concentração das atividades econômicas e definem o corpo teórico denominado de Nova Economia (NE). Diferentemente dos economistas da NEU, os trabalhos agrupados na NE não se utilizam de microfundamentos. Contudo, o ponto nevrálgico para melhor compreender o surgimento de novas centralidades urbanas está em agregar os fatores apontados por ambos grupos.

O problema locacional da firma em uma cidade com múltiplos centros passa pelo *trade-off* em se localizar no CBD ou no SBD. O surgimento desses subcentros está ligado aos benefícios derivados do espraiamento das firmas, apesar das vantagens oferecidas pela concentração em um único centro. Assim, em face dos resultados empíricos, modelos foram desenvolvidos considerando cidades policêntricas no fim da década de 70 do século XX. Segundo Anas, Arnott e Small (1999), os trabalhos de White (1976) e Dixit e Stiglitz (1976) são os seminais. Entretanto, Sheamur e Coffey (2002) indicam que os trabalhos mais relevantes foram os da década de 1980, tais como Fujita e Ogowa (1982), Greene (1980), Griffith (1981) e Getis (1983). O primeiro é o mais referenciado e utilizado como base para extensões, tal como em Henderson e Slade (1993) e Lucas e Rossi-Hansberg (2002).

A literatura teórica sobre cidades policêntricas pode ser dividida em duas categorias principais, como proposto por White (1999). A primeira categoria lida com CBD e SBD como um problema endógeno às economias de aglomeração e custos de transportes em uma abordagem de equilíbrio geral (FUJITA; OGAWA, 1982; FUJITA, 1988; HELSLEY; SULLIVAN, 1991; HENDERSON; SLADE, 1993; ANAS; KIM, 1996; LUCAS; ROSSI-HASNBERG, 2002; AHLFELDT *et al.*, 2016). A tratabilidade dos modelos policêntricos endógenos está embasada na hipótese de que a produção e o uso residencial do solo podem ocorrer em qualquer localização intra-urbana, mas é

dependente das viagens com intenção de consumo feitas pelos consumidores, de economias de aglomeração, de congestionamento, de considerações estratégicas, dos encadeamentos produtivos entre as firmas e dos custos de transporte de bens e de *commuting* (ANAS; KIM, 1996). Esses custos tornam a oferta de emprego e de mão de obra endógenas devido à escolha locacional das famílias, de onde residir, e das firmas, de onde produzir, simultaneamente. Todavia, os fatores básicos que causam a agregação de firmas no CBD ou SBD são economias de aglomeração externas e custos de transporte de bens relativamente maiores que o custo de *commuting* (WHITE, 1999).

Fujita e Ogawa (1982) usam pressupostos similares aos de Mills (1972), adicionando economias de aglomeração ao modelo. Para esses autores, a economia de aglomeração deriva da presença de externalidades não pecuniárias na produção decorrentes da proximidade entre firmas. Diferentes centralidades surgem devido ao *trade-off* entre custo de *commuting* e transporte dos bens e economias de aglomeração. Assim, se economias de aglomeração estão mais concentradas no CBD e decaem com o distanciamento, então as firmas se concentram no CBD e este é cercado por moradias. Todavia, se economias de aglomeração são constantes no espaço e o custo de *commuting* é relativamente mais elevado que o custo de transporte, tem-se um padrão disperso e os trabalhadores trabalham em casa. É possível que haja um padrão misto incompleto, isto é, famílias que trabalham em casa morando no CBD e famílias que se deslocam para trabalhar no CBD. Adicionalmente, pode existir apenas SBD ou SBD e CBD. Fujita (1988) mostra que múltiplos centros surgem em um contexto com interações entre firmas e consumidores, gerando ganhos puramente pecuniários sob concorrência monopolística.

Henderson e Slade (1993) estendem o modelo de Fujita e Ogawa (1982), propondo um jogo entre dois planejadores centrais que desenvolvem vizinhanças para habitação e centro de negócios. O jogo visa demonstrar a ineficiência derivada de modelos sem economias de aglomeração, pois a firma individual ignora os efeitos de seu comportamento no nível da aglomeração e sobre o custo das outras firmas, culminando na divisão do CBD antes mesmo que a cidade de cada um dos planejadores tenha suporte para tal, pois considera apenas a economia de aglomeração realizada pelas firmas em sua parte da cidade. Helsley e Sullivan (1991) propõem uma abordagem em que o planejador central da cidade aloca o crescimento populacional em duas regiões contíguas (CBD e SBD) de forma a maximizar o valor agregado social do produto. Todavia, infraestrutura pública

é requerida *a priori* para que uma localidade de produção possa ser desenvolvida. Outros ingredientes relevantes desse modelo são economias de aglomeração e deseconomias de transporte. Assim, subcentralidades derivam do *trade-off* entre economias de escala externas e custo de transporte, e podem ser afetados pelo custo fixo de infraestrutura pública, diferenças tecnológicas e interações entre localizações de produção.

No modelo desenvolvido por Anas e Kim (1996) não existem economias de aglomeração, mas tomam por pressuposto que as firmas vendem diretamente às famílias que vivem na área urbana, o que exige uma interação de vendas entre firmas e famílias. Tais interações são similares às que ocorrem entre firmas, que são a base dos modelos de economias de aglomeração segundo White (1999). A descentralização do emprego (firmas) deriva da heterogeneidade espacial dos bens e a demanda é afetada por custo de congestionamento, o que faz os trabalhadores consumirem em locais mais próximos.

Lucas e Rossi-Hansberg (2002), embasados no modelo de Fujita e Ogawa (1982), relaxam algumas hipóteses, tais como substituição de tecnologias de produção (trabalho e terra), o que culmina em diferentes aglomerações de trabalhadores espacialmente. No modelo, o equilíbrio espacial das firmas deriva dos benefícios gerados pelas economias externas dadas pela proximidade entre firmas e do custo de *commuting* dos trabalhadores. O equilíbrio do modelo permite uma cidade multinuclear, onde o emprego pode estar localizado em qualquer área da cidade. O modelo assume que toda a produção ocorre dentro da cidade, devido às economias de aglomeração, e que a produtividade é tanto mais elevada quanto maior for a concentração de emprego que circunscreve o trabalhador e/ou a firma. Segundo os autores, na presença de economias externas, os preços de mercado não são suficientes para fornecer incentivos locacionais eficientes para trabalhadores e firmas. Uma limitação desse modelo é que o vetor de preços é uma função exclusiva da âncora de preços, isto é, da distância ao CBD.

A segunda vertente dos modelos teóricos da NUE assume exogenamente o SBD (quantidade de subcentros, por exemplo) e o CBD. Nessa abordagem, a existência de um ou vários subcentros resulta da descentralização da atividade econômica e é tomada como dada. Fazem parte desse grupo Papageorgiou (1971), White (1976; 1990), Hartwick e Hartwick (1974), White (1976; 1988), Romanos (1979), Sullivan (1986), Wieand (1987), Sivitanidou e Wheaton (1992), Hotchkiss e White (1993), Yinger (1992), Ross e Yienger (1995) Henderson e Mitra (1996), Zhang e Komei

(1997; 2000) e Wrede (2015). Nesse grupo de modelos o interesse especial é avaliar o efeito da suburbanização do emprego sobre as residências e outros aspectos. Dito de outra forma, o foco desses modelos está em avaliar como as famílias/trabalhadores decidem onde morar e trabalhar e no padrão espacial resultante do preço da terra, da densidade populacional e do deslocamento pendular do trabalhador, dada a localização da firma. Os resultados indicam o decaimento da renda da terra e dos salários em função do distanciamento dos CBD e SBD, por exemplo.

White (1999) considera esses modelos, cujo padrão policêntrico é definido de forma exógena, como sendo um caso particular dos modelos endógenos. Ademais, devido à dificuldade resolutive desses modelos, a primeira vertente dos modelos policêntricos, que levam em consideração o efeito da aglomeração e do custo de transporte apresenta maior dificuldade em derivar as avaliações sobre renda da terra e salários.

Os autores da segunda vertente apontam para diferentes fatores que fomentam o espriamento do emprego para fora do centro principal (CBD), tais como políticas de imposto e de uso do solo (SULLIVAN, 1986; ZHANG; KOMEI, 1997, 2000). Henderson e Mitra (1996) sublinham o papel de agentes privados e públicos, os quais contribuem para o processo de descentralização do emprego e facilitam a migração das firmas. Os agentes privados escolhem considerando a capacidade que os distritos de negócios e a localização têm de maximizar o lucro das firmas. Tais escolhas dependem da capacidade histórica de oferta de emprego no SBD e, implicitamente, da população residencial. A questão principal trazida é que todos os agentes privados estão procurando pela melhor localização, e então, neste ambiente, há centenas de agentes agindo da mesma forma. Logo, políticas públicas desenvolvidas pelos agentes públicos atuariam, via políticas de regulação do uso do solo e subsídios, incentivando a localização ou realocação numa área específica no espaço intra-urbano.

Resumidamente, economias de aglomeração, retornos internos e externos às economias de escala da produção e deseconomias de transporte (congestionamento) são fatores abordados pela NEU para explicar a criação inicial do CBD e podem também influenciar o surgimento de subcentralidades. Firms concentradas no CBD se apropriam de economias de aglomeração e, no início da formação do CBD, não incorrem em fricções como baixo *pool* de mão de obra e dificuldade de deslocamento, as quais poderiam desencorajar as firmas de se localizarem no CBD

Entretanto, assim que o CBD alcança determinado tamanho que ative externalidades negativas (custo de congestionamento, elevado preço da terra, etc.) e que estes efeitos sobreponham as externalidades positivas da aglomeração, o processo de deslocamento do CBD para outra região se inicia.

Em áreas periféricas da cidade, menor preço da terra, acesso às rodovias intermunicipais e interestaduais induzem a mudança para essas localidades. A primeira reduz o custo fixo com aluguel e a acessibilidade não altera o acesso à oferta de trabalho, desde que o custo não seja elevado (MCMILLEN, 2001b). Embora a diversidade de atividades seja menor nos subúrbios, SBD maiores tendem a ter menores distâncias em relação ao CBD, desde que haja provisão de um sistema de transporte a custo baixo e que o preço da terra seja baixo próximo ao CBD. Outra vantagem associada ao tamanho de subcentralidades é que SBD grandes tendem a ofertar oportunidades de emprego e compras às famílias e, conseqüentemente, implicam na redução da dependência do CBD (MCMILLEN, 2001b).

Os trabalhos alinhados à Nova Economia consideram tecnologias de informação e comunicação (ICT³), desenvolvimento histórico, mercado de trabalho e utilidade dos trabalhadores e firmas como fatores capazes de concentrar e dissipar empregos. Kloosmeterman e Musterd (2001) argumentam que ICT contribui para a dispersão do emprego por atuar reduzindo o custo da transmissão e processamento da informação. Assim, os fluxos físicos podem ser substituídos por fluxos virtuais, reduzindo o custo do deslocamento pendular ou, de forma dual, o custo da proximidade é reduzido. Portanto, a partir da redução do custo do transporte e comunicação, as firmas podem arbitrar em termos locacionais no espaço intra-urbano e explorar vantagens comparativas que derivam de diferentes partes da cidade.

Entretanto, Castells, Hall e Hall (1994) e Graham e Marvin (1996) apontam para a complexidade das atividades baseadas em conhecimento e o importante papel da complexidade informacional que demandam relacionamentos face a face. Sob esse ponto de vista, os autores acreditam que, mesmo num cenário em que a ICT reduza custos, a complexidade das atividades baseadas no conhecimento atua impondo atrito ao espriamento dos postos de trabalho.

³ Sigla derivada do inglês para “Information and Communications Technologies”.

O desenvolvimento histórico da cidade é outro fator que pode explicar a concentração dos empregos no CBD e o surgimento de novos polos de concentração de emprego. Graham e Marvin (1996) apontam que cidades grandes reproduzem um processo auto reforçador de concentração dos postos de trabalho. A ideia por trás de sua hipótese é que as grandes cidades dispõem de oferta diversificada de mão de obra, alta qualificação e maior número de trabalhadores porque tais cidades fornecem sistema de transporte bem estruturado e tecnologias de informação e comunicação, e essas provisões são viáveis devido à maior oferta de trabalhadores do que em qualquer outra cidade.

Guiliano e Redfearn (2005) discutem a preferência das famílias como outro fator que leva à concentração do emprego. Como a oferta de serviços educacionais e culturais é maior nas grandes cidades do que nas cidades pequenas, os trabalhadores são atraídos por essas amenidades e pelo modo de vida da cidade também. Em seguida, as empresas assumem a utilidade dos trabalhadores e são estimuladas a se localizarem nessas grandes cidades para terem acesso aos trabalhadores altamente qualificados. No entanto, por outro lado, alguns trabalhadores preferem ambientes de vida de baixa densidade, estimulando a dispersão. Assim, com as ICT oferecendo mobilidade aos trabalhadores, as preferências das famílias tornam-se mais viáveis e os trabalhadores podem escolher a localização da residência em outro lugar, maximizando sua função de utilidade. Nessa função de utilidade tem-se como argumentos compras, escola das crianças e lazer (KLOOSMETERMAN; MUSTERD, 2001). Portanto, na medida em que a qualidade de vida é importante para os trabalhadores, as empresas internalizam a preferência dos trabalhadores na escolha de sua própria localização espacial, resultando em uma suburbanização do trabalho (GOTTLIEB, 1995; HALSTEAD; DELLER; 1997, BEYERS, 2000).

Resumidamente, os modelos monocêntricos não são compatíveis com as grandes cidades, as quais têm crescido tanto em termos populacionais quanto em densidade de emprego. Empiricamente, tem-se observado o espraiamento dos empregos em direção às periferias. O dinamismo econômico nas grandes cidades impacta diretamente sobre o efeito da aglomeração (economias de escala). Os efeitos positivos derivados da aglomeração são auto reforçados pelas características inerentes às grandes cidades e estimulam cada vez mais a aglomeração das firmas no espaço. Todavia, a partir do instante em que efeitos positivos da aglomeração crescem a taxas menores do que a taxa de crescimento do custo da aglomeração, efeitos centrífugos tendem a expulsar as firmas de tal

aglomeração, espalhando-as pela área intra-urbana ou levando a uma nova localização aglomerada fora do CBD. Como discutido, essa descentralização e a nova formatação espacial dos empregos refletem diretamente sobre o vetor de preços da economia e é dependente da relação das economias de aglomeração e função lucro da firma.

3.1.2. Abordagens Empíricas de Identificação de Subcentralidades

Os modelos teóricos citados acima têm focado essencialmente no equilíbrio da configuração espacial das cidades policêntricas, sem preocupação de gerar resultados empiricamente testáveis. No final dos anos 1980, vários pesquisadores lançaram esforços para identificar e avaliar a descentralização espacial do emprego na área intra-urbana (GORDON; RICHARDSON; WONG 1986; MCDONALD, 1987; CERVERO, 1989).

A literatura empírica especializada aponta para a descentralização dos postos de emprego para fora do CBD. Embora essa tendência possa justificar o papel decrescente do CBD na economia intra-urbana, bem como a perda relativa da participação de emprego no CBD ao longo do tempo (MCMILLEN, 2001b), isso não significa que o CBD tenha sido desmantelado. Resultados empíricos sustentam que a função do CBD tem se mantido e que os SBD têm se desenvolvido com base na especialização de diferentes atividades (SHEARMUR; COFFEY, 2002), caracterizando uma área suburbanizada.

Embora os trabalhos empíricos tenham buscado identificar subcentros, não existe uma estratégia empírica consolidada na literatura. Inicialmente, os estudos se valiam do conhecimento prévio da região em que se estudava (BENDER; HWANG, 1985; HEIKKILA; GORDON; KIM, 1989; RICHARDSON *et al.*, 1990). Segundo McMillen (2001b), essa escolha tende a não ser consistente com os dados observados.

Considerando apenas os procedimentos empíricos embasados em métodos estatísticos objetivos, pode-se dividi-los entre aqueles que consideram:

1) *Picos ou Valor de Corte*: propõe-se um valor de corte para identificar áreas centrais, isto é, são escolhidas as zonas que têm densidade de emprego ou populacional acima de certo *cut-off* pré-estabelecido. Fazem parte desse grupo os trabalhos de Greene (1980), McDonald (1987), McDonald e McMillen (1990), Giuliano e Small (1991, 1999), Small e Song (1994), Cervero e Wu

(1997), Bogart e Ferry (1999), Coffey e Shearmur, (2001), Anderson e Bogart, (2001), Shearmur e Coffey (2002), Boumont e Bourdon, (2002), Muniz, Galindo, García, (2003), López e Muñiz (2005), Pan e Ma (2006).

2) *Modelos Econométricos*: nesse grupo, os autores estimam equações de densidade de emprego e/ou população considerando abordagens metodológicas de econometria paramétrica e/ou não-paramétrica. (2.1) *Métodos Paramétricos*: Heikkila, Gordon e Kim (1989), McDonald e Prather (1994) e, (2.2) *Métodos Não-Paramétricos*: McMillen e McDonald (1997), McMillen e McDonald (1998), Craig e Ng (2001), McMillen (2001a, 2001b), McMillen e Smith (2003), Lopez (2006) e Redfearn (2007).

3) *Estatística Espacial*: os trabalhos desse grupo aplicam técnicas de análise exploratória espacial aos dados, tal como análise de *cluster* utilizando *I* de Moran: Modarres (2003), Baumont, Ertur e Gallo (2004), Guillain, Gallo e Boiteux-Oraian (2003), Ramos (2004) e Kneib (2008)

4) *Estratégia Híbrida*: os trabalhos desse grupo agregam diferentes metodologias para identificar os subcentros, objetivando acuracidade nas identificações. (4.1) *Modelos Não-parâmetros e Valor de Corte*: McMillen (2003) e Kane, Hipp e Kim (2016). (4.2) *Estatística Espacial e Valor de Corte*: Siqueira (2014). (4.3) *Estatística Espacial e Econometria Espacial*: Campos e Chagas (2017).

5) *Fluxos*: (5.1) *Método de grafos*: Bourne (1989) utilizando grafos avalia, visualmente, as relações de deslocamento; (5.2) *Método de Saldo Líquido*: baseia-se na avaliação do fluxo líquido de entrada e saída em determinada região para o motivo trabalho (BURNS; BOIX; ROCA, 2001); (5.3) *Cut-off de Fluxo*: identifica-se a zona central de negócios a partir da densidade de geração de viagem acima de determinado valor de corte (CLARK; KUIJPERS-LINDE, 1994; GORDON; RICHARSON, 1996); (5.4) *Atração de Fluxo*: Aguilera e Mignot (2004) utilizam os dados sobre pendularidade casa-trabalho para identificar áreas centrais.

O método de *valores de corte* mais discutido na literatura foi proposto por McDonald (1987) e estendido por Giuliano e Small (1991), embora Greene (1980) tenha apresentado uma abordagem de corte previamente. Segundo Greene (1980), um subcentro é considerado quando o desvio da densidade do emprego na área é duas vezes maior que a média da região metropolitana. Sobre a

abordagem discutida em McDonalds (1987), a identificação da subcentralidades decorre da concentração do emprego nas áreas adjacentes à analisada. Assim, o autor aplicou o método para a cidade de Chicago e identificou 4 subcentros, agregados em 44 zonas. A abordagem de Giuliano e Small (1991) busca ser mais objetiva. Para tanto, identificam como subcentro um conjunto de áreas contíguas que têm uma densidade mínima de 10 empregados por acre (4040 m²) e apresenta, conjuntamente, 10 mil empregados. Os autores aplicaram o método para a região de Los Angeles e identificaram 32 subcentros, utilizando o Censo de Origem-Destino de 1980⁴.

As abordagens de valor de corte são criticadas devido à discricionariedade do corte escolhido, à sensibilidade dos resultados à escolha de corte (MCMILLEN, 2001a; MCMILLEN, 2003) e à impossibilidade de generalização para qualquer cidade (ANAS; ARNOTT; SMALL, 1999; MCMILLEN, 2001b; BAUMONT; ERTUR; GALLO, 2004). Os resultados também são sensíveis ao tamanho da área quando a densidade do emprego é considerada como variável dependente (MCMILLEN, 2001a, 2001b). Portanto, sob tal abordagem empírica, a identificação está sujeita ao conhecimento local e às expectativas *a priori* para determinar o valor de corte previamente (MCMILLEN, 2003).

As abordagens econométricas geralmente estimam uma função simples de densidade do emprego para uma cidade monocêntrica $D(x) = D_0 e^{-\beta x + u}$, onde D é a densidade do emprego ou da população, x é a distância ao CBD, D_0 é a densidade do emprego ou da população média na distância zero, β é o gradiente de densidade, u é o termo de erro aleatório. McDonald (1987) e Heikkila, Gordon e Kim (1989) podem ser considerados os pioneiros em trabalhos paramétricos. McDonald (1987) estima uma função de densidade do emprego para uma cidade monocêntrica e considera os termos de erro estimados positivos e estatisticamente significativos para identificar a subcentralidades de emprego. Heikkila, Gordon e Kim (1989) consideram o valor da terra como a variável dependente e a distância aos candidatos a subcentro como regressores. Tais estratégias empíricas são um avanço em relação às abordagens anteriores, pois utilizam inferência estatística e buscam acomodar a abordagem aos modelos teóricos. Todavia, a falta de utilização de controles pode ter enviesado os resultados.

⁴ Para outros artigos relacionados à abordagem de valor de corte, ver Apêndice A.

Embasados em Gordon, Richardson e Wong (1986), Heikkila, Gordon e Kim (1989), McDonald e Prather (1994) propuseram identificar subcentralidades considerando os resíduos estimado derivados de regressões econométricas. Essa abordagem é limitada, pois os autores estimam densidade de emprego sem controles, não garantindo que as distribuições espaciais dos resíduos estimados pela especificação econométrica citada acima identificam subcentralidades. O método proposto por Craig e Ng (2001) propõe utilizar os resíduos estimados de uma regressão com especificação logarítmica da densidade do emprego. Para tanto, os autores utilizaram uma abordagem de regressão quantílica com foco no percentil 95 da distribuição de densidade de emprego. McMillen (2001b) aponta que a abordagem não paramétrica proposta pelos autores não é sensível à unidade de análise, como as que consideram a área. Todavia, McMillen (2001b) critica a estimação de uma função de densidade de população e/ou de emprego simétrica em torno do CBD, pois pode gerar identificação enviesada caso não haja simetria, isto é, podem existir diferentes gradientes em direção a leste e a oeste, por exemplo. Outra limitação, Craig e Ng (2001) consideram apenas os círculos concêntricos da cidade cuja alta densidade de emprego fora encontrada, embasados em conhecimento prévio da cidade. Nesse sentido, a abordagem é um retrocesso em termos metodológicos. Redfearn (2007) aponta para a sensibilidade dos resultados, dado que são pautados em valores de cortes, e para a ausência de inferência estatística.

McMillen (2001a, 2001b) propôs um modelo de dois estágios que dispensa o conhecimento prévio da cidade avaliada. O primeiro estágio é próximo ao proposto por McDonald (1987) e Craig e Ng (2001), isto é, estima-se o logaritmo da densidade de emprego contra a localização deste em relação ao CBD, todavia utilizando Regressões Ponderadas Localmente (*LWR – Local Weighted Regressions*). Os candidatos a SBD são aqueles cujos resíduos estimados formam *clusters* estatisticamente significativo a 5%. No segundo estágio, após considerar os potenciais subcentros, é utilizado um modelo semiparamétrico. A parte não-paramétrica da regressão ajustada por uma aproximação flexível de Fourier utiliza os resíduos do primeiro estágio como variável dependente e a distância ao CBD como variável explicativa. Os coeficientes estimados que apresentaram gradientes negativos com o distanciamento do CBD e estatisticamente significativos são identificados como SBD.

Redfearn (2007) também considera modelos não-paramétricos em suas estimações, que diferem das abordagens propostas por McMillen (2001a, 2001b). O autor propõe uma estimacão que considera a superfície da densidade de emprego localmente. Nessa abordagem, os candidatos a subcentro são áreas com máximo local, dado que as estimacões econométricas são locais. Empiricamente, o autor usa Regressões Ponderadas Localmente para estimar uma vizinhança de densidade de emprego ao redor do candidato a subcentralidade. Essas estimacões são divididas em três etapas, considerando a densidade de emprego: 1) estimacão dos picos globais; 2) estimacão dos picos relativos aos vizinhos e 3) *bootstrap* dos picos relativos para fazer inferência.

McMillen (2003) propõe um método híbrido para identificação de subcentros, combinando as abordagens desenvolvidas em Giuliano e Small (1991) e McMillen (2001a). Esta estratégia fornece uma solução potencial para o procedimento desenvolvido em McMillen (2001a). A estratégia empírica utiliza um método não-paramétrico para estimar Regressões Ponderadas Localmente. Os subcentros são áreas contíguas com resíduos positivos e com nível de significância estatística de 10 por cento, e que apresentem pelo menos 10.000 funcionários. Embora o método híbrido traga um dos dois elementos arbitrários de Giuliano e Small (1991), McMillen (2003) afirma que esse limiar é menos arbitrário do que a densidade mínima de emprego (segundo elemento arbitrário). McMillen e Smith (2003) utilizam a mesma abordagem não-paramétrica, sem considerar o valor de corte arbitrário no segundo estágio. Os autores propõem o valor corte derivado da distribuição do erro estimado, considerando áreas que compõem a distribuição menor ou igual a 5% dessa distribuição.

Também seguindo uma abordagem não-paramétrica em dois estágios, Kane, Hipp e Kim (2016) geocodificaram as firmas utilizando os centroides dos CEP em Los Angeles, EUA. Os autores criaram *grids* de 1 km por 1 km cobrindo a área urbana da região. O processo de identificação se dá em dois passos. No primeiro, seguindo McMillen (2001a), estimam regressões locais (LWR) para criar a superfícies de emprego predito para cada célula. No segundo passo, os autores consideram as células ao redor dos máximos locais que fazem fronteira com cada cluster de emprego utilizando matrizes de ponderação espacial, como sugerido em McMillen (2003).

Neste subgrupo de abordagens empíricas há um expressivo esforço em formular critérios objetivos de identificação. Todavia, considerar os resíduos estimados pode levar a resultados enviesados, pois nem todo o desvio para cima dos erros estimados identificam subcentralidades. Regras de

corde mais objetivas podem ser verificadas, mas é aplicada aos resíduos estimados. Ademais, regras de corte *ad hoc* também se fazem presente neste subgrupo.

Considerando trabalhos cuja abordagem empírica é estatística espacial, as principais técnicas aplicadas são *Kringing*⁵ e Análise Exploratória de Dados Espaciais (AEDE). Modarres (2003) usa *Kringing* para identificar subcentralidades em Los Angeles e aponta para a possibilidade de distinguir entre centros maiores dentre os subcentros. No processo de interpolação, o autor considera os centroides dos distritos de Los Angeles (EUA) e então calcula a proximidade espacial utilizando a técnica *nearest neighbor hierarchical* (NNH), agrupando o total de emprego associado a cada centroide no espaço. Embora a abordagem de *Kringing* forneça resultados hierarquizados, esta requer valores de *cut-off* para a definição de subcentralidades *a priori*. Tal abordagem assume estacionariedade espacial do processo gerador dos dados, isto é, dados igualmente distribuídos no espaço. Logo, se a base de dados que utiliza essa abordagem não observa tal hipótese, os resultados ficam prejudicados. Como a identificação das subcentralidades se dá via concentração de emprego não isotrópicas, a hipótese não é observada.

Baumont, Ertur e Gallo (2004), Guillin, Gallo, Boiteux-Orain (2006), Kneib (2008) e Siqueira (2016) usam a abordagem do tipo AEDE para identificar subcentralidades de emprego. Enquanto os dois primeiros aplicaram o método para a aglomeração de Dijon e região Ile-de-France na França, Kneib (2008) concentra-se na cidade de Manaus e Siqueira (2016) estudou a área metropolitana de São Paulo.

Baumont, Ertur e Gallo (2004) propõem uma abordagem em dois estágios utilizando estatística e econometria espacial. Os autores consideram como variáveis dependentes o nível e a densidade do emprego. O primeiro estágio consiste em estimar o *I* de Moral Local, considerando os *clusters* alto-alto como candidatos à subcentralidade. No segundo estágio, utilizam-se os candidatos a SBD para estimar um modelo econométrico considerando a função de crescimento populacional, como demonstrado acima. Consideram subcentros as áreas cujos coeficientes estimados sejam negativos e estatisticamente significativos. Guillin, Gallo e Boiteux-Orain (2006) aplicam o primeiro estágio

⁵ É também conhecido como Processo Gaussiano de Regressão. O método utilizado em geoestatística consiste em interpolar dados, criando uma superfície constante contínua para os dados (Modarres, 2003).

apontado em Baumont, Ertur e Gallo (2004) utilizando como variável dependente a relação total de trabalhador por total de população (taxa trabalhador-população) por área de ponderação do censo francês. Nessa abordagem empírica, a identificação está restrita à correlação espacial da taxa trabalhador-população e seus os vizinhos.

Destacando os trabalhos desenvolvidos na literatura nacional, os artigos acadêmicos têm despendido pouca atenção à questão do espraiamento urbano quando comparados à literatura internacional. Parte disso está ligado à ausência de base de dados em nível tão detalhado. Ademais, a maior parte das pesquisas com foco na área intra-urbana é conduzida por pesquisadores de Planejamento Urbano e pouco conversa com o debate empírico internacional e com o campo de Economia Urbana. Ramos (2004) é o primeiro artigo empírico que debate a questão de subcentralidade intra-urbana no Brasil utilizando uma abordagem com inferência estatística data de 2004, 24 anos após a publicação do artigo seminal de Green (1980) nos Estados Unidos.

A produção acadêmica relacionada à identificação de subcentralidades no país pode ser dividida em seis categorias. Essa subdivisão apresenta pouca relação com a divisão acima, pois no Brasil muito dos trabalhos foram desenvolvidos fora das Ciências Econômicas. Os cinco primeiros subgrupos seguem a subdivisão proposta por Kneib (2008). No primeiro grupo estão os trabalhos relacionados a pesquisadores que aplicaram variadas técnicas espaciais, o que Kneib (2008) denomina de Síntese Espacial. Esses trabalhos têm como foco a funcionalidade, morfologia do sistema de vias urbanas, a conexão histórica e geométrica do CBD. No segundo grupo estão trabalhos que utilizam informação do uso da terra e do sistema de transporte como as principais variáveis para identificar subcentralidades de emprego, tais como informações sobre o número de rotas de ônibus, criação potencial de viagens, concentração de atividades econômicas, acessibilidade e topografia. O grupo três agrega conceitos sobre fluxo de passageiros, fluxo de chamadas telefônicas, convergências de pessoas e comunicação, considerando apenas o nível relativo desses fluxos na identificação. A quarta categoria condensa trabalhos que usam análises de grafos como abordagem empírica para avaliar a região que mais atrai pessoas, independente do motivo. Na quinta e sexta categorias estão os artigos que utilizam de estatísticas e econometria espacial e a técnica de especialista (ET), respectivamente.

Entre os doze trabalhos citados por Kneib (2008) relacionados às cidades brasileiras, o mais antigo a discutir subcentralidade data de 1971 e utiliza dados sobre o fluxo de passageiros. É importante

destacar que a abordagem é regional e considera os municípios do estado de Minas Gerais. Ramos (2004) apresentou o primeiro trabalho publicado a utilizar inferência estatística, considerando as zonas de tráfego da Região Metropolitana de São Paulo para o ano de 1997.

Considerando o debate internacional quanto às abordagens empíricas de identificação de subcentralidades, não discutiremos os quatro primeiros subgrupos apontados acima por Kneib (2008) devido à aplicação de formas empíricas pouco convencionais na literatura internacional e nacional em Economia Regional e Urbana. Assim, consideram-se os poucos artigos acadêmicos nacionais que utilizam métodos de estatística espacial (RAMOS, 2004; KNEIB, 2008; SIQUEIRA, 2014) e econometria espacial (CAMPOS; CHAGAS, 2017), e, recentemente, trabalhos que utilizam a Técnica de Especialistas tem sido utilizados (Kneib, 2008; Barros et al., 2014 e Kneib, 2016)⁶.

Embora Ramos (2004) faça várias análises utilizando informações georreferenciadas aplicando indicadores espaciais locais, dentre eles a concentração de emprego, sendo o primeiro estudo que demonstra a concentração de emprego de forma mais objetiva e com testes estatísticos no Brasil, o objetivo principal da pesquisa não era identificar subcentralidades, mas discutir como a diferente agregação dos dados e utilização de variável dependente afetam os resultados.

Kneib (2008) e Siqueira (2014) têm por objetivo a identificação de subcentralidades na Região Metropolitana de Manaus e na Região Metropolitana de São Paulo, respectivamente. Em ambos trabalhos são estimados *I* de Moran Local para identificar subcentralidades, considerando o nível de emprego como variável de análise. Kneib (2008) considera subcentrais as áreas cujos *clusters* sejam do tipo alto-alto, isto é, áreas com elevada concentração de emprego circundada por áreas

⁶ Por ser uma abordagem pouco usual em Economia, a discussão se limitará a essa nota de rodapé. A Técnica de Especialistas agrega várias abordagens empíricas, tais como Método do Especialista, Método Delphi e Método de Escala de Pontos. Os Métodos do Especialista e Delphi coletam opiniões de especialistas sobre as características das áreas da cidade, considerando áreas centrais aquelas pontuadas pelos entrevistados. Na abordagem embasada no Método de Escala de Pontos, os pesquisadores formulam uma classificação de importância para cada candidato a subcentro, previamente estabelecido. Assim, as regiões que se destacam entre as principais posições da classificação são consideradas centrais. Essa abordagem foi primeiramente aplicada em Kneib (2008) e seguida por Barros et al. (2014) e Kneib (2016). Tal abordagem empírica é subjetiva quanto aos valores de corte pré-estabelecidos. A necessidade de contratar especialistas para responder questionários impõe mais uma impedância à análise. Assim, para cada cidade, o pesquisador teria que encontrar um especialista local.

de mesma configuração. Siqueira (2014) considera os *clusters* estatisticamente significantes do tipo alto-alto (primeiro estágio) e que apresentem no mínimo 31.500 postos de emprego por área de ponderação e que a relação emprego-população ocupada seja no mínimo igual a 2 (segundo estágio). Para Siqueira (2014), o segundo estágio é uma medida que visa reduzir o risco de falso positivo. Como resultado, o autor identifica 42 áreas como subcentros.

Campos e Chagas (2017) também consideram uma abordagem em dois estágios para identificar subcentralidade no município de São Paulo. O primeiro estágio se aproxima da estratégia de identificação de Boumont, Ertur e Gallo (2004), i.e., considerando como candidatas as regiões que formam *clusters* do tipo alto-alto e alto-baixo e sejam estatisticamente significativas. No segundo estágio os autores estimam um modelo de preços hedônicos usando econometria espacial, aproximando-se da abordagem de Heikkila, Gordon e Kim (1989). Os subcentros são áreas que são correlacionadas negativamente com gradiente de preços dos imóveis.

A abordagem desenvolvida por Siqueira (2014) também aplica valores de cortes para identificação da centralidade e assumidos de forma discricionária. Outra limitação está em considerar *clusters* do tipo alto-alto apenas, podendo haver *clusters* do tipo alto-baixo como uma representação de subcentros em áreas periféricas da região avaliada. Essa mesma limitação está presente em Baumont, Ertur e Gallo (2004). A estratégia empírica desenvolvida por Campos e Chagas (2017) depende de base dados do mercado imobiliário. Nesse sentido, informações adicionais, muitas vezes não disponíveis, são requeridas.

Portanto, assim como o conceito de SBD, a identificação de subcentralidades não apresenta ponto pacífico na literatura empírica, sendo diversas as abordagens utilizadas. Em face dessa questão, busca-se neste trabalho considerar uma combinação de métodos que sejam capazes de responder à questão posta de forma objetiva e generalizável e acomodar a base de dados a ser utilizada.

Outra questão relevante é a variável dependente. Nos artigos empíricos, comumente os pesquisadores utilizam nível da população, densidade populacional (total da população dividido pela área), nível de emprego, densidade de emprego (total de emprego dividido pela área), nível de viagens por motivo de trabalho, densidade de viagens por motivo de trabalho (total de viagens dividido pela área) e preço da terra.

Segundo McDonalds (1989) e McMillen (2001a), os resultados que derivam de modelos de população, considerando total de população e/ou densidade populacional como variáveis dependentes, tendem a apresentar vieses, pois não utilizam controles para as preferências locais das famílias. Tais especificações econométricas assumem como hipótese que as famílias minimizam custo de transporte e têm preferências por viver próximo às áreas centrais. Versões expandidas do modelo de densidade populacional controlam por percentual de idosos, negros e proprietários de imóveis, nível de renda, tamanho da família, valor da terra e acesso à região de compras (MCDONALDS, 1989).

A identificação das subcentralidades usando o nível ou a densidade do emprego também pode ter variáveis omitidas. Nesse caso, a concentração espacial das empresas pode estar relacionada à infra-estrutura oferecida na região, como o acesso às rodovias, aeroportos, mercados de consumo, conforme discutido por Heikkila, Gordon e Kim (1989), e do retorno da aglomeração, conforme discutido nos modelos teóricos. Por esse motivo, os trabalhos empíricos em sua grande maioria buscam identificar áreas centrais ao invés de identificar determinantes da formação de tais áreas. Todavia, estimações econométricas do modelo exponencial de concentração de emprego também podem incorrer em omissão de variáveis relevantes. A estratégia empírica utilizada é considerar médias locais a partir de modelos não-paramétricos, como discutido acima, tomando as áreas centrais como exógenas. Assim, o importante para esse grupo de pesquisadores é entender a posição relativa da concentração média de emprego espacialmente.

Heikkila, Gordon e Kim (1989) e Campos e Chagas (2017) utilizam o preço de imóveis como variável dependente para identificar subcentralidade. A hipótese de identificação é que os subcentros são áreas cujo gradiente de preço é decrescente em função da distância. Nessas especificações, os autores utilizam da abordagem de preços hedônicos, controlando por características dos imóveis e amenidades. A utilização de preço dos imóveis requer muitos controles para evitar falsos positivos na identificação de subcentralidade, advindos de variáveis omitidas. Outro limite imposto por essa abordagem para o Brasil é a dificuldade de acesso a bases de dados de preços de imóveis e amenidades georreferenciadas.

Nas estratégias empíricas que utilizam como variável dependente o sistema de transporte, o fluxo de viagem por motivo trabalho, o nível e/ou densidade de emprego, os autores dispensam os

determinantes da concentração, restringindo a análise à identificação de regiões que recebem maior fluxo de trabalhadores e/ou apresentam maior concentração de emprego. Assim, como nos modelos teóricos exógenos, os subcentros são dados. Esses autores se aproximam dos pesquisadores que utilizam concentração de empregos.

Este trabalho utiliza como variável dependente a concentração de emprego por área de 1 km². A utilização dessa estratégia de identificação exige um esforço adicional em função da base de dados a ser utilizada. As abordagens e estratégia empíricas são apresentadas na próxima subseção.

3.2. METODOLOGIA

Esta subseção apresenta a metodologia adotada para a identificação de subcentralidades neste trabalho. Primeiramente é apresentada a base de dados e o tratamento dado para o georreferenciamento e, em seguida, os ajustes feitos para a estimação do modelo econométrico. As subseções seguintes apresentam a estratégia empírica e os modelos econométricos e métodos utilizados para identificar as questões postas neste capítulo.

3.2.1. Base de Dados

Os dados administrados pelo Ministério do Trabalho sobre o mercado de trabalho são fornecidos pelo Cadastro Geral de Empregados e Desempregados (CAGED), a Relação Anual das Informações Sociais (RAIS) ou a Pesquisa de Emprego e Desemprego (PED) de forma agregadas por município. Para a Região Metropolitana de São Paulo, a cada 10 anos é realizada pela companhia do Metrô uma pesquisa de origem-destino (OD), gerando uma base de dados amostral por área de ponderação. A decisão de trabalhar com a microlocalização das firmas apresenta um passo adiante, por superar a dificuldade de identificar subcentralidades devido à ausência de bases de dados que localizem o emprego numa escala mais fina, bem como supera a limitação trazida por normalização do total de empregos pela extensão de área (densidade bruta), como discutido acima. Para tanto, parte do esforço está em geocodificar a RAIS.

A RAIS é um relatório de informações socioeconômicas solicitado compulsoriamente pelo Ministério do Trabalho (MT) aos estabelecimentos formais⁷ (públicos e privados) no âmbito nacional. As firmas devem preencher o relatório anualmente⁸, e cabe ao Ministério do Trabalho a

⁷ Foi instituída pelo Decreto nº 76.900, de 23 de dezembro de 1975 (MT, 2017).

⁸ As declarações são prestadas entre janeiro e fevereiro, e referem-se ao ano anterior (MT, 2017).

responsabilidade de gestão das informações (acompanhamento, controle e disseminação). O objetivo da RAIS é servir como instrumento para o acompanhamento do mercado de trabalho formal, bem como ter o controle da arrecadação de benefícios sociais e trabalhistas⁹.

Cada registro da RAIS identifica uma relação de trabalho formal (constituída por um trabalhador e uma firma) via Consolidação das Leis do Trabalho (CLT). Os estabelecimentos estão registrados no Cadastro Nacional de Pessoa Jurídica (CNPJ) ou são identificados pelo número de matrícula no Cadastro Específico do Instituto Nacional do Seguro Social - INSS (CEI). A base de dados consolida 99% do universo do mercado de trabalho formal do Brasil (MT, 2017) e, nesse sentido, pode ser considerada um censo dos trabalhadores formais. Os microdados estão organizados de tal forma que permite acompanhar a trajetória dos trabalhadores no nível geográfico (município, estado), setorial e ocupacional ao longo do tempo, trazendo também informações selecionadas da firma. Adicionalmente, variáveis-chaves permitem que os microdados de trabalhador e estabelecimento possam ser combinados.

A base de microdados da RAIS é um diferencial em relação às demais supracitadas por seu formato longitudinal, que permite estudar questões relevantes do mercado de trabalho sob uma abordagem de dados em painel, o que tem sido recorrentemente explorado nos diversos trabalhos sobre o mercado de trabalho formal no Brasil. Um potencial diferencial não explorado nessa base de dados é sua característica espacial intra-urbana. A partir das informações de endereçamento dos estabelecimentos é possível identificar latitudes e longitudes, explorando uma abrangência geográfica nunca utilizada com a base de dados da RAIS em trabalhos acadêmicos. Uma desvantagem da base de dados é a falta de padronização quanto à informação de endereços e a demanda de suporte computacional avançado para armazenamento, tratabilidade dos microdados e estimação de modelos.

Para a análise intra-urbana colocada nesse trabalho, será considerada apenas a Região Metropolitana de São Paulo, abrangendo 7.744.879 empregos e 442.388 empresas em 2014. Em termos agregados, segundo os dados extraídos da RAIS, a RMSP teve o emprego formal ampliado em 57,8% entre 2002 e 2014 – muito próximo ao resultado derivado pela PNAD no capítulo 2

⁹ Tais como, Fundo de Garantia por Tempo de Serviço (FGTS), Programa de Integração Social (PIS) e Programa de Formação do Patrimônio do Servidor Público (PASEP).

deste trabalho (59,1%), enquanto o estado de São Paulo (ESP) e o Brasil apresentaram taxas de crescimento de 63,9% e 72,8%, respectivamente. Em termos de participação, o total de empregos formais na RMSP representa 39% do total de empregos do ESP em 2002 e 37,6% em 2014. Quando comparado ao Brasil, essa relação é de 17,1% em 2002 e 15,6% em 2014, demonstrando uma queda de participação relativa. Importante destacar que dessas proporções, o município de São Paulo (MSP) tem a maior parcela do total de emprego da RMSP, representando 68,5% nos anos de 2002 e 2014. Essa relação é de 39% (2002) e 37,6% (2014) do total de emprego do ESP e, quando comparado ao Brasil, o total de emprego do MSP representa 11,7% (2002) e 10,7% (2014). Portanto, nota-se a alta participação relativa do emprego formal da RMSP tanto no contexto estadual e nacional, cabendo ao MSP a maior relevância.

Nas próximas subseções estão descritos os procedimentos utilizados para o georreferenciamento da base de dados, apontando as etapas e pressupostos assumidos. Em seguida são descritos os procedimentos para a limpeza dos microdados e para a formulação de um painel de dados, explorando a característica longitudinal da RAIS.

3.2.2. Padronização e Geocodificação da RAIS

A informação dos endereços das firmas é a relevante neste trabalho, permitindo a identificação de centralidades e subcentralidades no espaço intra-urbano. Para alcançar o objetivo central dessa tese, é necessário obter a microlocalização dos agentes que demandam trabalho. Portanto, o primeiro e principal desafio que se coloca é o georreferenciamento das firmas. Devido às limitações de recursos financeiros e capital físico geradas pela quantidade de observações, associado à complexidade do processo de geocodificação, justifica-se o porquê da microlocalização se dar apenas no âmbito da RMSP.

Em função do requerimento primordial de localizar a firma no espaço intra-urbano, faz-se necessário padronizar conceitos e nomenclaturas, objetivando facilitar o entendimento. Primeiro, segundo a Empresa Brasileira de Correios e Telégrafos (Correios)¹⁰, o endereço postal no Brasil tem a seguinte estrutura: *tipo do logradouro + nome do logradouro + número + complemento* (se

¹⁰ Trata-se de uma empresa pública federal responsável pela execução do sistema de envios e recebimentos de correspondências no Brasil.

houver) + *nome do bairro* + *município* + *sigla da unidade da federação* + *código de endereçamento postal* (CEP).

Alguns esclarecimentos são necessários. Entende-se por *tipo do logradouro* ruas, travessas, avenidas, rodovias, etc. Para melhor compreensão o *nome do logradouro* é subdividido em *título* e *nome do logradouro*. Em muitos casos o *nome do logradouro* é precedido por *títulos* como doutor, mestre, engenheiro, professor, etc. Adicionalmente, define-se a nomenclatura utilizada na dissertação dessa tese. Adota-se *endereço* como sendo a junção das informações de *tipo*, *título* e *nome do logradouro*. A nomenclatura *endereço completo* refere-se à junção de *endereço* e *número*.

O código postal possui orientação geográfica, hierarquizada e classificada em função do destinatário, seguindo a padronização dos Correios. O CEP é composto por um vetor de 8 algarismos e está estruturado segundo o sistema decimal, seguindo a estrutura: *Região* + *Sub-região* + *Setor* + *Subsetor* + *Divisor de Subsetor* + *Identificadores de Distribuição*. Esquematicamente: XXXXX-XXX. Em que cada X pode conter algarismos inteiros do intervalo 0-9. Os três últimos algarismos são conhecidos como sufixo de distribuição.

O escalar da coluna *Região* identifica uma das dez regiões postais do Brasil (divisão derivada dos Correios), sendo a Região Metropolitana de São Paulo a Região 0 (zero). Assim, todos os CEP da RMSP iniciam pelo algarismo 0 (zero). Cada região postal foi dividida em 10 sub-regiões, identificada por *Sub-região*. As sub-regiões postais por sua vez também foram subdivididas em 10 subsetores, sendo identificadas pela coluna *Setor*. O quarto algarismo representa a divisão de cada *Setor*, dada origem do *Subsetor* (o subsetor é a cidade postal sede). Cada *Subsetor* também é dividido em 10 divisores de subsectores, que identifica a coluna *Divisor de Subsetor*. Os três últimos algarismos que identificam os *Identificadores de Distribuição* destinam-se à identificação precisa da localidade, logradouro, Códigos Especiais e Unidades dos Correios (logradouros: 000 a 899; códigos especiais: 900 a 959; CEPs promocionais: 960 a 969; Unidades dos Correios: 970 a 989 e 999; Caixas Postais Comunitárias: 990 a 998). Resumidamente, o código postal (5 primeiros dígitos) da RMSP varia entre 01000-09999 (Correios, 2018).

Assim, por *endereço postal* se referirá à inclusão do CEP ao *endereço completo*. Esses conceitos são essenciais para explicitar cada estágio do processo de georreferenciamento dos dados e a discussão sobre a utilização das coordenadas de latitude e longitude.

Como o objetivo oficial da compilação das informações trazidas pela RAIS não é avaliar movimentos intra-urbanos das firmas, as informações que permitem identificar a posição espacial não são padronizadas. Assim, informações como *tipo* e *título* de logradouro não seguem um padrão em sua grafia, podendo ser escritas por extenso ou abreviadas. A mesma questão se apresenta quanto ao *nome do logradouro*, que pode iniciar com letras maiúsculas ou minúsculas e pode ou não conter acentos gráficos. Adicionalmente, nomes de logradouros que são numéricos, como Rua 25 de Março, podem ser escritos por extenso, com números arábicos ou romanos. A numeração e complementos geralmente procedem o *tipo*, *título* e *nome do logradouro* e podem estar separados por vírgulas, traços ou espaços em branco. Não reportar ou reportar de forma incompleta o número do CEP também pode ser observado na base de dados. Portanto, a padronização da base de dados demanda esforço em tornar as composições do *endereço postal* adequadas para o processo de georreferenciamento.

Adicionalmente à questão da padronização dos dados, muitas firmas reportam seu endereçamento postal de forma parcial, ou não reportam. Comumente, encontram-se apenas o endereço completo ou o nome do logradouro, mas sem numeração ou apenas o CEP. Em diferentes estágios do processo de georreferenciamento tentou-se completar o endereço postal, como será discutido adiante. Informações que são sempre reportadas são a localização municipal e estadual das firmas.

No processo de microlocalização das firmas na Região Metropolitana de São Paulo foram utilizados os arquivos cartográficos georreferenciados (*shapefiles*) dos logradouros da RMSP disponibilizados pelo Centro de Estudos da Metrópole¹¹ (CEM). A base geocodificada fornece informação da estrutura viária da região metropolitana, possibilitando identificar praças, parques e canais hidrográficos. Esses *shapefiles* são essenciais para identificar a localização e a geocodificação pontuais das firmas.

O *shapefile Logradouro da Região Metropolitana de São Paulo edição 2016* (CEM, 2016) totaliza 326.912 segmentos (cada trecho viário, considerando-o de esquina a esquina). Deste total, 89%

¹¹ Trata-se de uma instituição de pesquisa sediada na Universidade de São Paulo e no Centro Brasileiro de Análise e Planejamento (CEBRAP). O CEM investiga temáticas relacionadas às desigualdades e à formulação de políticas públicas nas metrópoles brasileiras contemporâneas. A instituição é financiada pela Fundação de Amparo à Pesquisa do Estado de São Paulo) e conta com um grupo multidisciplinar de pesquisadores das áreas de Demografia, Ciências Políticas, Sociologia, Geografia, Economia e Antropologia.

possuem *endereços postais*. Claramente, esse é um número relevante, dada a dinâmica da região metropolitana mais populosa do país.

A uniformização dos *tipos e títulos* seguiu, principalmente, o menu de possibilidades adotado pelo CEM (2016). Nesse arquivo são destacadas as várias possibilidades de apresentação gráfica dos tipos e títulos. Adicionalmente, consideraram-se as grafias mais observadas nas bases de dados da RAIS e conectamos à base do CEM (2016), seguindo a mesma padronização. O passo seguinte foi parear a base de dados do CEM (2016) estendida com as múltiplas formas de *tipos e títulos* apresentadas na base da RAIS, gerando a uniformização necessária para as composições em questão.

Outro importante passo foi padronizar os *nomes dos logradouros*. Nessa fase, o foco esteve em retirar todos os acentos gráficos e tornar todos os caracteres maiúsculos, como são apresentados na base de arruamento do CEM.

Ainda na etapa da padronização, buscou-se retirar todos os complementos do endereço completo. Isto é, informações sobre o conjunto, o andar, a sala, etc. Essas informações não geram perdas à identificação das centralidades, dado que essas serão identificadas a partir da concentração de emprego na microlocalização considerando sua posição ao longo dos logradouros, não importando sua posição vertical (andar que ocupa no prédio). Logo após foram concatenadas todas as informações, formando os *endereços completos*. Todos os pontos, vírgulas, duplos espaços e espaços no início do endereço foram eliminados.

O último passo que precede a obtenção do *endereço postal* é ajustar as informações dos CEP. A base de dados da RAIS para a RMSP, em sua grande maioria, fornece números de CEP com 7 algarismos. A regra estabelecida pela empresa postal oficial do Brasil (Correios) é que o CEP contenha 8 algarismos, como discutido acima. Essa frequência ocorre porque os municípios da RMSP estão localizados na macrorregião postal cujos CEP iniciam com o algarismo zero. A fim de padronizar, adicionaram-se zeros à esquerda de todos os CEP com 7 algarismos em sua composição.

Para o ajuste da base de dados são seguidas seis etapas de tratamento dos dados. Dado que a informação do trabalhador é essencial para o estudo, o primeiro passo para o tratamento dos microdados da RAIS-firma foi excluir todas as firmas cujo identificador de RAIS negativa fosse

igual a 1, ou seja, firmas que não tiveram trabalhadores durante o ano-base (Etapa I). A taxa de exclusão média, ao longo dos doze anos em análise, foi de 64,4% do estoque de firmas registradas pela RAIS. Outra exclusão importante deriva da impossibilidade de identificar a verdadeira firma quando mais que um estabelecimento apresenta CEI e CNPJ idênticos no mesmo ano-base avaliado¹² (Etapa II) – a taxa média de exclusão foi de 0,5%. Comparando os códigos de *Região* do CEP com os municípios declarados na base RAIS-firma, parte das observações divergiam em ambas informações. Em função da impossibilidade de identificar qual informação é correta, optou-se por excluir as firmas com informações divergentes (Etapa III), reduzindo o estoque em uma taxa média de 8,9%, em relação à Etapa II. Após esses ajustes dos microdados, a base inicial que agregava o montante 4.906.071 (2002) e 7.744.894 (2014) de trabalhadores ativos foi reduzida para 4.348.157 (2002) e 7.155.690 (2014). O Apêndice B traz o total de observações após cada etapa de ajuste da base de dados anualmente.

Após as subetapas no estágio de padronização dos *endereços completos* e códigos postais, tornou-se possível geocodificar as firmas. O georreferenciamento foi conduzido utilizando ArcMap 10.5.1. Utilizaram-se *address locators* com o estilo brasileiro considerando o *shapefile* de arruamento disponibilizado pelo CEM (2016) e, de forma complementar, o arruamento da base de dados do GeoNames (2017)¹³.

A etapa de georreferenciamento (EG) segue duas subetapas. Na primeira subetapa criaram-se os seguintes *address locators* no estilo brasileiro: 1) *endereço completo + CEP* e 2) *endereço completo + município*. Em seguida, georreferenciaram-se os dados considerando o ajuste da etapa de padronização.

Parte das observações não foi geocodificada na primeira etapa de georreferenciamento porque não constava endereço completo, a grafia do *endereço* apresentava algum caracter divergente do *shapefile* de arruamento do CEM (2016) ou esse deixou de cobrir ruas citadas na base de dados da RAIS. Na segunda subetapa do georreferenciamento, buscando superar essa limitação, foi utilizado o *locator World* do ArcGIS 5.1. Nessa nova composição utilizaram-se as seguintes

¹² Esse problema pode ocorrer caso os estabelecimentos filiais registrem o mesmo CNPJ da matriz ou devido erros no preenchimento da declaração.

¹³ GeoNames é uma base de dados fornecida pelo ArcGIS que compila o arruamento para todo o globo terrestre. A base está disponível no serviço *online* de georreferenciamento do ArcGIS e sua utilização é taxada em função da quantidade de elementos georreferenciados.

composições para formar o *address locator*: 1) *endereço completo + CEP*, 2) *endereço completo + município* e 3) *CEP (8 dígitos e 5 dígitos)*.

Address locator que consideram o *endereço postal* é a forma mais consistente de geocodificar dados. Como discutido acima, do *endereço postal* consta nome do logradouro, número e o código postal do imóvel. O número do imóvel é uma informação relevante, sobretudo em logradouros longos, pois localiza a posição exata da firma ao longo da via. Os nomes do logradouro e o CEP reforçam a qualidade do georreferenciamento ao impor mais duas restrições, considerando o número do imóvel, no processo de localização geográfica. Georreferenciamentos derivados de *address locators* que consideram apenas o CEP fornecem o logradouro exato quando utilizados os CEP 8 dígitos, mas não a posição no logradouro. O CEP com 5 dígitos identifica apenas o *Divisor de Subsetor*, isto é, sub-áreas dentro do município.

Mesmo que não tão precisa quanto os *address locators* derivados do endereço postal, *address locators* derivados apenas do CEP identificam microlocalizações posicionando as firmas no centroide do logradouro (CEP 8 dígitos) ou no centroide da faixa de numeração, para o caso de logradouros longos, e no centroide de sub-áreas (CEP 5 dígitos). Utilizaram-se latitudes e longitudes derivadas de *locators* que usam, exclusivamente, CEP 8 dígitos somente quando o município possuir arruamento oficial, isto é, municípios que apresentam *suífixo de distribuição* diferentes de 000 identificam os logradouros¹⁴.

Entre os 39 municípios que compõem a RMSP, 18 municípios¹⁵ têm CEP único. A estratégia utilizada para município que tenha oficializado o arruamento municipal durante o período estudado é considerar o ano da alteração feita pelos Correios, e então atribuir as mesmas latitudes e longitudes do ano referência às firmas que estavam na RAIS-firma nos períodos anteriores à alteração. A hipótese assumida é que as firmas não teriam alterado sua localização nos anos

¹⁴ Segundo os Correios, a implementação de CEP por logradouro ocorre em municípios com mais de 50 mil habitantes e com arruamento oficial e cartografado.

¹⁵ Arujá, Biritiba Mirim, Caeiras, Cajamar, Embu Guaçu, Francisco Morato, Guararema, Juquitiba, Mairiporã, Pirapora do Bom Jesus, Rio Grande da Serra, Salesópolis, Santa Isabel, São Lourenço são municípios com CEP único entre 2002 e 2012. A ressalva é para o município de Arujá que a partir de 2012 tornou oficial o arruamento do município.

anteriores ao ano referência¹⁶. As firmas de municípios com CEP único que não foram georreferenciadas pelo *endereço postal* foram excluídas da análise (Etapa IV). Em relação à Etapa III, o total de firmas e empregos reduziu na média do período avaliado em 2,8% e 5,2%, respectivamente.

Após as duas etapas de georreferenciamento, checkou-se a aderência das latitudes e longitudes aos municípios. Primeiramente, plotaram-se os pontos sobre o *shapefile* da Região Metropolitana de São Paulo (IBGE, 2017), atribuindo a cada ponto o código derivado do *shapefile*. No segundo momento, compararam-se os códigos de municípios derivados do *shapefile* da RMSP com os atribuídos pela RAIS (MTE). A diferença não nula dessas duas regiões identifica erros de georreferenciamento. Observações que apresentaram diferenças não nulas também foram excluídas (Etapa V), representando uma redução de 1,2% em relação à Etapa IV.

Outro importante tratamento está relacionado à atividade que a firma desenvolve. Há atividades dispersas que estão cadastradas na sede, por motivos diferentes. Um dos motivos decorre da normalização do administrador, como o caso de empreendimentos da administração pública. Outro é decorrente da própria atividade, tais como motoristas de ônibus, táxi, mestres de obras, etc. Considerar estabelecimentos cujos registros sejam conectados à sede ou cuja atividade do trabalhador não seja desenvolvida *in loco* leva a erro de medida e enviesada o resultado da estimação.

Estabelecimentos da administração pública registram os trabalhadores com o endereço da sede do governo (endereço da prefeitura ou do palácio do governo). Assim, há uma superestimação do emprego na RMSP, puxado pelo município sede do governo do estado (município de São Paulo), e, mais precisamente, nas microlocalizações do distrito da Sé (sede da prefeitura) e no distrito do Morumbi (sede do governo do estado). Por esse motivo, estabelecimentos sob a administração pública¹⁷ foram excluídos e, conseqüentemente, seus trabalhadores.

É importante também destacar atividades de elevada mobilidade espacial durante a jornada de trabalho, isto é, atividades relacionadas ao setor de transportes, limpeza urbana, construção civil, e

¹⁶ Essa estratégia foi utilizada apenas para o município de Arujá e Caeiras, pois foram os únicos que oficializaram o arruamento, ambos em 2012.

¹⁷ Administração pública em geral, regulação das atividades sociais e culturais, regulação das atividades econômicas, atividades de apoio à administração pública, relações exteriores, defesa, justiça, segurança e ordem pública, defesa civil e seguridade social.

captação, tratamento e distribuição de água. Essa característica da atividade impõe viés à identificação de subcentralidades, pois agrega todos os funcionários na sede do estabelecimento enquanto o desenvolvimento do trabalho ocorre em outras localidades da área intra-urbana. A exclusão de tais atividades foi embasada na classificação CNAE 1.0. Em relação ao setor da construção civil, é importante destacar que atividades como incorporação e compra e venda de imóveis são agregados no setor de comércio e administração de imóveis, valores mobiliários e serviços técnicos.

Após as exclusões setoriais (Etapa VI), o total de estabelecimentos na amostra é de 238.918 (2002) e de 348.807 (2014), uma redução de 5,7% e 8,2% respectivamente em relação à Etapa V. Em relação ao total de empregos, este reduziu em 23% em 2002 e 24,5% em 2014 quando comparado à etapa de ajuste anterior. As variações do total de observações, tanto na base de firmas quanto na base de trabalhadores anuais, considerando as etapas de tratamento podem ser verificadas no Apêndice B.

Adicionalmente, o Apêndice B traz informações sobre a parcela de firmas e empregos georreferenciados sobre o total de firmas e de empregos, considerando o Município de São Paulo (MSP), Municípios com CEP Único (MCU) e Municípios com Múltiplos CEPs (MMC)¹⁸. Em relação ao total de firmas, em média, o MSP teve 83% dos estabelecimentos georreferenciados, enquanto os MCU e MMC tiveram, respectivamente, 55% e 78%. Em relação ao total de emprego, 72% foram geolocalizados nos MMC, 66% no MSP e 47% nos MCU. Esses resultados descritivos apontam para a concentração do emprego no funcionalismo público, etapa da limpeza da base de dados que mais reduz o número de empregos. Os MCU apresentam outro corte importante relacionado à identificação da microlocalização. Todavia, essa perda no total de empregos da RMSP representa, na média do período, 2,8% em termos de total de firmas e 5,2% em termos de emprego total quando comparado aos valores totais da Etapa III. (Tabela B.1 e Tabela B.2, Etapa IV).

Após os estágios descritos acima, a base de dados RAIS-firma traz informações que identificam a posição das firmas dentro do espaço intra-urbano. O procedimento seguinte foi anexar as

¹⁸ Nessa categoria é excluído o município de São Paulo.

informações do trabalhador às firmas para fins de caracterização das subcentralidades. Utilizou-se apenas trabalhador cujo vínculo empregatício é não nulo. Assim, informações dos trabalhadores foram extraídas, divididas por intervalos e somadas por estabelecimento, de tal sorte que se obtém a quantidade de trabalhador com suas características: escolaridade, experiência, gênero, renda medida em salários mínimos e tipo de contrato de trabalho. A informação sobre o total de trabalhadores também foi coletada a partir dos microdados dos trabalhadores, pois na RAIS-firma as informações apresentavam erro de reportagem, como estabelecimentos com mais de 2 milhões de empregados. Para a construção da variável massa salarial da firma (total de salário pagos) foi utilizado o valor da remuneração nominal¹⁹ de cada trabalhador recebido no mês de dezembro.

A variável educação é dividida por grau de instrução, sendo 1) analfabetos, 2) até o 5º ano do ensino fundamental incompleto, 3) 5º ano do ensino fundamental completo, 4) 6º a 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 completo²⁰. As variáveis com intervalos não definidos pela RAIS foram discretizadas. A variável experiência foi dividida por anos de trabalho no mesmo emprego: 1) com menos de 1 ano no emprego; 2) entre 1 + 2 anos; 3) 2 + 3 anos; 4) 3 + 4 anos; 5) 4 + 5 anos; 6) 5 + 6 anos; 7) 6 + 7 anos; e 8) com mais de 8 anos. A variável renda, medida em salários mínimos (SM), foi particionada em 9 intervalos: 1) menor que 1 SM; 2) entre 1 + 2 SM; 3) entre 2 + 3 SM; 4) entre 3 + 4 SM; 5) entre 4 + 5 SM; 6) entre 5 + 7 SM; 7) entre 7 + 10 SM; 8) entre 10 + 15; e 9) maior que 15 SM. Informações sobre o setor de atividades foram obtidas na base de estabelecimentos, considerando as 25 subdivisões derivadas da classificação IBGE 80. Os setores seguem a seguinte ordem: 1) extrativa mineral; 2) indústria de produtos minerais não metálicos; 3) indústria metalúrgica; 4) indústria mecânica; 5) indústria do material elétrico e de comunicações; 6) indústria do material de transporte; 7) indústria da madeira e do mobiliário; 8) indústria do papel, papelão, editorial e gráfica; 9) indústria da borracha, fumo, couros, peles, similares, indústrias diversas; 10) indústria química de produtos farmacêuticos, veterinários, perfumaria; 11) indústria têxtil do vestuário e artefatos de tecidos; 12)

¹⁹ A remuneração do trabalhador é compreendida como a massa salarial do trabalhador que compreende rendimentos derivados dos salários, ordenados, vencimentos, honorários, vantagens, adicionais e gratificações. Excluem-se rendimentos derivados do 13º salário (MTE, 2010).

²⁰ A partir de 2006 o questionário da RAIS permite identificar graduados, mestres e doutores. Todavia, para manter a consistência, agregou-se os três graus de instrução em conformidade com os microdados prévios ao ano de 2006.

indústria de calçados; 13) indústria de produtos alimentícios, bebidas e álcool etílico; 14) serviços industriais de utilidade pública; 16) comércio varejista; 17) comércio atacadista; 18) instituições de crédito, seguros e capitalização; 19) comércio e administração de imóveis, valores mobiliários, serviço técnico; 20) transportes e comunicações; 21) serviços de alojamento, alimentação, reparação, manutenção, redação; 22) serviços médicos, odontológicos e veterinários; 23) ensino; e 25) agricultura, silvicultura, criação de animais, extrativismo vegetal.

3.2.3. Modelo Econométrico, Método De Estimação e Estratégia de Identificação

O procedimento padrão para modelar cidades policêntricas tem sido embasado na abordagem de econometria clássica. Todavia, uma forma funcional incorretamente determinada tende a enviesar a inferência da regressão paramétrica (MCMILLEN; MCDONALDS, 1997). Ademais, estimações paramétricas têm baixo poder em áreas metropolitanas complexas, devido aos erros de especificação quando os candidatos a subcentros são espacialmente correlacionados ou heterogêneos²¹ (GIULIANO *et al.*, 2005). *Local Weight Regressions* (LWR) têm sido recorrentemente utilizadas na identificação de subcentralidades. Tais estimações não-paramétricas apresentam-se como alternativas ao procedimento clássico, oferecendo vantagens significativas sobre o método de regressão linear simples. Estimadores não-paramétricos são flexíveis, permitindo que o coeficiente estimado varie entre as áreas de agregação, fornecem maior acurácia em relação às centralidades principais (subcentros na área central da cidade) e às centralidades periféricas quando comparados a outras abordagens, além de permitirem uma visualização em mapas das centralidades estimadas (MCMILLEN; MCDONALD, 1997; MCMILLEN, 2001b; REDFEARN, 2007).

Modelos não-paramétricos fazem inferência sobre a superfície a partir da regressão dos dados, suavizando as observações e mantendo a complexidade necessária para fornecer estimativas imparciais. Ao lidar com modelos não-paramétricos, algumas decisões arbitrárias também são

²¹ Algumas outras abordagens espaciais poderiam ser aplicadas, tais como Kriging ou Econometria Espacial. Entretanto, as relações espaciais (o mecanismo que conecta unidades geográficas, ou a matriz W) são quase sempre medidas com estatísticas de autocorrelação global, como matriz espacial, e o resultado derivado de tais modelos fornece um conjunto de parâmetros globais (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002).

necessárias. O pesquisador deve escolher o tamanho da banda e o *kernel* ao estimar as regressões ponderadas localmente. De acordo com McMillen (2003), o número de subcentralidades identificadas pelo procedimento não-paramétrico e o nível de significância são influenciados pelo tamanho da janela. Um tamanho menor de banda leva a um menor alisamento da superfície da densidade do emprego, o que tende a reduzir a magnitude dos resíduos ao aumentar a variância estimada. O efeito líquido dessas tendências opostas não é claro (MCMILLEN, 2003).

A escolha por um estimador não-paramétrico neste trabalho deriva de sua vantagem em controlar os padrões globais de concentração de emprego ao identificar subcentralidades, dado que, em regiões metropolitanas, a concentração de emprego segue diferentes distribuições no espaço, demandando um procedimento que considere tal heterogeneidade espacial. De acordo com McMillen (2003), regressões ponderadas localmente lidam com tais questões, usando critérios estatísticos padrão para determinar o nível de emprego cuja concentração varia espacialmente dentro da cidade e em toda a área metropolitana. Em suma, a abordagem local fornece uma desagregação local das estatísticas globais, lidando com não-estacionariedade espacial e permite avaliações espaciais dos dados ao longo da área em questão (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002; BIVAND; PEBESMA; GÓMEZ-RUBIO, 2008).

Portanto, para a identificação de subcentralidades, visando superar as limitações derivadas da especificação econométrica discutida acima, a estratégia empírica está embasada na estimação da média espacial da concentração do emprego por uma abordagem não-paramétrica, seguindo McMillen (2001a, 2003) e Kane, Hipp e Kim (2016), como discutido na subseção seguinte.

3.2.3.1. Especificação Econométrica e Métodos de Estimação

A proposta econométrica considerada nesse trabalho segue uma categoria específica de LWR, isto é, Regressões Ponderadas Geograficamente (GWR – *Geographically Weighted Regression*). Para tornar tratável a modelagem, optou-se por agregar o total de trabalhadores por células do grid. Inicialmente, para a área da RMSP, foram criadas i células com 1 km² de área e os trabalhadores foram somados para cada célula, simbolizado por e_i . Essa estratégia empírica associa cada centroide da célula i a um total de trabalhadores, reduzindo significativamente o total de coeficientes locais a serem estimados. Adicionalmente, lida com as críticas em relação às regiões geográficas com áreas distintas (áreas de ponderação do Censo, áreas de mercado de trabalho, etc.)

e aos problemas decorrentes de densidade bruta e líquida do emprego (MCDONALDS, 1987, MCMILLEN, 2001a). Mieszkowski e Smith (1991) apontam que o uso da densidade bruta tende a superestimar os coeficientes estimados, dado que áreas mais afastadas do centro histórico tendem a concentrar maiores extensões de terra não desenvolvidas e Redfearn (2007) destaca que dados agregados em áreas grandes mascaram as concentrações de empregos locais independentes.

A Figura 3.1 apresenta a formatação do grid. Devido à irregularidade da área oficial da RMSP, as células possuem área menor que 1 km² nas franjas da região. Essa diferença é pouco relevante, uma vez que pouco emprego é observado nessas partes da Região Metropolitana.

Figura 3. 1 – Grid da Região Metropolitana de São Paulo

Fonte: Elaboração própria a partir dos *shapefiles* do IBGE. Nota: células com 1 km² de área

A estimativa do emprego médio espacial em um contexto de regressão local é dada por:

$$e_i = \beta_0(u_i, v_i) + \varepsilon_i \quad (3.1)$$

Onde e_i é o total de emprego na célula i , o vetor (u_i, v_i) denota as coordenadas geográficas (latitude e longitude) dos i -ésimos centroides associados às células, $\beta_0(u_i, v_i)$ é a média espacial no ponto i usando métodos embasados em *kernels* e ε_i é o termo de erro aleatório.

A estimativa GWR para o parâmetro $\beta_0(u_i, v_i)$ da equação (1) é dada por:

$$\hat{\beta}_0(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) e \quad (3.2)$$

Onde $\hat{\beta}_0$ representa a média estimada de β_0 , $W(u_i, v_i)$ denota uma matriz de pesos $n \times n$, cujos elementos fora da diagonal principal são nulos e os elementos da diagonal principal são pesos geográficos de cada uma das n observações contidas na regressão para o ponto i (ou célula i), X é um vetor $n \times 1$ contendo valores unitários, e o sobrescrito T denota que o vetor está na transversal.

$$W(i) = \begin{bmatrix} w_{i1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_{in} \end{bmatrix} \quad (3.3)$$

A matriz de pesos tem função essencial na estimação de modelos do tipo GWR. $W(u_i, v_i)$ é a matriz de pesos que conecta as observações para cada sub-amostra local. Tais conexões são especificadas por funções de pesos, em que observações mais próximas da célula i são ponderadas com maior peso quando comparadas a observações mais longínquas.

Assim, a GWR busca obter um estimador não-paramétrico para funções contínuas de localização (u_i, v_i) usando *kernels*. A *log-likelihood* para cada conjunto particular de estimativas não fornece solução única. A maneira de ajustar a otimização é considerar *log-likelihoods* locais e assumir observações próximas a i (BOWMAN; AZZALINI, 1997, FOTHERINGHAM; BRUNSDON; CHARLTON, 2002).

A estimação de Regressões Ponderadas Geograficamente envolve a seleção de bandas (ou janelas) para uma função de peso *kernel* isotrópica espacialmente. Dentre as diferentes funções de ponderação que podem ser usadas para modelar o processo de dependência espacial estão as funções gaussiana, tricúbica e quadrática.

As funções de peso podem ser fixas ou adaptativas. A primeira considera o tamanho da banda constante no espaço, não importando o tamanho da amostra ao redor de i , enquanto o segundo método adapta o tamanho da banda de acordo com o tamanho da amostra. Problemas potenciais podem derivar de funções de peso fixas, tais como calibração considerando poucos dados, dando origem a estimativas de parâmetros com elevados erros-padrões e, por conseguinte, gerando superfícies pouco suaves. Em situações extremas, a estimação dos parâmetros pode ser comprometida devido à variação insuficiente em pequenas amostras (FOTHERINGHAM; BRUNSDON; CHARLTON, 2002). A segunda abordagem busca superar tal limitação, ajustando o tamanho das janelas considerando a variação de densidade da amostra.

No caso desse trabalho, bandas calculadas pelo método adaptativo comprometem os resultados da identificação. Como regiões periféricas tendem a ter pouco emprego, ajustes nas bandas considerando o tamanho da amostra para tais regiões resultaria numa banda excessivamente grande, e menores no centro. Nesse sentido, apenas as bandas fixas serão exploradas. Por exemplo, regiões pouco representativas tenderão a não rejeitar a hipótese nula de média espacial diferente de zero, devido à elevado erro-padrão do coeficiente estimado ou, no caso extremo, não será possível estimar a média espacial.

De acordo com Fotheringham, Brunsdon e Charlton (2000), os resultados derivados de regressões locais são sensíveis à escolha do tamanho da banda. A escolha ótima do tamanho da banda envolve um *trade-off* entre viés e variância: bandas pequenas geram elevada variância, enquanto tamanhos de banda grandes enviesam os resultados de regressões locais. Buscando superar escolhas arbitrárias²², considera-se o critério de informação AIC (*Akaike Information Criterion*) para a seleção das bandas²³.

Os resultados da otimização AIC são sensíveis à função *kernel* escolhida. McMillen e Smith (2003) sugerem a utilização da função *kernel* tricúbica. Todavia, em função da sensibilidade dos resultados às funções escolhidas, como robustez às bandas calculadas serão utilizadas diferentes formas funcionais de *kernel*, tais como gaussiana, biquadrada e tricúbica, cuja escolha estará sujeita ao critério de informação AIC. Matematicamente,

(A) Função Kernel Gaussiana

$$w_{ij} = \exp[-0.5(d_{ij}/b)^2] \quad (3.4)$$

(B) Função Kernel Biquadrada

²² Em McMillen (2001), Redferan (2007) e Kane, Hipp e Kim (2016) o tamanho das bandas é arbitrariamente escolhido. O primeiro autor escolhe 50% da amostra, enquanto Redferan (2007) escolheu 1% das observações mais próximas de cada uma das n regiões avaliadas. Kane, Hipp e Kim (2016) selecionaram o valor que seleciona 120 *grids* como vizinhos.

²³ Os métodos CV (*cross validation*) e GCV (*generalized cross validation*) não foram utilizados devido às limitações apresentadas, tais como elevada sensibilidade a pequenas amostras, outliers e pontos de influência. Ademais, o problema de otimização dos métodos CV e GCV apenas fornecem mínimo global e nenhuma informação relacionada a estimações locais (FABER; PAÉZ, 2007).

$$w_{ij} = \begin{cases} [1 - (d_{ij}/b)^2]^2 \\ 0, \text{ caso contrário} \end{cases} \quad (3.5)$$

(C) Função Kernel Tricúbica

$$w_{ij} = \begin{cases} [1 - (d_{ij}/b)^3]^3 \\ 0, \text{ caso contrário} \end{cases} \quad (3.6)$$

Onde w_{ij} é uma função contínua de d_{ij} , isto é, a distância entre i e j e b é o tamanho da banda. Utiliza-se a distância euclidiana $d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$.

De forma prática, o modelo econométrico calcula para cada célula do grid médias espaciais do emprego total, considerando funções *kernel* e distâncias de bandas definidas por critérios de otimização. Serão considerados subcentralidades áreas cujas médias espaciais de emprego sejam estatisticamente significantes ($\hat{\beta}_0^{**}$) e sejam maior ou igual a valor crítico c . Buscando diminuir o papel da discricionariedade do valor crítico, este será igual ao percentil χ da distribuição de $\hat{\beta}_0$. Para o exercício, utilizam-se os percentis 99, 95 e 90. Embora tal corte continue ainda a ser discricionário, sua vantagem está em não exigir conhecimento prévio da região e permitir uma adequação temporal desse corte em função da própria distribuição da amostra das médias espaciais estimadas e estatisticamente significativas. Matematicamente,

$$SBD_i = \hat{\beta}_0^{**} \geq c \quad (3.7)$$

A regra de identificação proposta pela equação 3.7 dispensa a utilização do valor de corte pré-definido e estático intertemporalmente. Tal flexibilidade decorre do fato das médias espaciais estimadas serem representativas da concentração do emprego local e, logo, dispensa a tentativa de ajuste por um valor de corte sobre o total de emprego, como ocorre em estimações derivadas de densidade de emprego (GUILIANO; SMALL, 1991, 1999; MCMILLEN, 2002, 2003; KANE; HIPPI; KIM, 2016). Outra vantagem decorre da escolha endógena do valor de corte para a identificação, isto é, o *cut-off* deriva da própria distribuição das médias estimadas, cabendo ao pesquisador a escolha do nível de significância para a inferência estatística. Essa endogeneização da escolha do valor de corte permite comparabilidade temporal dos resultados, mesmo com a variação em nível do total de emprego por célula.

Diferentes resultados podem derivar de diferentes agregações espaciais dos dados, mesmo utilizando mesmas bases de dados e abordagem empírica (GIULIANO; SMALL, 1999; MCMILLEN, 2001a; REDFERN, 2007; KANE; HIPPI; KIM, 2016). Isso pode ser observado ao comparar os resultados estimados por McDonalds (1987) e McDonalds e Prather (1994) e ao avaliar a robustez dos resultados de Kane, Hipp e Kim (2016). Tal questão é conhecida na literatura como Problema de Unidade de Área Modificável (MAUP – *Modifiable Area Unit Problem*).

O MAUP surge devido às múltiplas variedades de agregação dos dados e à disponibilização desses por órgãos competentes, muitas vezes limitando a pesquisa à agregação prévia dos dados. Segundo Wrigley (1996), efeitos de escala e de zoneamento compõem o MAUP. O efeito de escala está relacionado à tendência de se obter resultados estatísticos distintos considerando o mesmo conjunto de dados quando a agregação geográfica é alterada. O efeito de zoneamento diz respeito à variabilidade dos resultados estatísticos em função de uma nova agregação derivada de zoneamentos para a coleta dos dados (por exemplo, alteração do número de setores censitários para a mesma região em diferentes períodos de coleta). Nesse sentido, neste trabalho as médias espaciais da concentração do emprego são também calculadas para *grids* com células de 2 e 4 km² de área, para fins de averiguar a sensibilidade dos resultados.

O acesso a dados sobre o total de área construída utilizada pelas firmas por unidade espacial não é uma informação trivial de se obter, e estimar modelos econométricos em áreas muito diferentes considerando o nível de emprego também comprometeria os resultados. Dado o georeferenciamento realizado nesse estudo, é possível criar áreas tão pequenas e homogêneas quanto o pesquisador requeira, com tamanhos idênticos ao longo da área metropolitana²⁴, superando o viés potencial da normalização²⁵. Em outras palavras, tal medida é uma maneira de

²⁴ Kane, Hipp and Kim (2016) utilizaram a estratégia de grids, todavia a microlocalização se deu por zonas de CEPs. Tal medida pode gerar erros de medida, dado que as zonas de CEPs divergem em extensão, a depender da região em questão. Ademais, o georeferenciamento feito sob essa estratégia agrega todas as firmas no centroide da zona, gerando pesos artificiais em determinadas zonas de CEPs, em termos de concentração de emprego. Nesse sentido, o georeferenciamento considerando o endereço completo das firmas é mais adequado.

²⁵ É importante destacar que o processo de criar *grids* idênticos obedeceu aos limites geográficos pré-estabelecidos pelo IBGE. Isto é, nas periferias municipais e distritais, os *grids* têm área menores do que 1 km². Todavia, tal questão não prejudica os resultados. Pelo contrário, impõe restrição extra, em termos de área para ser considerado uma subcentralidade.

superar as limitações decorrentes da normalização do total de emprego ou da ausência de normalização. Além de trabalhar com medidas mais flexíveis, a variável dependente não será sensível a diferentes valores no denominador a ponto de reverter os resultados.

3.3. RESULTADOS

3.3.1. Características da Amostra

As estatísticas descritivas apresentadas nessa subseção resultam dos microdados ajustados conforme procedimentos descritos acima. A Tabela 3.1 traz informações do total de emprego, número total de firmas, massa salarial e total de empregados por faixas de experiência, tipo de salário, grau de instrução, salário mínimo e gênero.

A partir da Tabela 3.1, que apresenta apenas anos selecionados, nota-se que houve espraiamento das firmas na RMSP. Em 2002, as firmas ocupavam 3.901 das 9.071 células criadas (área de 1 km² cada), isto é, 43% das áreas quadriculadas apresentavam pelo menos um estabelecimento formal com pelo menos um vínculo ativo. Em 2008 a proporção foi ampliada para 46%; em 2014, foi de 48%. Em termos de taxa de crescimento, nos interregnos 2002-2008 e 2008-2014 a área da região metropolitana com emprego formal expandiu em 8% e 4%, respectivamente. Ao subtrair o total de células ocupadas em 2008 do total de células em 2002, o emprego formal foi espraiado dentro de 311 Km². Utilizando a mesma abordagem para os anos 2008 e 2014, o espraiamento ocorreu numa área de 171 Km². Os empregos formais passaram a estar localizados em áreas que até então eram ocupadas por residências e/ou emprego informal, totalizando uma área de expansão de 482 Km² entre 2002-2014.

No ano de 2002 a média de emprego era de 340 vínculos formais ativos por célula. A média é elevada para 496 e 589 vínculos nos anos de 2008 e 2014, respectivamente. Mesmo em um contexto de expansão do total de vínculos ativos e espraiamento do emprego formal, a comportamento concentrador de empregos na RMSP é observado. Essa tendência fica mais clara ao observar o crescimento do desvio-padrão do total de emprego por células. Nota-se que a dispersão da massa de emprego por célula apresentou maior crescimento no período 2002-2008, saindo de 1.532 empregos formais para 2.291, crescendo a uma taxa de 49,4%. Entre 2008 e 2014 o desvio aumenta a uma taxa mais suave (16,6%), alcançando a marca de 2.672 empregos formais. O máximo de empregos é outra característica relevante da amostra: em 2002 apenas uma célula concentrava

54.607 trabalhadores formais, esse valor se eleva para 87.707 e 89.476 em 2008 e 2014, respectivamente.

Embora a identificação de subcentralidades considere a variável emprego, também foram avaliadas as dimensões de total de firmas e massa salarial por célula do *grid*. Em 2002, a média era de 26 firmas por célula, alcançando as marcas de 32 e 38 firmas/célula em 2008 e 2014, respectivamente. Os desvios das médias são mais expressivos no interregno 2002-2008 comparativamente ao período 2008-2014, com taxas de crescimento de 16% e 12%, respectivamente. Os máximos globais da amostra apontam para elevada concentração de firmas espacialmente. Em 2002, 4.272 firmas formais estavam concentradas em apenas uma célula do grid, em 2008 e 2014 os valores aumentam para 4.450 e 4.919, respectivamente.

Mesmo em face do espraiamento espacial do emprego na RMSP, é possível notar características de concentração tanto do emprego quanto dos estabelecimentos, demonstrando uma tendência de aglomeração. Tal comportamento espacial do mercado de trabalho na região metropolitana sugere a existência de efeitos positivos de aglomeração na região, uma vez que se observam movimentos centrípetos ao invés de centrífugos. Ademais, taxas de crescimento de emprego mais elevadas que taxas de crescimento das firmas apontam para o surgimento de firmas com elevada capacidade de contratação.

Em relação à massa salarial, os dados apontam para a presença de desigualdade salarial no espaço intra-urbana. Essa desigualdade pode derivar tanto da elevada concentração de trabalhadores em determinada região e/ou da não aleatoriedade espacial das firmas e trabalhadores mais produtivos, isto é, *sorting*. O desvio-padrão da massa salarial em 2002 era de R\$ 2.460.355, ampliando para 5.397.076 em 2008 e 10.092.028 em 2014. Descontando a inflação e trazendo a valor de 2014 com o Índice de Preços ao Consumidor da Fundação Instituto de Pesquisas Econômicas (IPC-Fipe), os desvios-padrões, em termos monetários são, para os respectivos anos, R\$ 403 mil, R\$ 10.119 milhões e R\$ 10.092 milhões.

Tabela 3.1 – Estatísticas descritivas

continua

Variáveis	2002			2008			2014		
	Média	Desvio	Máx.	Média	Desvio	Máx.	Média	Desvio	Máx.
Emprego	340	1.532	54.607	496	2.291	87.707	589	2.672	89.476
Total de Firms	26	117	4.272	32	136	4.450	38	152	4.919
Massa Salarial (R\$)	408.171	2.460.355	88.542.751	876.598	5.397.076	176.594.369	1.640.874	10.092.028	324.799.763
<i>(em total de empregados)</i>									
<i>Experiência (anos)</i>									
< 1 ano	111	531	20.199	192	950	41.584	213	1.012	36.926
1-2 anos	62	272	9.118	88	403	14.812	103	459	15.122
2-3 anos	41	180	5.679	51	229	7.373	65	290	8.413
3-4 anos	25	111	3.271	34	155	4.761	46	210	5.720
4-5 anos	18	82	2.709	24	106	3.272	33	148	4.276
5-6 anos	16	73	2.458	17	75	1.932	22	106	3.782
6-7 anos	12	57	1.712	14	64	1.511	19	85	3.073
> 8 anos	11	51	1.723	12	58	1.804	14	62	1.685
<i>Tipo de Salário</i>									
mensal	292	1.410	52.775	439	2.158	85.478	539	2.542	87.195
quinzenal	0,4	5	343	0	18	1.641	0	0	14
semanal	0,1	1	64	0	2	97	0	2	63
diário	0,1	4	338	0	3	181	0	2	99
horário	45	222	10.067	52	247	9.311	43	203	7.811
tarifa	0,5	16	1.284	0	8	424	1	86	8.165
outros	2	20	843	4	28	1.067	6	33	1.107
<i>Grau de Instrução</i>									
Analfabeto	3	20	1.306	1	8	386	1	4	119
até 5º ano incompleto	14	73	2.366	10	47	1.681	8	52	2.654
5º ano completo	29	129	4.284	20	86	2.369	14	98	6.624
6º a 9º ano	40	159	3.900	34	135	3.157	25	119	3.542
fundamental completo	60	237	5.550	64	246	5.838	52	204	4.653
médio incompleto	35	148	5.534	41	161	5.763	43	171	6.670
médio completo	96	462	18.686	217	1.030	45.144	292	1.239	44.925
superior incompleto	20	140	5.758	30	205	7.936	29	183	6.500
superior ou mais	44	324	12.110	79	595	21.820	125	891	25.352
<i>Salário Mínimo</i>									
< 1 SM	10	56	2.155	25	130	5.489	34	172	6.695
1-2 SM	50	259	8.292	181	821	36.127	232	970	38.388
2-3 SM	98	416	14.467	110	475	14.122	127	550	13.553

Variáveis	2002			2008			2014			conclusão
	Média	Desvio	Máx.	Média	Desvio	Máx.	Média	Desvio	Máx.	
3+4 SM	48	205	6.325	51	234	8.003	57	267	7.298	
4+5 SM	29	130	4.337	30	147	4.757	34	176	4.931	
5+7 SM	34	162	5.074	33	178	6.424	37	221	6.517	
7+10 SM	25	135	4.472	25	163	5.348	26	183	5.166	
10+15 SM	20	126	4.286	19	147	4.599	20	166	5.576	
> 15 SM	27	228	9.117	23	214	7.287	21	210	8.506	
<i>Gênero</i>										
homem	203	861	27.001	212	1.118	46.240	312	1.351	41.199	
mulher	137	699	27.606	284	1.214	41.467	277	1.363	48.277	
Número de células ocupados		3.901			4.212			4.383		
Número total de células		9.071			9.071			9.071		

Fonte: elaboração própria a partir dos microdados da RAIS

A maior concentração de trabalhadores está na faixa de experiência com menos de 1 ano de trabalho, com taxa de crescimento das médias de 72,4% e 10,8% para os períodos 2002-2008 e 2008-2014, respectivamente. Nessa categoria é que se encontra a maior variância ao longo dos anos. À medida em que se elevam os anos de experiência, menos trabalhadores compõem as respectivas categorias. Todavia, as taxas de crescimento das médias de trabalhadores por célula do grid no período 2002-2008 variam entre 8% e 42% entre as demais categorias, o que reflete o elevado grau de *turnover* no mercado de trabalho na RMSP nesse interregno, não divergindo do comportamento nacional. Para o período 2008-2014, as taxas variam entre 10% e 38%, havendo menor rotatividade do trabalhador. No intervalo entre 1 e 2 anos, a taxa de crescimento da média foi subiu 16%, refletindo a transição de parte dos trabalhadores da primeira faixa de experiência para a segunda.

A participação média dos trabalhadores formais com contrato de recebimento mensal é de aproximadamente 90% nos anos avaliados. Os 10% restantes dizem respeito aos trabalhadores formais cujo contrato de pagamento é por hora. Em relação ao grau de instrução dos trabalhadores, a média da amostra para 2002 é mais elevada no grupo com ensino *médio completo*, seguida pelo grupo de trabalhadores que apresentam *fundamental completo* e pelo grupo que agrega trabalhadores com ensino superior completo ou mais (mestrado e/ou doutorado). Em 2008, as categorias que apresentam as maiores médias de trabalhadores por célula do *grid* são ensino médio completo (1.030 trabalhadores), ensino superior (595 trabalhadores) e ensino fundamental completo (246 trabalhadores). O mesmo padrão é observado no ano de 2014. Ao comparar taxas de crescimento de concentração de emprego por faixas de grau de instrução, percebe-se queda na participação de trabalhadores com grau de instrução abaixo do ensino fundamental. A avaliação simples das médias de trabalhadores aponta para a elevação da qualificação, em termos de ano de estudos, dos trabalhadores formais.

Outra característica importante da amostra é a concentração de trabalhadores recebendo entre 2 e 3 salários mínimos, seguidos pelos grupos que agregam trabalhadores que ganham entre 1 e 2 e 3 e 4 salários mínimos – considerando o ano de 2002. O crescimento do trabalho formal na RM, considerando o ano de 2014, concentrou trabalhadores cuja remuneração esteve entre 1 e 3 salários mínimos. A maior redução de trabalhadores é observada no grupo de trabalhadores que recebem

acima de 15 salários mínimos. A média nessa faixa cai de 27 trabalhadores por célula do grid em 2002 para 23 trabalhadores em 2008 e 21 trabalhadores em 2014. Em relação ao gênero, a participação média dos homens é superior à participação média das mulheres no mercado formal de emprego em 2002 e em 2014, essa diferença é reduzida ao longo dos anos.

As distribuições espaciais dos empregos, firmas e massa salarial, por locais dos municípios, podem ser melhor avaliadas em mapas. Os tamanhos das barras verticais de cada figura são comparáveis anualmente, pois seguem o mesmo peso e a mesma regra dos valores de corte para formatação dos agrupamentos criados, salvo para a massa salarial, que está em valor nominal. Importante ressaltar que os valores de corte são discricionários, tendo por limites superior e inferior os valores apresentados pela série de microdados da RAIS (2012-2014). Devido à dificuldade em sobrepor os nomes das regiões e as barras verticais nos mesmos mapas, criaram-se gráficos em 2D (Apêndice C). Assim, é possível enxergar e compreender melhor a não aleatoriedade espacial dos empregos, firmas e massa salarial por localidade da RMSP.

As Figuras 3.2-3.4 trazem a microlocalização espacial por nível de emprego. Avaliando a expansão do emprego sobre a área metropolitana, nota-se maior expansão no interregno 2002-2008, comparativamente ao período 2008-2014. A partir da Figura 3.2, pode-se verificar a elevada concentração do emprego no centro expandido do Município de São Paulo. E, ao longo dos períodos avaliados, percebe-se que existe um padrão de aglomeração do emprego em detrimento do espraiamento.

As novas localizações de emprego surgem na proximidade de firmas pré-existentes. A taxa de adensamento por km² é mais elevada do que a taxa de dispersão espacial do emprego. Essa afirmativa se torna mais clara ao comparar a taxa de crescimento da área ocupada por emprego formal *vis-à-vis* a taxa de expansão do emprego sobre a RMSP (como discutido acima e demonstrado na Tabela 3.1) e o crescimento das barras verticais que medem a concentração de emprego por km².

Embora os empresários apresentem preferência pela aglomeração (adensamento é relativamente mais expressivo do que o espraiamento do emprego), os dados apontam para a concentração em lugares específicos. A avaliação intertemporal das figuras apresenta a alteração da estrutura urbana da RMSP quando se avaliam os máximos de concentração de emprego. No Município de São Paulo

estão as duas regiões que mais concentram empregos formais relativamente ao restante dos municípios da RMSP. Considerando a divisão de distritos do MSP, como discutida no capítulo 2, essas regiões estão localizadas entre os limites dos distritos da República (REP) e Sé (SE) e dos distritos de Jardim Paulista (JDP) e Bela Vista (BVI), mais precisamente, na Avenida Paulista. Para o ano de 2002, a primeira região agrega entre 45.001–65.000 postos de empregos formais e a segunda, entre 35.001–45.000. As zonas Central e Sul 1 do Município de São Paulo, ao longo dos períodos, apresentaram forte crescimento concentrado de empregos formais, expandindo em direção ao sul da cidade. Após oito anos, as mesmas regiões ranqueiam entre os dois primeiros intervalos de divisão do total de emprego da amostra. Em 2014, a zona Central do MSP continua concentrando emprego, e o devido destaque deve ser dado à concentração de emprego no distrito do Itaim Bibi (IBI), que se tornou tão relevante quanto a concentração JDP/BVI. Pode-se notar, na Figura 3, a formação de uma cordilheira intra-urbana de emprego concentrando entre 15.000–115.600 empregos formais passando pelos distritos Bresser (BRE), Santa Cecília (SCE), República (REP), Sé (SE), Consolação (CON), Jardim Paulista (JDP), Bela Vista (BVI), Liberdade (LIB), Pinheiros (PIN), Itaim Bibi (IBI), Santo Amaro (SAM).

Áreas relevantes em termos de emprego também podem ser observadas fora da capital. Na Sub-região Leste destaca-se a aglomeração formada no Município de Guarulhos, apresentando concentrações de empregos formais tão elevadas quanto determinadas regiões do MSP que formam as cordilheiras intra-urbana de empregos. Na Sub-região Sudeste destacam-se as concentrações formadas na Região do ABCD (Santo André - SAD, São Bernardo – SBC, São Caetano do Sul – SCS, e Diadema - DIA). Dentre esses, é importante destacar a maior representatividade do Município de São Caetano do Sul quanto à concentração de empregos formais. Desde 2002 o município de SCS apresenta aglomeração entre 15.001–35.000 empregos. Em 2014, o município de SAD também alcança o patamar de pelo menos uma célula do *grid* compondo o intervalo 15.001–35.000 de empregos formais.

Destacam-se na Sub-região Oeste os municípios de Barueri (BAR) e Santana do Parnaíba (STP). Barueri apresenta expansão tanto horizontal quanto vertical do total de emprego por km². Ao comparar as distribuições espaciais do nível de emprego, novas células de emprego apareceram ao redor das áreas mais adensadas. Comparativamente, a partir da Figura 3.2, percebe-se que no município de Barueri concentra mais empregos do que qualquer outro município da RMSP,

excetuando a capital. É importante destacar o município de Santana do Parnaíba, pois esse perde elevada parcela do emprego que se concentrava no limite geográfico do município de Cajamar no ano de 2002. Tal perda está associada às saídas de indústrias da região. Na Sub-região Leste da RMSP, Guarulhos, Poá e Moji das Cruzes se destacam por apresentarem as maiores concentrações de emprego para os três anos analisados. Todavia, o município de Guarulhos tem a região com maior concentração de emprego dos três municípios destacados.

Figura 3.2 – Distribuição espacial do nível de emprego (ano 2002)

Fonte: Elaboração própria a partir dos microdados da RAIS, 2002.

Figura 3.3 – Distribuição espacial do nível de emprego (ano 2008)

Fonte: Elaboração própria a partir dos microdados da RAIS, 2008

Figura 3.4 – Distribuição espacial do nível de emprego (ano 2014)

Fonte: Elaboração própria a partir dos microdados da RAIS, 2014

As distribuições espaciais de firmas estão fortemente correlacionadas com as respectivas distribuições de emprego, como se verifica nas figuras 3.5 – 3.7. Pode-se também apontar para a elevada relação trabalhador-firma nas áreas de elevada aglomeração, ao considerar os limites superiores do total de emprego e de firma, aludindo à elevada concentração de trabalhadores por firmas. Essa característica sugere que nos locais com alta concentração de emprego, poucas firmas demandam elevada quantidade de força de trabalho. Embora mais firmas passem a se alocar em regiões com elevada aglomeração, tal crescimento não altera a característica central dos empregos formais na RMSP, isto é, elevado nível de emprego por firma.

As células localizadas nas zonas Centro e Sul 1 e Oeste apresentam as maiores concentrações de firmas em 2002, com destaque para as áreas localizadas entre os distritos República (REP) e Sé (SE), que agregam entre 3.501-5.000 firmas. Áreas entre os distritos REP e Brésler (BRE) e REP e Consolação (COM) ranqueiam entre 2.501-3.500. Os distritos de Jardim Paulista (JDP), Pinheiros (PIN), Itaim Bibi (IBI), Saúde (SAL) e Vila Mariana (VMN) (zona Sul 1) são representativos na categoria entre 701-2.000 firmas. Nos demais períodos demonstrados nas figuras 3.6 e 3.7 houve uma expansão ao redor da célula do grid localizada entre os distritos da República e Sé e um crescimento em direção à zona Sul 1 da cidade, quando avaliado apenas o Município de São Paulo. Destaque para o ganho de relevância de áreas no distrito de Santo Amaro (SAM) no ano de 2014 (zona Sul 2), Perdizes (PER) e Lapa (LAP) (zona Oeste). Essa trajetória de crescimento vai ao

encontro do projetado por Meyer, Grostein e Biderman (2004). As regiões do ABDC, Guarulhos, Barueri e Mogi das Cruzes também tiveram o nível de firmas ampliados nos três períodos avaliados, alcançando patamar tão relevante quanto os distritos do extremo da zona Oeste e Sul 1 e 2.

Figura 3. 5 – Distribuição espacial das firmas (ano 2002)

Fonte: Elaboração própria a partir dos microdados da RAIS, 2002

Figura 3. 6 – Distribuição espacial das firmas (ano 2008)

Fonte: Elaboração própria a partir dos microdados da RAIS, 2008

Figura 3. 7 – Distribuição Espacial das Firms (ano 2014)

Fonte: Elaboração própria a partir dos microdados da RAIS, 2014

A massa salarial (soma de todos os salários recebidos na célula) é outra variável importante, pois permite identificar as áreas com maiores concentrações salariais. Os valores nominais plotados nas figuras comprometem a comparação intertemporal. O período 2002-2014 é um período de redução da inflação. Segundo dados do IPC-FIPE, a inflação acumulada no ano de 2002 foi de 1,8 (9,92% de alta em relação a 2001) alcançando a marca de 0,16% em 2008 e 0,3% em 2014. O que justifica dizer que, em alguma medida, tal aumento da massa salarial não deriva essencialmente de processo inflacionário, mas do aumento do nível do emprego e/ou do aumento do nível salarial dos trabalhadores.

Em 2002, a área com maior massa salarial relativa está localizada no distrito de Jardim Paulista (Avenida Paulista) – ver Figura 3.8. Essa característica está relacionada tanto à quantidade de trabalhadores na região, como discutido acima, quanto à elevada massa salarial média na região, alcançando a marca de R\$ 88.542.751. Na área entre os distritos República e Sé, a segunda maior massa salarial é de R\$ 78.155.222. Importante notar que a cordilheira de salários segue em direção às zonas Oeste-Sul 1 (Pinheiros, Morumbi, Itaim Bibi, Santo Amaro). Em 2008, na Figura 3.9, a região da Avenida Paulista (Bela Vista-Jardim Paulista-Consolação) apresenta duas células mais relevantes no que diz respeito à massa salarial, sendo uma delas a maior concentração entre todas as demais áreas, com massa salarial de R\$ 176.594.368. A célula localizada entre REP/SE ocupa a terceira posição do *rank*, com mais de R\$ 150 milhões de massa salarial. Outro destaque fica a

cargo da região localizada no Itaim Bibi, representando a quarta mais elevada massa salarial do ano avaliado, R\$ 126.458.657. No ano de 2014, o distrito do Itaim Bibi se consolida como a região com maior massa salarial, mesmo não apresentando a maior aglomeração de empregos – ver Figura 3.10. A região REP/SE perde participação relativa quanto a massa salarial, bem como a região da Avenida Paulista. No restante da RMSP, os municípios de Barueri, São Bernardo do Campo e São Caetano são os mais relevantes em termo de massa salarial em 2008. O município de Barueri ranqueia no topo da distribuição da massa salarial entre os municípios da RMSP (excluindo a capital). Esses resultados sugerem a existência de núcleos produtividade espacialmente heterogêneos.

Figura 3. 8 – Distribuição espacial massa salarial (ano 2002)

Fonte: Elaboração própria a partir dos microdados da RAIS, 2002

Figura 3. 9 – Distribuição espacial da massa salarial (ano 2008)

Fonte: Elaboração própria a partir dos microdados da RAIS, 2002

Figura 3. 10 – Distribuição Espacial Massa Salarial (ano 2014)

Fonte: Elaboração própria a partir dos microdados da RAIS, 2014

A partir da distribuição espacial do nível de emprego, firmas e massa salarial, observa-se forte concentração nas três dimensões no Município de São Paulo, mais precisamente no distrito Jardim Paulista, região onde se localiza a Avenida Paulista. Avaliando a dimensão nível de emprego, o crescimento e adensamento de emprego partem do foco central do Município de São Paulo (centro histórico) e de alguns municípios mais relevantes na RMSP, tais como Barueri, Guarulhos, Santo André, São Bernardo, São Caetano, Diadema e Mogi das Cruzes. Todavia, em termos de massa salarial, essas regiões ganham alguma expressividade, relativamente ao município de São Paulo, em 2014. Portanto, a partir das dimensões avaliadas nestas estatísticas descritivas, há espaço para discorrer sobre concentração espacial de empregos e da renda do trabalho no espaço intra-urbano da RMSP.

3.3.2. Identificação dos Subcentros da RMSP

Esta seção apresenta os resultados da estimação econométrica da concentração de empregos considerando um grid com células com 1 km² de área. A partir do critério metodológico desenvolvido e discutido acima, consideram-se subcentros áreas contíguas cujas médias espaciais estimadas ($\hat{\beta}_0$) sejam estatisticamente significativas e atendam ao critério de corte: quantis de ordem 99, 95 e 90, dado a distribuição das médias estimadas. Nesse contexto, os valores de corte

são identificados de forma endógena, ou seja, são derivados da estimação do modelo econométrico não-paramétrico.

Como discutido anteriormente, os resultados derivados de modelos não-paramétricos são sensíveis ao tamanho das bandas selecionadas. A forma objetiva de selecionar o tamanho ótimo da banda, derivado da otimização da função de máxima log-verossimilhança, é comparar critérios de informação. A Tabela 3.2 apresenta os critérios de informação para cada função kernel utilizada (gaussiana, biquadrada e tricúbica) considerando os seguintes anos selecionados²⁶ de 2002, 2008 e 2014 da RAIS. Como resultado, a função *kernel* tricúbica apresenta menor AIC para todos os anos, indo ao encontro dos resultados da literatura, como discutido em McMillen e Smith (2003). Para o ano de 2002, considera-se uma banda com 2,3 km. Nos anos de 2008 e 2014, as bandas que minimizam o critério de informação são de 2,09 e 2,06 km, respectivamente. Nota-se que os tamanhos das bandas se reduzem, em função do espraiamento do emprego na RMSP.

Tabela 3. 2 – Critério de informação AIC e tamanho das bandas

Função Kernel	2002		2008		2014	
	AIC	Banda (Km)	AIC	Banda (Km)	AIC	Banda (Km)
Gaussiana	150.453,9	0,910	157.403,6	0,879	159.668,7	0,871
Biquadrada	150.289,6	2,271	157.169,2	2,157	159.429,8	2,129
Tricúbica	150.197,6	2,309	157.033,3	2,095	159.279,6	2,063

Fonte: elaboração própria, a partir dos dados da RAIS (2002; 2008; 2014).

Para estimar as regressões, consideraram-se as bandas que minimizam o critério de informação. A estimação de modelos não-paramétricos exige como *input* a definição dos pesos espaciais definidos por funções kernel. Assim, estimam-se funções kernel do tipo gaussiana, biquadrada e tricúbica. Novamente, dentre todos os modelos estimados, a escolha deriva das estimações com menor AIC.

A Tabela 3.3 apresenta os resultados por quantil das estimações para os anos de 2002, 2008 e 2014 com menor AIC²⁷. Primeiramente, é possível sublinhar a heterogeneidade espacial da concentração de empregos na RMSP a partir dos valores mínimos (*Min*), 1º Quartil (*1º Q*), mediana (*Med*), 3º Quartil (*3º Q*) e máximos (*Máx*). Adicionalmente, estimou-se a média do emprego global considerando o estimador OLS (*Global*).

²⁶ Os anos foram escolhidos visando capturar possíveis mudanças mais expressivas na estrutura espacial da localização de empregos.

²⁷ As estimações utilizando diferentes funções kernel podem ser verificadas no Apêndice D.

Nos três momentos avaliados, baixos valores estimados das médias espaciais de empregos podem ser observados. Valores de mínimos e 1º quantil dos coeficientes estimados são iguais a zero para os três anos em análise. Esses resultados decorrem da utilização de todo o espaço da RMSP, tais como áreas verdes, rios, áreas tipicamente residenciais, etc.²⁸ As medianas dos empregos médios estimados também são baixas, apontando para uma distribuição de emprego espacialmente homogênea entre o mínimo e a mediana. No 3º quartil, a média estimada é de 97 empregos para o ano de 2002. Esse valor se eleva ao longo dos anos, alcançando as marcas de 147 em 2008 e 175 em 2014. Os resultados mostram que até o quantil de ordem 50, a variação espacial é pouco relevante, observando-se elevado salto no quantil de ordem 75. Em relação ao valor máximo de emprego médio estimado, existe uma notável heterogeneidade da distribuição espacial dos empregos na RMSP, mesmo com a suavização espacial das funções *kernel*. Os modelos que consideram o estimador OLS desconsideram a não isotropia espacial dos empregos formais na RMSP e superestima os valores das médias estimadas. Para os anos de 2002, 2008 e 2014, o estimador de OLS calcula médias de 340, 496 e 589 empregos, respectivamente. Os coeficientes globais de emprego tendem, portanto, a “esconder” a estrutura espacial heterogênea dos resultados.

Tabela 3.3 – Regressões ponderadas geograficamente

	2002	2008	2014
	I	II	III
Função Kernel	Tricúbica	Tricúbica	Tricúbica
Min.	0,00	0,00	0,00
1º Quartil	0,00	0,00	0,00
Mediana	1,42	2,39	3,46
3º Quartil	96,93	146,63	175,35
Max.	21.883,75	34.296,66	41.130,67
Global	339,99	495,97	589,14
AIC	150.197,6	157.033,3	159.279,6
Quase-global R ²	0,7363	0,7624	0,7795
Num. Obs.	9.071	9.071	9.071

Fonte: elaboração própria, a partir dos dados da RAIS (2002; 2008; 2014). Nota: os quartis se referem a distribuição dos coeficientes estimados pelo estimado GWR.

Diferentemente dos trabalhos que consideram valores de corte *ad hoc* e fixos no tempo, a identificação das subcentralidades deriva da distribuição da amostra do ano em questão,

²⁸ Utilizar toda a área da RM incorpora maior realismo da distribuição espacial do emprego, a qual é limitada por fatores geográficos e políticas públicas. Tais fatores, por consequência, influenciam a própria aglomeração de emprego e seu espraiamento. Portanto, considerar apenas áreas de trabalho com ao menos um trabalhador resultaria em resultados enviesados, sobretudo em áreas de municípios pequenos, que tendem a apresentar apenas uma única concentração de emprego e arredores vazios.

considerando os limites da distribuição dos quantis 99, 95 e 90 e a significância estatística correspondente. Para o ano de 2002, os valores de corte são de 5.764, 1.876 e 819 empregos, respectivamente. Ou seja, as células pertencentes ao conjunto das 1% mais densas têm número de empregos superior ou igual a 5.764. Os valores de corte se alteram substancialmente em relação aos anos de 2008 e 2014. Para o ano de 2008, passam a 8.484 (quantil de ordem 99), 2.654 (quantil de ordem 95) e 1.161 (quantil de ordem 90). No ano de 2014, tais valores são 10.070 (quantil de ordem 99), 3.132 (quantil de ordem 95) e 1.380 (quantil de ordem 90). Para o primeiro corte, independentemente do ano, identificam-se 91 células como SBD; o segundo valor de corte identifica 363 novas células, e o terceiro corte identifica 454 novas áreas, como demonstrado na Tabela 3.4 (na tabela os valores são apresentados como estoque, não incrementos). Cabe destacar que as áreas são robustas às variações no grau de significância estatística, identificando o mesmo total de células.

Tabela 3.4 – Valores de corte e número de SBD identificados

Quantis	Valor de Corte	Nº. de SBD a 10%*	Nº. de SBD a 5%*	Nº. de SBD a 1%*
2002				
99	5.764	91	91	91
95	1.876	454	454	454
90	819	908	908	908
2008				
99	8.484	91	91	91
95	2.654	454	454	454
90	1.161	908	908	908
2014				
99	10.070	91	91	91
95	3.132	454	454	454
90	1.380	908	908	908

Fonte: elaboração própria, a partir dos dados da RAIS (2002; 2008; 2014). * nível de significância estatística considerado na identificação das células do grid.

Embora o total de células identificadas não se altere intertemporalmente, as áreas não são as mesmas, como será discutido abaixo. Tal fato aponta para a estabilidade do comportamento estrutural do mercado de trabalho na RMSP em termo de ocupação espacial, ou seja, mesmo em face do espraiamento do emprego formal na região metropolitana, o total de área (em km²) identificado com SBD se mantém²⁹. Nota-se um deslocamento para a direita da distribuição das médias espaciais estimadas, o que altera, conseqüentemente, o valor de corte. Essa alteração deriva

²⁹ Ao considerar quantis de ordens 0,85 e 0,80 o total de áreas identificadas se altera intertemporalmente. Esse resultado decorre da sensibilidade da distribuição inferior dos empregos formais em áreas menos densas. No caso da RMSP, em áreas periféricas da RM, como esperado.

das diferentes taxas de crescimento do emprego em cada célula ao longo do tempo, alterando a distribuição relativa entre áreas dos *grids*. As diferentes células identificadas intertemporalmente decorre do comportamento dinâmico do mercado de trabalho – fechamentos e abertura de firmas, expansão e retração da economia, etc. Ou seja, SBD podem desaparecer por perda de participação relativa na composição de empregos formais da região metropolitana.

Definem-se subcentros de emprego como uma região que agrega células do grid que seguem os critérios de identificação discutidos acima. Nesse sentido, a totalidade de SBD é identificada a partir da quantidade de *clusters* discretos distribuídos na RMSP. Tal discussão é melhor qualificada quando se avalia a localização espacial das subcentralidades. A nomenclatura dos SBD se dará em função do município ou conjunto de municípios em que se localiza o *cluster*, quando a mancha contínua não se localizar estritamente nas fronteiras político-administrativas do município, isto é: *SBD + nome do município*.

As Figuras 3.11 – 3.22 trazem as disposições espaciais das subcentralidades para os anos de 2002, 2008 e 2014 por ordem do quantil, mostrando apenas as áreas estatisticamente significantes. Estas figuras ilustram os centros que emergiram, morreram, tiveram suas áreas aumentadas e/ou reduzidas, contrastando com as avaliações mais agregadas que são recorrentes na literatura.

As Figuras 3.11 – 3.14 dispõem as posições geográficas das médias espaciais de emprego para o ano-base 2002 para diferentes ordens de quantis. Os resultados derivados do valor de corte igual a 5.764 (quantil de ordem 99) identificam três subcentros: SBD de Barueri (SBD-BAR), SBD de São Caetano do Sul (SBD-SCS) e SBD de São Paulo (SBD-SAO). O SBD-BAR está distribuído numa área de 5 km² e o pico mais elevado de toda a área concentra uma média espacial entre 7.000 e 10.000 empregos formais. O SBD-SCS compreende uma área de 7 km², contendo ao menos uma área de 1 km² com uma média espacial entre 7.000 e 10.000 empregos formais. O SBD-SAO ocupa 79 km² de área, desde a zona central, passando por parte das zonas Oeste, Sul 1 e Sul 2 (ver Figura 3.14 – mapa 2D). Nessa área estão localizados dois dos picos mais relevantes em termos de quantidade de emprego. O primeiro e mais importante apresenta entre 20.000 e 35.000 empregos, localizando-se entre os distritos da República (REP), Consolação (COM) e Bela Vista (BVI). O segundo pico dessa distribuição concentra entre 10.000 e 15.000 empregos no distrito do Itaim Bibi

(IBI). Em termos de concentração de emprego, a região do IBI é mais aglomerada em comparação com os SBD-BAR e SBD-SCS.

Ao relaxar o valor de corte para a identificação dos subcentros (quantil de ordem 95, com valor de corte igual a 1.876), a mancha de empregos se espalha substancialmente ao redor dos SBD identificados com o corte mais restrito – ver Figura 3.12. Em alguns casos, forma uma única grande mancha uniforme, por exemplo o SBD-SCS, que se estende em direção aos municípios de Santo André (SAD) e de São Paulo (SAO). Subcentralidades não conectadas às regiões pré-identificadas surgem na Sub-região Leste da RMSP, nos municípios de Guarulhos (GRU), Poá (POA) e Moji das Cruzes (MJC), e São Bernardo do Campo (SBC) e Diadema (DIA), na sub-região Sudeste. Na Sub-região Oeste, dois SBD passam a ser observados, nos municípios de Santa do Parnaíba (STP) e Osasco (OSA) e um SBD na Sub-região Norte (Cajamar- CJM). Os municípios de Taboão da Serra (TBS) e Embu das Artes (EMB), na Sub-região Sudoeste, também apresentam concentrações relevantes de médias de empregos formais. Todavia, essas regiões não são homogêneas em termos de concentração de emprego. Os SBD-SAD, SBD-SBC, SBD-GRU e SBD-POA apresentam maior relevância em termos de aglomeração de emprego, com pelo menos uma das células entre 3.000 e 7.000 empregos formais, enquanto os SBD-MJC, SBD-STP, SBD-CJM, SBD-OSA, SBD-TBS e SBD-EMB situam-se no intervalo entre 1.000 e 3.000 empregos. O SBD-SAO cresce ao redor do mega SBD-SAO identificado com o valor de corte mais restrito, direcionando-se às áreas periféricas do MSP e distribuindo-se numa nova área de 240 Km².

Novamente, como esperado, ao assumir o valor de corte igual a um quantil de menor ordem, os SBD se espraiam em direção às franjas da RMSP, como demonstrado na Figura 3.13. Percebe-se a formatação de uma única macha de SBD unindo municípios da região do ABCD, São Paulo, Guarulhos, Osasco, Taboão da Serra e Embu. A área SBD-BAR é ampliada e a nova configuração tende a diminuir a distância entre os SBD-BAR e a grande mancha formatada. Cabe destacar, na capital, o aparecimento de SBD separado da grande mancha no distrito de Cidade Dutra, na Zona Sul 1 (SBD-CDU) e outro, nos limites políticos dos distritos de Itaquera, Parque do Carmo na Zona Leste 2 (SBD-ITQ-PQC).

Para o ano-base 2008, as figuras 3.15 – 3.18 trazem as novas áreas que compõem os SBD. Considerando a Figura 3.15 (2008) e a Figura 3.11 (2002), três SBD são identificados e nos

mesmos municípios, sob o valor de corte mais restritivo. Todavia, tanto o SBD-BAR quanto o SBD-SCS perdem área relativamente ao ano de 2002. O SBD-BAR passa agora a estar alocado numa área de 4 km², enquanto o SBD-SCS, em 1 km², sendo que este apresenta menor concentração de emprego (intervalo 7.000-10.000) do que o SBD-BAR (intervalo 10.000-15.000). Dado que no ano de 2008 se observa o mesmo total de área de SBD comparativamente a 2002, esse resultado aponta para o crescimento relativo dos empregos formais no município de São Paulo, os quais passaram a ocupar uma área de 86 km². A média espacial de emprego alcança o pico entre 25.000 e 35.000, concentrando-se espacialmente entre os distritos da República, Sé, Consolação, Boa Vista e Jardim Paulista. A região do distrito do Itaim Bibi também observa significativo crescimento, posicionando-se na categoria de empregos entre 15.000 e 20.000 (média espacial). Importante destacar que o espraiamento das regiões consideradas SBD se dá em direção ao sul (Santo Amaro) e oeste (Lapa) da cidade (ver Figura 3.18), sugerindo a direção em que a descentralização do emprego está ocorrendo. Nota-se um movimento centrífugo do emprego, mas concentrado.

Valores de corte iguais ao quantil de ordem 95, o SBD é composto por uma grande área localizada entre os municípios de Osasco, São Paulo, São Caetano do Sul e Santo André, no ano 2008. Duas áreas menores estão nos municípios de Guarulhos (SBD-GRU I e SBD-GRU II) e Poá (SBD-POA), na Sub-região Leste. Na Sub-região Sudoeste, identifica-se uma subcentralidade formada entre os distritos de Diadema e São Bernardo do Campo, formando o SBD-SBC-DIA, e o SBD-TBS no município de Taboão da Serra. Os SBD identificados em 2002 na Sub-região Oeste não mais são observados, o que aponta para o baixo crescimento relativo dos empregos formais nas áreas em questão *vis-à-vis* as demais áreas.

Ainda em relação ao período de 2008, ao considerar como valor de corte o quantil 90, a mancha forma um mega SBD, agregando municípios das sub-regiões Oeste (OSA), Sudoeste (TBS, EMB), Leste (GRU) e Sudeste (SAD, SBC, SCS, DIA – ABDC). Nesse contexto, são identificados, além do mega SBD, três SBD na região metropolitana (SBD-BAR, SBD-POA e SBD-MJC) e vários SBD satélites (por exemplo, SBD-EMB, SBD-STP, SBD-COT), isto é, pequenas regiões cuja concentração de emprego (médias espaciais) não ultrapassam a marca de 3.000 empregos formais.

Para o ano de 2014, as figuras 3.19 – 3.22 apresentam os resultados estimados. Iniciando com valor de corte de 10.070 empregos (quantil de ordem 99), a redução do total de SBD na RMSP é notória.

Dois SBD são identificados, a saber SBD-BAR e SBD-SAO – ver Figura 3.19. O SBD localizado em Barueri tem a área de localização ampliada em 1 km², comparativamente ao período de 2008, totalizando 5 km². Em relação ao SBD-SAO, as médias foram elevadas substancialmente, alcançando o patamar máximo, entre 35.000 e 45.000, e localizadas nos distritos da República, Sé, Consolação e Jardim Paulista. No distrito do Itaim Bibi é possível identificar significativo crescimento da concentração de empregos, tanto de uma célula específica quanto de sua vizinhança, alcançando o segundo mais elevado posto do ranque (entre 25.000 e 35.000 empregos). Relativamente aos anos anteriores, nota-se a consolidação do espraiamento dos empregos na capital, iniciada nos anos 2000.

Ao relaxar o valor de corte (quantil de ordem 95), os SBD identificados nos anos de 2008 aparecem elencados em 2014, tais como SBD-TBS, SBD-OSA, SBD-GRU I e SBD-GRU II, SBD-MJC, SBD-SBC-DIA – ver Figura 3.20. O destaque fica por conta da concentração de emprego em São Caetano do Sul, cuja área passa a ser identificada apenas com um valor de corte menor. Esse resultado aponta para o baixo crescimento da concentração de emprego relativo aos SBD-BAR e SBD-SAO. Outra mudança relevante ocorre na zona Leste 2 do MSP, onde na região de Itaquera pode ser identificada uma subcentralidade (SBD-ITQ). Essa região era identificada no ano de 2008 sob o critério de corte menos restritivo (quantil de ordem 90). Seguindo o padrão dos anos anteriores, ao utilizar o valor de corte menos restritivo, novamente um mega SBD é observado. Outro destaque relevante diz respeito ao SBD-POA, localizado entre os municípios de Poá e Suzano, o qual é identificado apenas com um valor de corte mais frouxo, diferentemente dos anos anteriores.

A partir dessa avaliação, o padrão encontrado é que novas centralidades derivadas do relaxamento do valor de corte se dão substancialmente ao redor das áreas com maior força de aglomeração de trabalhadores. Poucos são os SBD que surgem em áreas distantes ou isoladas, e poucos os que são identificados por cortes mais restritivos. Esse resultado deriva claramente da abordagem empírica utilizada, onde a concentração de emprego dos vizinhos atua elevando ou reduzindo a aglomeração de empregos média estimada, mesmo em face de funções *kernel* ajustando o distanciamento da célula às suas vizinhas.

Outro ponto relevante, derivado da metodologia desenvolvida, é poder identificar a relativa perda de influência dos SBD, em termos de médias de emprego, à medida que os quantis têm ordens menores. A Tabela 3.5 traz o total de células que movem para uma das três classes (c): morte, persistência e nascimento de novas áreas consideradas SBD (usando células com 1 km² de área). As taxas de mobilidade e imobilidade das áreas identificadas são calculadas como abaixo:

$$Tx\ Mob = \frac{(DIFERENÇA_{t-(t-1)}^{Quantil\ 99})_{c_i}}{(DIFERENÇA_{t-(t-1)}^{Quantil\ 99})_{Totalc_i}} \quad (1)$$

$$Tx\ Imob = 1 - Tx\ Mob \quad (2)$$

Onde, $t = 2002, 2008, 2014$ e c_i indica uma das três classes indicadas acima.

Observando a Tabela 3.5, para os valores de cortes mais restritos, o total de áreas dos SBD que morreram é de 10 km² na comparação 2008 e 2002. Como o total de áreas identificadas é constante, novas áreas surgiram com mesma dimensão. Na comparação 2014 e 2008, 6 km² de áreas perderam participação relativa. Das 10 células que perderam o *status* de SBD em 2008, apenas uma retoma a qualidade de área relevante na concentração de emprego em 2014, representando 10% de mobilidade. Das regiões estáveis no período 2002 e 2008, 5 km² de área surgem do ganho de participação relativa, apresentado uma taxa de mobilidade de 0,1%. Como discutido acima, o diferencial do primeiro período deriva da perda de participação relativa das áreas do SBD-BAR e SBD-SCS e ampliação da área do SBD-SAO. O diferencial entre os anos 2008 e 2014 decorre das perdas no SBD-SCS, ampliando tanto o SBD-BAR quanto SBD-SAO. A mesma avaliação para quantis de ordens mais baixas mostra que a taxa de mobilidade se eleva comparativamente ao quantil anterior, alcançando a marca de 23% no quantil 95 e 18% no quantil 90.

Quanto ao dinamismo do mercado de trabalho na RMSP, observa-se, na capital, a dispersão para áreas mais periféricas, principalmente Zona Oeste e Sul 1, com concomitante consolidação das regiões com elevada concentração de emprego. Fora da capital, nos municípios de Guarulhos (GRU) e São Caetano do Sul (SCS), eleva-se o total de emprego, formando-se *clusters*. Todavia, o SBD-SCS apresenta taxas de crescimento menores do que as principais áreas identificadas com valor de corte mais restrito. Os SBD identificados no limite dos municípios de Santana do Parnaíba e Cajamar perdem sua relevância relativa, deixando de elencar como SBD.

Tabela 3. 5 – Análise de persistência dos SBD identificados

<i>DIFERENÇA^{Quantil 99%}_{2014–2008}</i>				
<i>DIFERENÇA^{Quantil 99%}_{2008–2002}</i>	Morte	Persistência	Nascimento	Total
Morte	0	9	1	10
Persistência	5	9041	5	9051
Nascimento	1	9	0	10
Total	6	9059	6	9071
<i>DIFERENÇA^{Quantil 95%}_{2014–2008}</i>				
<i>DIFERENÇA^{Quantil 95%}_{2008–2002}</i>	Morte	Persistência	Nascimento	Total
Morte	0	27	8	35
Persistência	27	8937	37	9001
Nascimento	18	17	0	35
Total	45	8981	45	9071
<i>DIFERENÇA^{Quantil 90%}_{2014–2008}</i>				
<i>DIFERENÇA^{Quantil 90%}_{2008–2002}</i>	Morte	Persistência	Nascimento	Total
Morte	0	56	13	69
Persistência	28	8864	41	8933
Nascimento	26	43	0	69
Total	54	8963	54	9071

Fonte: elaboração própria, a partir dos resultados do modelo GWR.

Assim, dada a elevação dos valores de corte ao longo do tempo, determinados endogenamente a partir da estimação dos modelos GWR, associada ao efeito pouco relevante do espraiamento para novas áreas da região metropolitana, é possível caracterizar o caminho espacial do emprego na RMSP por dois movimentos simultâneos, a saber: consolidação do adensamento e descentralização concentrada do emprego. Claramente, o efeito do adensamento domina sobre o efeito da descentralização, tanto na capital como nos demais municípios, e por esse motivo denomina-se como uma desconcentração concentrada. Como discutido, a descentralização (novas áreas para aninhar o emprego formal) não altera a característica concentradora de empregos das centralidades, tendo pouco impacto na RMSP como um todo. Entretanto, em uma avaliação mais focalizada e restrita ao município de São Paulo, podem-se encontrar indícios de descentralização. Isto é, empregos formais se direcionam às áreas menos aglomeradas, todavia não distantes das áreas com concentração expressiva, salvo a concentração em Itaquera, e isso com valor de corte mais flexível. Tal padrão pode estar relacionado à necessidade de manter os retornos da aglomeração, mantendo o padrão clássico exposto pelos modelos teóricos de Economia Urbana, onde as firmas procuram lugares com menores custos, mas próximos dos lugares centrais principais.

Figura 3. 11 – Subcentralidades da RMSP (ano-base 2002 – quantil 99%)

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR.

Figura 3. 12 – Subcentralidades da RMSP ano-base 2002 – quantil 95%

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR.

Figura 3. 13 – Subcentralidades da RMSP (ano-base 2002 – quantil 90%)

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR.

Figura 3. 14 – Distribuição Espacial dos SBD por quantil

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR.

Figura 3. 15 – Subcentralidades da RMSP (ano-base 2008 – quantil 99%)

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR

Figura 3. 16 – Subcentralidades da RMSP (ano-base 2008 – quantil 95%)

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR.

Figura 3. 17 – Subcentralidades da RMSP (ano-base 2008 – quantil 90%)

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR.

Figura 3. 18 – Distribuição espacial dos SBD por quantil

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR.

Figura 3. 19 – Subcentralidades da RMSP ano-base 2014 – quantil 99%

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR

Figura 3. 20 – Subcentralidades da RMSP ano-base 2014 – quantil 95%

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR.

Figura 3. 21 – Subcentralidades da RMSP ano-base 2014 – quantil 90%

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR.

Figura 3. 22 – Distribuição Espacial dos SBD por quantil

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR.

3.3.3. Caracterização dos Subcentros da RMSP

Embora a identificação das áreas sub-centrais seja feita com base nas médias espaciais do emprego, os SBD possuem características e dinâmicas diferenciadas, a depender das atividades que desenvolvem, tanto entre quanto intra-SBD. Portanto, nessa subseção são caracterizados os SBD em função da concentração de trabalhadores por setores de atividades econômicas, gênero, grau de instrução, faixa de remuneração, tipo de contrato de trabalho e experiência. E, seja SBD^o o SBD identificado segundo a ordem do quantil o , podendo ser $Q99$, $Q95$ e $Q90$.

A divisão espacial dos SBD segue a mesma delimitação utilizada acima, exceto para o município de São Paulo. Buscando capturar as diferentes dinâmicas do mercado de trabalho intra-SBD do MSP, foram adotadas subdivisões dessa centralidade de emprego. Tal decisão deriva do papel central que a capital desempenha na concentração de emprego, como demonstrado pelos resultados, pela elevada participação do emprego alocado na capital (Tabela B.3) e pela extensão territorial das áreas identificadas. Essas características do SBD-SAO, por consequência, agregam relevantes diferenças intra-SBD, motivando uma avaliação desagregada. Para melhor compreensão, a nomenclatura dos SBD-SAO considerará a divisão por distritos do MSP, os quais aninham os empregos. Assim, os SBD-SAO serão renomeados como $SBD + sigla do distrito$. As áreas das células não respeitam a divisão administrativa pré-definida, portanto, as células que se localizam entre dois ou mais distritos receberão a *sigla do distrito* em que o centroide da célula estiver localizado. Por exemplo, células localizadas do SBD-SAO localizadas entre os distritos Sé e República serão consideradas como SBD-REP se o centroide estiver localizado no distrito REP.

A Tabela 3.6 apresenta as estatísticas descritivas para total de empregos, firmas e massa salarial concentrados em cada SBD considerando os três valores de cortes propostos. Em relação aos SBD^{Q99} , identificados com o corte mais restritivo, o total de empregos soma 934.834, distribuídos entre 71.544 firmas, em área de 91 km² no ano de 2002. Em 2008, o número total de empregos se eleva em 54,7% (1.446.631 postos de emprego), alocados em 83.519 firmas (aumento percentual de 16,7% em relação a 2002). Em relação ao período seguinte, a taxa de crescimento do emprego acumulado nos SBD^{Q99} foi de 16,4% (1.684.185 postos de emprego) em relação a 2008. Como demonstrado na Tabela 3.6, o máximo de postos de empregos em uma célula é de 139.220 em 2002, alcançando as marcas de 228.516 em 2008 e 252.890 em 2014. Esses resultados

correspondem ao SBD-JDP, o qual apresenta a taxa de crescimento mais elevada quando comparada à taxa de crescimento do total dos postos de emprego conjuntos dos SBD^{Q99} . Em relação ao SBD com menor concentração de emprego, os resultados se alteram intertemporalmente. Em 2002, o total mínimo de emprego corresponde ao SBD-VPR (Vila Prudente), com 3.467 postos de empregos distribuídos entre 181 firmas. O SBD-VPR deixa de elencar nessa categoria nos anos de 2008 e 2014, apresentando o SBD-CMB (Cambuci) com 4.551 postos de empregos alocados em 632 firmas em 2008 e 6.127 empregos e 648 firmas em 2014 nessa categoria. Para as demais posições dos SBD^{Q99} , ver Tabela E.1.

Para os demais quantis são considerados apenas as células que emergem a partir do relaxamento do valor de corte, excetuando todas as células identificadas pelo valor de corte mais restrito, ou seja, apenas células incrementais³⁰. A Tabela 3.6 apresenta a estatística descritiva para as três categorias de SBD.

Considerando as áreas definidas como SBD^{Q95} , a extensão da área é de 363 Km², concentrando 1.149.074 postos de emprego em 2002, 1.566.672 em 2008 e 1.875.782 em 2014. Ao considerar a média de empregos, os SBD^{Q95} apresentam valores menores relativamente ao valor calculado na categoria SBD^{Q99} . No interregno 2002-2008, observa-se que a taxa de crescimento do total de empregos (36,3%) nos SBD^{Q95} é relativamente menor quando comparada à mesma taxa nos SBD^{Q99} (54,7%). Esse resultado sugere um crescimento concentrado dos empregos, localizados nas áreas identificadas com valores de corte mais restritivos. No período 2008-2014, a taxa de crescimento do emprego é de 19,7% nos SBD^{Q95} , sendo três pontos percentuais acima da taxa de crescimento nos SBD^{Q99} para o mesmo interregno.

Entre as áreas que compõem os SBD^{Q95} , os máximos correspondem ao SBD-SBC, aninhando 110.017, 166.309 e 136.459 emprego numa área de 36, 40 e 30 km² nos anos 2002, 2008 e 2014, respectivamente. Em relação aos valores mínimos, as áreas são alteradas em cada ano avaliado. Em 2002, o total de emprego mínimo está localizado na Sub-Região Leste da RMSP, mais precisamente em Itaquaquecetuba (SBD-IQQ). Em 2008, o SBD-SDO representa o limite inferior da distribuição de empregos, localizado na Zona Norte 2 do MSP, enquanto o SBD-IQQ deixa de

³⁰ Por exemplo, no SBD^{Q95} contém exclusivamente células identificadas com valor de corte igual ao quantil 95 exceto as células que compõem o SBD^{Q99} e no SBD^{Q90} contém as células identificadas pelo quantil 90 menos as células que compõem os SBD^{Q99} e SBD^{Q95} .

elencar como SBD^{Q95} . Em 2014, o SBD-SDO (São Domingos) se consolida como área com importância relativa na concentração de empregos e o SBD-PEN (distrito da Penha, Zona Leste 1) surge como nova área de concentração de negócios, devido ao menor crescimento relativo das demais regiões, especificamente, as regiões fora da capital. Para as demais posições dos SBD, ver Tabela E.2.

Tabela 3.6 – Total de empregos, firmas, massa salarial e área por quartil

Estatísticas	2002			2008			2014		
	Empregos	Firmas	Massa Salarial (R\$)*	Empregos	Firmas	Massa Salarial (R\$)*	Empregos	Firmas	Massa Salarial (R\$)*
<i>SBD^{Q99} – 91 células</i>									
Total	934.843	71.544	1.372.410	1.446.631	83.519	3.169.201	1.684.185	89.962	6.179.594
Média	35.956	2.752	52.785	53.579	3.093	117.378	67.367	3.598	247.184
Desv. Pad.	32.526	2.532	61.736	51.969	3.015	143.389	64.781	3.462	294.543
Mínimo	3.467	181	3.036	4.551	256	5.406	6.127	268	12.418
Máximo	139.220	9.991	242.061	228.516	11.690	534.020	252.890	12.985	1.017.398
<i>SBD^{Q95} – 363 células</i>									
Total	1.149.074	77.403	1.396.829	1.566.762	91.767	2.762.459	1.875.782	113.568	4.771.076
Média	19.476	1.312	23.675	27.487	1.610	48.464	31.793	1.925	80.866
Desv. Pad.	19.322	1.270	26.354	28.331	1.506	57.988	28.590	1.804	73.435
Mínimo	1.407	25	907	2.212	36	3.335	1.253	47	2.010
Máximo	110.017	5.673	145.122	166.309	7.216	363.859	136.459	7.176	353.959
<i>SBD^{Q90} – 454 células</i>									
Total	547.674	47.219	537.894	797.115	58.996	1.142.014	932.349	67.793	2.192.148
Média	7.502	647	7.368	11.722	868	16.794	13.711	997	32.237
Desv. Pad.	7.238	614	7.837	13.670	836	23.552	19.404	1.045	54.389
Mínimo	69	9	75	110	10	200	100	18	315
Máximo	42.470	2.431	42.970	88.362	4.392	142.203	131.097	5.855	324.281

Fonte: elaboração própria a partir dos microdados da RAIS. * Os valores estão a preços correntes e devem ser multiplicados por 1.000.

Em relação aos SBD^{Q90} , os empregos estão distribuídos em 454 Km², representando a menor densidade de emprego por área quando comparados aos demais SBD . Os SBD^{Q90} concentram 547.674 postos de emprego em 2002, 797.115 em 2008 e 932.349 em 2014. Em termos de taxa de crescimento, o total de empregos apresenta taxa de crescimento de 45,5% no primeiro interregno (2002-2008) maior do que os SBD^{Q95} , mas menor do que aqueles aninhados no SBD^{Q99} , como demonstrado na Tabela 3.7. No período entre 2008-2014, o total de emprego nos SBD^{Q90} cresceu à taxa de 17%, maior do que nas áreas do SBD^{Q99} e menor que no SBD^{Q95} . Os valores máximos de postos de emprego estão no SBD-GRU (Guarulhos), ao longo dos três períodos avaliados, enquanto o mínimo está no SBD-SAM (Santo Amaro), na Zona Sul 2 da capital.

A Tabela E.4 apresenta informações sobre o gênero dos trabalhadores nos SBD para os diferentes valores de corte. A maioria dos trabalhadores é composta por homens, exceto no ano de 2008, em todos os casos, mas a diferença se reduz ao longo do tempo. A diferença composicional entre os SBD^{Q99} , SBD^{Q95} e SBD^{Q90} não é elevada, como se observa na Tabela E.4. No ano de 2014, a parcela de homens dos SBD^{Q99} é 50%, enquanto nos SBD^{Q95} e SBD^{Q90} as parcelas são de 52% e 54%, respectivamente. Em 2008, a diferença entre SBD na parcela de homens no total de emprego se manteve praticamente inalterada, quando comparada ao período anterior, a diferença da média dos SBD^{Q99} para os SBD^{Q95} e para os SBD^{Q90} é de 5 pontos percentuais. É importante notar que tal diferença foi reduzindo ao longo dos anos, independente do valor de corte utilizado. A maior diferença pode ser observada nas áreas que compõem os SBD^{Q90} , enquanto nos SBD^{Q99} a composição é quase a mesma, considerando o ano de 2014 como referência.

Tabela 3.7 – Diferença do total de empregos, firmas, massa salarial e área por quantil

Estatísticas	2002-2008			2008-2014		
	Emprego	Firma	Massa Salarial (R\$)	Emprego	Firma	Massa Salarial (R\$)
<i>SBD^{Q99} – 91 células</i>						
Total	54,7%	130,9%	16,7%	16,4%	95,0%	7,7%
Média	49,0%	122,4%	12,4%	25,7%	110,6%	16,3%
Desvio Padrão	59,8%	132,3%	19,1%	24,7%	105,4%	14,8%
Mínimo	31,3%	78,1%	41,4%	34,6%	129,7%	4,7%
Máximo	64,1%	120,6%	17,0%	10,7%	90,5%	11,1%
<i>SBD^{Q95} – 363 células</i>						
Total	36,3%	97,8%	18,6%	19,7%	72,7%	23,8%
Média	41,1%	104,7%	22,7%	15,7%	66,9%	19,6%
Desvio Padrão	46,6%	120,0%	18,6%	0,9%	26,6%	19,8%
Mínimo	57,2%	267,9%	44,0%	-43,4%	-39,7%	30,6%
Máximo	51,2%	150,7%	27,2%	-17,9%	-2,7%	-0,6%
<i>SBD^{Q90} – 454 células</i>						
Total	45,5%	112,3%	24,9%	17,0%	92,0%	14,9%
Média	56,2%	127,9%	34,1%	17,0%	92,0%	14,9%
Desvio Padrão	88,9%	200,5%	36,1%	41,9%	130,9%	25,0%
Mínimo	59,4%	168,8%	11,1%	-9,1%	57,3%	80,0%
Máximo	108,1%	230,9%	80,7%	48,4%	128,0%	33,3%

Fonte: elaboração própria a partir dos microdados da RAIS.

A participação relativa dos trabalhadores por faixas de anos trabalhando na mesma firma (experiência) pode ser observada na Tabela E.5. Em 2014 mais de um terço dos trabalhadores estavam na faixa com até um ano de experiência, e cerca de 13% estavam na faixa com nove anos de experiência, para quaisquer valores de corte utilizados. No grupo SBD^{Q99} , o SBD-BAR e SBD-SCS são as regiões com maior percentual de trabalhadores nessa categoria (44% do total de trabalhadores), em 2002. Nos períodos de 2008 e 2014, os valores de máximo são observados nos SBD-SAM, que tem 49,7% dos seus trabalhadores com até 1 ano de experiência, e SBD-BAR, com

Ao considerar os grupos SBD^{Q95} e SBD^{Q90}, a diferença entre valores de máximo e mínimos dos trabalhadores na categoria com até um ano de experiência é mais elevada quando comparado ao grupo SBD^{Q99}. Essas cifras revelam a heterogeneidade nessas regiões quanto à rotatividade dos trabalhadores. Ademais, é possível encontrar, inclusive, heterogeneidade quanto à permanência do trabalhador na mesma firma dentro do próprio distrito, como no caso do distrito de Santo Amaro. Nesse distrito, há ao menos uma célula que concentra elevado número de trabalhadores na primeira faixa de experiência, compondo o SBD^{Q99}, e outra área com concentração relativamente menor, que compõe o SBD^{Q90}.

Diferentemente da heterogeneidade existente em relação à rotatividade do trabalhador, intra e entre grupos de SBD, não se notam diferenças relevantes quanto ao tipo contrato de trabalho firmado. A maior parte dos trabalhadores firmaram contratos com remuneração mensal. Os subcentros que compõem o grupo SBD^{Q99} são os que apresentam a maior parcela dos trabalhadores nessa categoria, crescendo intertemporalmente: 93,4% (2002), 95,1% (2008) e 96,5% (2014). A menor composição relativa ao total de trabalhadores é observada no grupo SBD^{Q95}, onde 84,0% dos trabalhadores compunham a categoria de recebimento mensal em 2002; em 2014, essa parcela elevou-se para 91,6%. O comportamento crescente também é observado nas áreas do SBD^{Q90}. Em 2002, 85,3% dos trabalhadores eram remunerados mensalmente; nos anos de 2008 e 2014 essa relação se alterou para 87,3% e 90,5%. Esse crescimento deriva da redução do tipo de contrato com remuneração por hora. Essa categoria apresenta um comportamento declinante para qualquer valor de corte utilizado. Todavia, existem subcentros com parcela relevante de trabalhadores que firmaram contratos de recebimento por hora. Em 2002, as regiões de SBD-SCS (20,9%, no grupo SBD^{Q99}), SBD-SNO (47,1%, no grupo SBD^{Q95}) e SBD-ERM (46,7%, no grupo SBD^{Q90}). Em 2014, regiões como SBD-BRS (6,9%, no grupo SBD^{Q99}), SBD-STP (24,9%, no grupo SBD^{Q95}) e SBD-JAC (31,1%, no grupo SBD^{Q90}). Tal característica apresenta forte relação com as atividades econômica desenvolvidas nessas áreas.

Os gráficos 3.3 e 3.4 mostram a parcela de trabalhadores em cada uma das categorias de tipo contrato salarial para o SBD^{Q99}, considerando os anos 2002 e 2014. As menores parcelas no total de trabalhadores para o tipo de contrato mensal encontram-se no SBD-SCS (78%, em 2002) e no SBD-LAP (89%, em 2014). Os SBD-JDP (97%, em 2002) e SBD-CMB (99%, em 2014) elencam com as maiores parcelas de contrato de trabalho mensal.

Gráfico 3.3 – Parcela de trabalhadores por tipo de contrato salarial nos SBD (base 2002), por distrito

Fonte: Elaboração própria a partir dos dados da RAIS (2002).

Gráfico 3.4 – Parcela de trabalhadores por tipo de contrato salarial nos SBD (base 2014), por distrito

Fonte: Elaboração própria a partir dos dados da RAIS (2014).

Embora os tipos de remuneração salarial apresentem baixo desvio à média da parcela de trabalhadores no agregado, tanto entre os grupos de SBD quanto intertemporalmente, o mesmo não é observado quando se divide a variável por categorias de salário mínimo. Considerando o total de trabalhadores dos grupos SBD^{Q99}, SBD^{Q95} e SBD^{Q90} divididos em nove faixas de remuneração, observa-se, a partir da Tabela E.6 que a maior parcela dos trabalhadores formais recebe salários que variam entre 2 e 3 salários mínimos (SM3) no ano de 2002 para os três grupos de SBD. Essa concentração se altera no ano de 2014, com a maior parcela de trabalhadores na categoria que varia entre 1 e 2 salários mínimos (SM2). Quanto à parcela de trabalhadores recebendo salários acima de 7 salários mínimos, por exemplo, os subcentros do grupo SBD^{Q99} agregam a maior parcela desses trabalhadores. Os gráficos 3.4 e 3.5 mostram as decomposições para os anos de 2002 e 2014. Esses resultados de remuneração refletem, em alguma medida, a desigualdade espacial da remuneração entre os diferentes grupos de SBD, que pode ser explicada tanto pelo nível educacional dos trabalhadores quanto pelos setores de atividades.

Gráfico 3.5 – Parcela de trabalhadores por salário mínimo nos SBD (ano 2002), por distrito

Fonte: Elaboração própria a partir dos dados da RAIS (2002). Nota: SM1 (menor que 1 SM); SM2 (entre 1 e 2 SM); SM3 (entre 2 e 3 SM); SM4 (entre 3 e 4 SM); SM5 (entre 4 e 5 SM); SM6 (entre 5 e 7 SM); SM7 (entre 7 e 10 SM); SM8 (entre 10 e 15 SM); SM9 (maior que 15 SM).

Gráfico 3.6 – Parcela de trabalhadores por salário mínimo nos SBD (ano 2014), por distrito

Fonte: Elaboração própria a partir dos dados da RAIS (2014). Nota: SM1 (menor que 1 SM); SM2 (entre 1 e 2 SM); SM3 (entre 2 e 3 SM); SM4 (entre 3 e 4 SM); SM5 (entre 4 e 5 SM); SM6 (entre 5 e 7 SM); SM7 (entre 7 e 10 SM); SM8 (entre 10 e 15 SM); SM9 (maior que 15 SM).

A Tabela E.8 apresenta informações sobre o grau de instrução dos trabalhadores para os grupos de SBD. Utilizando a categorização disponibilizada pelos microdados da RAIS foi possível dividir o grau de instrução dos trabalhadores em nove categorias, como exposto acima. Em linhas gerais, as parcelas de trabalhadores com grau de escolaridade nas categorias iniciais foram reduzidas quando se comparam os dados de 2002 aos anos de 2008 e 2014 nos três grupos de SBD, enquanto as parcelas de trabalhadores com ensino médio completo cresceram substancialmente, sobretudo no interregno 2002-2008.

Os SBD^{Q99} apresenta maior concentração de trabalhadores com ensino superior incompleto ou com graduação, mestrado e/ou doutorado completo. Em 2002, 22% do total de trabalhadores nos SBD localizados nos municípios de Barueri, São Paulo e São Caetano do Sul compunham essas duas categorias. Nos anos de 2008 e 2014, as parcelas de trabalhadores nessas duas categorias alcançaram as marcas de 27% e 33%, respectivamente. Nos outros dois grupos, as parcelas de trabalhadores mais qualificados são relativamente menores e o crescimento da composição de trabalhadores nessas categorias é inferior ao grupo anterior. Em números, os SBD^{Q95}, em 2002, tem 16% do total dos trabalhadores com o nível educacional maior que o ensino médio. Em 2008 e 2014 as parcelas alcançaram, para as mesmas categorias, 20% e 23%, respectivamente. Os

reduzindo a participação nos demais setores. As atividades relacionadas às instituições de crédito perdem participação relativa e não elencam entre os principais setores de atividade, dando espaço às atividades relacionadas ao setor de comércio atacadista. Todavia, em 2014 voltam a ganhar participação relativa, elencando entre as principais atividades. Nesse ano, o total de empregos agregado nesses cinco principais setores se eleva e alcança a marca de 76% do emprego total. Os gráficos 3.9 e 3.10 apresentam os principais setores que compõem o grupo SBD^{Q99} para os anos de 2002 e 2014.

Gráfico 3.9 – Parcela de trabalhadores por setor nos SBD (top 5)

Fonte: Elaboração própria a partir dos dados da RAIS (2002). Nota: S1 (comércio, administração de imóveis, valores mobiliários); S3 (alimentação e bebidas); S4 (alojamento e alimentação); S6 (comércio atacadista); S7 (Comércio varejista); S9 (ensino); S12 (indústria mecânica); S13 (indústria metalúrgica); S14 (indústria química); S15 (indústria têxtil); S16 (instituições financeiras); S17 (madeira e mobiliário); S18 (indústria do material de transporte); S19 (serviços médicos, odontológicos e veterinários); S22 (transportes e comunicações).

Gráfico 3.10 – Parcela de trabalhadores por setor nos SBD (top 5)

Fonte: Elaboração própria a partir dos dados da RAIS (2014). Nota: S1 (comércio, administração de imóveis, valores mobiliários); S3 (alimentação e bebidas); S4 (alojamento e alimentação); S6 (comércio atacadista); S7 (Comércio varejista); S9 (ensino); S12 (indústria mecânica); S13 (indústria metalúrgica); S14 (indústria química); S15 (indústria têxtil); S16 (instituições financeiras); S17 (madeira e mobiliário); S18 (indústria do material de transporte); S19 (serviços médicos, odontológicos e veterinários); S22 (transportes e comunicações).

Em relação ao SBD^{Q95}, ano de 2002, a característica setorial é similar ao SBD^{Q99}, exceto por apresentar entre os cinco setores mais relevantes as atividades de comércio atacadista (S4) e indústria química (S14). Em 2008, os mesmos setores são elencados, ampliando minimamente a parcela relativa de trabalhadores em cada setor – ver Tabela E.9. Em 2014, o setor de serviços médicos, odontológicos e veterinários (S19) passa a elencar entre os cinco principais, e o setor S14 deixa o *rank*. Em relação aos SBD^{Q90}, as atividades estão relacionadas ao setor de comércio

varejista, saindo do patamar de 23,3% do total de trabalhadores em 2002 e alcança as marcas de 23,5% em 2008 e 24,9% em 2014. Os demais setores que elencam no *rank* são: S1, S4, S14 e S19 (2002); S1, S4, S6, S7, S14 (2008) e S1, S4, S14 e S19 (2014).

A partir das avaliações agregadas por grupos de SBD é possível notar desvios das parcelas de trabalhadores formais para as características avaliadas, que apontam à não homogeneidade intra-SBD. Observando esses dois últimos indicadores, observa-se elevada discrepância entre SBD do mesmo grupo, havendo SBD cuja concentração no setor S1 é quase nula, por exemplo. Em relação à remuneração e ao grau de instrução mais elevado, os trabalhadores formais localizados nos SBD^{Q99} concentram maior parcela de trabalhadores. Essas diferenças espaciais dos SBD tornam-se mais relevantes devido à sua concentração numa área de 91 km² e densidade média de emprego variando de 10 mil trabalhadores/km² em 2002 para 18 mil trabalhadores/km² em 2014, restritos a três municípios dos 39 que compõem a RMSP. Nesse sentido, justifica-se uma análise mais desagregada desse grupo de SBD.

Os diferenciais de concentração de renda por total de trabalhadores entre os subcentros estão expostos nos Gráficos 3.5 e 3.6. SBD-MOR (39%, em 2002), SBD-IBI (25%, em 2002) e SBD-JDP (23%, em 2002) são os SBD concentram maior parcela de trabalhadores com salários mensais maior ou igual a dez salários mínimos e localizados no SBD-SAO. Essas áreas apresentam as mais elevadas parcelas de trabalhadores com no mínimo ensino superior completo (classe G8, gráficos 3.7), enquanto, em média, 10% dos seus trabalhadores estão na categoria de ensino superior incompleto, apontando para a qualificação dos trabalhadores nessas localidades. Em 2008, os SBD-IBI (21%), SBD-MOR (29%) e SBD-VAN (43%) aparecem novamente como os principais centros de negócios com maior remuneração para seus trabalhadores e localizados na capital. Esses distritos apresentam as maiores participações de trabalhadores com ensino superior completo, mestrado e/ou doutorado, alcançando a parcela de 64% no SBD-VAN, 42% no SBD-MOR e 33% no SBD-IBI. No ano de 2014, os mesmos SBD estão entre os que melhor remuneraram os trabalhadores, mesmo em face de redução das parcelas quando comparadas ao ano de 2008: SBD-VAN (37%), SBD-MOR (25%) SBD-IBI (18%).

Os SBD com menores remunerações estão concentrados nos SBD-BAR e SBD-SCS, que concentram, respectivamente, 26% e 35% dos trabalhadores formais recebendo entre 1 e 2 salários

mínimos em 2002. Em 2008, os SBD-SCS e SBD-MOO (distrito da Mooca) passam a aparecer nessa categoria, com 67% e 61% de sua mão de obra. No ano de 2014, os SBD que pior remuneravam seus trabalhadores foram SBD-REP e SBD-SCE com 57% e 55%, respectivamente. Esses centros de negócios com menor remuneração salarial concentram trabalhadores com baixa qualificação. Em 2002, os SBD-BAR e SBD-SCS apresentam 22% e 36% dos seus trabalhadores com ensino fundamental incompleto (soma das parcelas das categorias G1 a G4). Enquanto, no mesmo ano, o SBD-MOR tem apenas 10% de sua mão de obra formal com o mesmo grau de instrução. Os SBD-SCS, em 2008, eleva a participação de trabalhadores pouco qualificados em termo de anos de estudo para 25% e o SBD-MOO apresenta a maior concentração entre todos os SBD identificados, com 31% do total de trabalhadores na mesma categoria. Enquanto apenas 10% e 4,5% dos funcionários desses SBD têm ensino superior completo, mestrado e/ou doutorado. SBD-REP e SBD-SCE, os dois piores SBD em termos de remuneração em 2014, apresentam parcelas de 7% e 8% dos trabalhadores na mesma faixa educacional que os SBD anteriormente analisados.

Como discutido acima, houve elevação do nível educacional dos trabalhadores dos SBD da RMSP, em todos os três grupos de SBD, com redução significativa da parcela de trabalhadores com até o ensino fundamental incompleto. Todavia, a elevação do nível de escolarização dos trabalhadores não foi capaz de alterar na mesma proporção a parcela de trabalhadores recebendo acima de 2 salários mínimos. Essas características divergentes dos SBD e suas dinâmicas, em alguma medida, estão relacionadas às atividades desenvolvidas em cada centralidade de negócios.

Entre os SBD com maior parcela de trabalhadores recebendo acima de 10 salários mínimos, em 2002, as atividades desenvolvidas estão essencialmente relacionadas ao setor de serviços, especialmente nas atividades de comercialização e administração de imóveis, valores mobiliários e serviços técnicos (S1), comércio varejista (S7), serviços de alojamento, alimentação, reparação, manutenção e redação (S4), e instituições de crédito, seguros e capitalização (S16). Serviços de transporte e comunicação (S22) se destacam apenas nos SBD localizados nos distritos de Itaim Bibi e Morumbi, e serviços médicos, odontológicos e veterinários (S19), no SBD localizado no distrito da Saúde. Em 2008, os SBD identificados permanecem apresentando as cinco atividades supracitadas e empregando mais que 75% do total dos trabalhadores formais. Ainda em 2008, o SBD localizado no distrito de Vila Andrade aparece como o que melhor remunera os trabalhadores. As estruturas setoriais do SBD-IBI permanecem inalteradas, exceto no SBD-MOR e SBD-VAN,

onde o setor de comércio atacadista e instituições de crédito, seguros e capitalização passam a se destacar entre as cinco principais atividades do centro de negócios.

Entre os SBD que apresentam maior parcela de trabalhadores recebendo até 2 salários mínimos, as posições foram preenchidas por SBD diferentes para cada ano avaliado, demonstrando maior dinamismo quando comparado às parcelas de trabalhadores recebendo acima de 10 salários mínimos. Os SBD localizados em Barueri e São Caetano do Sul têm suas economias locais concentradas nos setores S1, S4, S7, S19 e S22, muito semelhante aos SBD que apresentam maiores concentrações de trabalhadores ganhando salários acima de 10 salários mínimos. Esses resultados sugerem, em alguma medida, que os serviços prestados nos SBD da capital são mais produtivos, permitindo ofertarem maior remuneração, quando comparados aos SBD-BAR e SBD-SCS. Em 2008, o SBD-MOO, SBD-REP e SBD-SCE são os que pior remuneraram seus trabalhadores. A estrutura setorial do SBD-SCE é similar ao SBD-MOO, exceto pela parcela de 45% dos trabalhadores alocados na indústria têxtil.

Os resultados derivados dessa caracterização dos centros de negócios na RMSP revelam a heterogeneidade do mercado de trabalho da região. Mesmo em face de certa homogeneidade da economia local dos SBD, é possível identificar elevadas diferenças salariais pagas aos trabalhadores e no que diz respeito à “qualidade” da mão de obra demandada em cada centro de negócios local. Pode-se também notar que as aglomerações de empregos, aqui denominadas SBD, estão majoritariamente concentradas no setor de serviços, tendo uma baixa parcela de trabalhadores em indústrias. Os únicos SBD que têm na sua composição setorial atividades relacionadas à indústria estão localizadas no distrito de Vila Andrade (8% do total dos trabalhadores na indústria química) e nos distritos de Bom Retiro (14%), Brás (7%), Pari (30%), Santa Cecília (45%) e Santana (2%) na indústria têxtil, no ano de 2014.

3.3.4. Teste de Sensibilidade

A comparação entre áreas centrais é informativa e a análise do espriamento intertemporal da concentração de empregos fornece insumo importante para entender a variação da aglomeração, em termos de emprego, e sua dispersão no espaço intra-urbano. Todavia, esses resultados podem

ser sensíveis à unidade geográfica adotada na análise. Essa questão é trazida à literatura como o Problema de Unidade de Área Modificável (MAUP – *Modifiable Areal Unit Problem*).

O MAUP aponta que as correlações entre variáveis podem divergir em função da escala e formato das áreas (OPENSHAW; TAYLOR, 1979). A escala diz respeito à resolução espacial (ou tamanho) das unidades que aninham as informações, enquanto o formato se refere ao desenho dos limites das áreas tomadas para uma dada resolução espacial (ANDERSSON; KLAESSON; LARSSON, 2016). O debate sobre MAUP tem sido extensivamente discutido na literatura de Geografia, enquanto em Economia Espacial pouco tem sido abordado nas análises (BURGER, VAN OORT E VAN DER KNAAP, 2010; BRIANT, COMBES E LAFOURCADE, 2010).

Para averiguar se os resultados podem ser influenciados pelo tamanho escolhido para as áreas, realizaram-se testes de sensibilidade que consistem na ampliação das áreas das células. Áreas de 2 km² e 4 km² foram construídas e regressões ponderadas geograficamente foram reestimadas. Considerando as células de 2 km² de área, criou-se um grid formado por 4.611 células. No caso de células com 4 km² de área, 2.342 células formam o grid. Para o tamanho da banda, novamente o critério de informação AIC para diferentes funções kernel foi empregado. Em todos os anos, para ambos tamanhos de células, a função *kernel* tricúbica apresenta o menor AIC. Os testes e seus respectivos valores podem ser observados na Tabela 3.8.

Tabela 3. 8 – Critério de Informação AIC e Tamanho das Bandas

Função Kernel	Critério de Informação AIC					
	2002		2008		2014	
	AIC	Banda (Km)	AIC	Banda (km)	AIC	Banda (Km)
Células com 2 km ²						
Gaussiana	81.793	1,2335	85.192	1,2685	86.573	1,2592
Biquadrada	81.685	2,9479	85.114	3,0281	86.483	3,0543
Tricúbica	81.605	2,8863	85.036	2,9182	86.403	2,9183
Células com 4 km ²						
Gaussiana	44.093	1,6664	45.901	1,6647	46.664	1,6221
Biquadrada	43.978	4,2328	45.781	4,2479	46.539	4,0656
Tricúbica	43.940	4,2479	45.742	4,2754	46.492	4,0206

Fonte: elaboração própria, a partir dos dados da RAIS (2002; 2008; 2014).

No grid formado por células com áreas de 1 km², cada regressão estimada, em média, considerava uma área de 25 km², dado uma janela de 2 km em média. Portanto, ao dobrar a área, cada regressão GWR considerará o dobro de áreas, comparada ao parâmetro anterior. Isto é, 50 km² e 100 km², embora mantenha fixa o número de células para cada regressão.

Tabela 3. 9 – Teste de sensibilidade para regressões ponderadas geograficamente

	Células com 2 Km ²			Células com 4 Km ²		
	2002	2008	2014	2002	2008	2014
	I	II	III	IV	V	VI
Função Kernel	Tricúbica	Tricúbica	Tricúbica	Tricúbica	Tricúbica	Tricúbica
Min	0,00	0,00	0,00	0,00	0,00	0,00
1º Quartil	0,00	0,00	0,02	0,35	0,51	0,83
Mediana	5,18	3,47	12,78	21,62	36,07	42,66
3º Quartil	239,97	353,88	414,92	572,07	740,72	972,64
Max	35.652,49	55.238,39	64.651,00	56.490,29	86.777,00	104.840,00
Global	668,85	975,69	1.159,00	1.316,90	1.921,00	2.281,80
AIC	81.604,9	85.036,2	86.403,2	43.939,8	45.742,3	46.491,6
Quase-global R ²	0,7757	0,7833	0,7892	0,7990	0,8072	0,8208
Num. Obs.	4.611	4.611	4.611	2.342	2.342	2.342

Fonte: elaboração própria, a partir dos dados da RAIS (2002; 2008; 2014).

Como nos resultados derivados de áreas menores, as médias espaciais do 1º quartil e mediana não se alteram significativamente. Os resultados do 3º quartil apresentam maior concentração em relação às células menores, em termos de área. Como esperado, os resultados apresentam maior concentração de trabalhadores, mas a densidade da média espacial estimada é muito próxima das estimações anteriores. Dividindo o resultado do 3º Quartil (Tabela 3.9, coluna I) pela área total, isto é, 2 km², encontra-se o número de 120 trabalhadores por km², resultado próximo ao encontrado quando estimado com área da célula de 1 km² (97 trabalhadores/km², Tabela 3.3, coluna I). Nesse sentido, o MAUP aparenta ser pouco expressivo no desvio dos resultados estimados.

Tabela 3. 10 – Valores de Corte e total de células identificadas como SBD

Quartil	2002		2008		2014	
	Valor de Corte	SBD*	Valor de Corte	SBD*	Valor de Corte	SBD*
Células com 2 km ²						
99	10.544	47	15.697	47	18.914	47
95	3.668	231	5.397	231	6.351	231
90	1.628	462	2.292	462	2.806	462
Células com 4 km ²						
99	18.861	24	28.890	24	33.843	24
95	7.482	118	11.410	118	12.572	118
90	3.310	235	4.677	235	5.622	235

Fonte: elaboração própria, a partir dos dados da RAIS (2002; 2008; 2014). SBD*: refere-se ao total de células do *grid* identificadas a 1% de significância estatística. As mesmas células são identificadas ao relaxar a significância estatística.

Os valores de corte utilizados no processo de identificação também variam, como observado na Tabela 3.10. Os cortes são mais elevados para todos os anos e para as duas composições de áreas utilizadas. Quanto ao grid com células de 2 km², são identificadas, em termo de quantidade, pouco mais da metade das células identificadas utilizando células com 1 km² de área, independentemente do valor de corte utilizado. Para o grid com células de 4 km², o total de células identificadas é quase um terço do total de células identificadas pelo modelo *benchmark*.

O próximo passo é avaliar o efeito do MAUP nos resultados encontrados face aos derivados dos modelos de referência. Os modelos com áreas menores identificam quantidades menores de células como SBD. A Tabela 3.11 expõe a aderência dos novos SBD em relação aos SBD *benchmark*.

Tabela 3. 11 – Matriz de aderência dos SBD por diferentes tamanho de área das células

	2002			2008			2014		
	$SBD_{1Km^2}^{Q99}$	$SBD_{2Km^2}^{Q99}$	$SBD_{4Km^2}^{Q99}$	$SBD_{1Km^2}^{Q95}$	$SBD_{2Km^2}^{Q95}$	$SBD_{4Km^2}^{Q95}$	$SBD_{1Km^2}^{Q90}$	$SBD_{2Km^2}^{Q90}$	$SBD_{4Km^2}^{Q90}$
$SBD_{1Km^2}^{Q99}$	100,0%	-	-	100,0%	-	-	100,0%	-	-
$SBD_{2Km^2}^{Q99}$	75,0%	100,0%	-	77,1%	100,0%	-	76,9%	100,0%	-
$SBD_{4Km^2}^{Q99}$	63,8%	78,3%	100,0%	72,5%	86,5%	100,0%	64,9%	82,9%	100,0%
$SBD_{1Km^2}^{Q95}$	100,0%	-	-	100,0%	-	-	100,0%	-	-
$SBD_{2Km^2}^{Q95}$	86,1%	100,0%	-	86,1%	100,0%	-	83,9%	100,0%	-
$SBD_{4Km^2}^{Q95}$	79,1%	85,7%	100,0%	80,4%	86,8%	100,0%	80,2%	85,4%	100,0%
$SBD_{1Km^2}^{Q90}$	100,0%	-	-	100,0%	-	-	100,0%	-	-
$SBD_{2Km^2}^{Q90}$	88,7%	100,0%	-	89,3%	100,0%	-	87,9%	100,0%	-
$SBD_{4Km^2}^{Q90}$	81,5%	87,1%	100,0%	82,2%	87,8%	100,0%	82,7%	87,5%	100,0%

Fonte: elaboração própria a partir dos resultados das estimações GWR.

É possível observar para os SBD^{Q99} que a divergência dos resultados encontrados se reduz ao longo dos anos. Em 2002, 21 células que eram identificadas como SBD no modelo *benchmark* deixam de apresentar tal característica ao ampliar a área da célula para 2 km²; enquanto ao ampliar para 4 km², 31 células perdem tal característica. Por outro lado, 25 novas células ganham relevância em termos de concentração de emprego, no grid com células de 2 km²; enquanto a 4 km², 36 novas células são identificadas como áreas centrais. Esses valores são reduzidos para 29 células (a 2 km²) e 35 células (a 4 km²) em 2014, como pode ser observado na Tabela E.10. Em termos de aderência, as novas áreas centrais representam 75% das áreas de negócios identificadas com células de 1 km². Comparando as áreas centrais identificadas com áreas de 4 km² com as áreas identificadas pela estrutura do grid *benchmark*, a sobreposição espacial das áreas é de 63%. As taxas de aderência se alteram ao longo do tempo, alcançando valores maiores, relativamente ao ano de 2002.

As figuras 3.23 a 3.28 demonstram as novas localizações para as áreas na categoria SBD^{Q99} , considerando as novas dimensões das células. No ano de 2002, uma redução da área dos SBD localizados em Barueri e em São Caetano do Sul é observada e o SBD do MSP é o que apresenta maior aderência em relação ao modelo *benchmark*. Ao expandir a área das células para 4 km²,

SBD-SCS e SBD-BAR não são identificados. Nesse caso, forma-se apenas um mega SBD na capital e, em relação aos SBD identificados no modelo *benchmark*, as áreas identificadas como centro de negócios localizadas na Zona Sul 1 da cidade deixam de compor o SBD-SAO. Em 2008, em geral, com células dos grids expandidas, identifica-se apenas o subcentro no município de São Paulo. No ano de 2014, o SBD-BAR volta a ganhar participação relativa no total de empregos da RMSP. Todavia, ao ampliar a área da célula para 4 km², somente o SBD-SAO é identificado.

Os SBD fora da capital são sensíveis ao tamanho das áreas das células. Esses resultados derivam do padrão espacial dos empregos na capital, que são concentrados e cerceiam os grandes SBD. Assim, ampliar a área das células, implica em acréscimos maiores para essas localidades relativamente aos vizinhos dos SBD fora do MSP. Portanto, as médias espaciais são relativamente menores quando considerando células maiores nas regiões fora do MSP, devido à baixa extensão do deslocamento do emprego.

Figura 3. 23 – Distribuição espacial dos SBD por quantil – grid 2 km²

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR (2002).

Figura 3. 24 – Distribuição espacial dos SBD por quantil – grid 2 km²

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR (2008).

Figura 3. 25 – Distribuição espacial dos SBD por quantil – Grid 4 km²

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR (2002).

Figura 3.26 – Distribuição espacial dos SBD por quantil – Grid 4 km²

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR (2008).

Figura 3.27 – Distribuição espacial dos SBD por quantil – Grid 2 km²

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR (2014).

Figura 3.28 – Distribuição espacial dos SBD por quantil – Grid 4 km²

Fonte: Elaboração própria a partir dos resultados derivados dos modelos GWR (2014).

3.4. CONCLUSÃO

Este capítulo apresenta alguns passos adiante na discussão espacial do mercado de trabalho. Buscando uma avaliação mais focalizada da concentração espacial do emprego, o georreferenciamento da base de dados da RAIS constituiu um grande desafio e forneceu o insumo necessário para discutir sobre o espraiamento do emprego formal e privado na RMSP, identificação de subcentralidades e suas respectivas características. O resultado amplia o conhecimento que se tem da distribuição das atividades econômicas no espaço metropolitano e, por seu aspecto temporal, torna possível avaliar as mudanças ocorridas no período considerado.

Outro importante avanço está na proposta de uma metodologia capaz de identificar centralidades de emprego. A proposta baseada em Regressões Ponderadas Geograficamente e a utilização dos coeficientes estimados, estatisticamente significantivos, é superior às demais metodologias apontadas por superar a subjetividade, a necessidade de conhecimento prévio do objeto de estudo por parte do pesquisador e valores de corte fixos intemporalmente. A nova abordagem proposta pode ser vista como uma adequação dos valores de corte à dinâmica do mercado de trabalho, permitindo endogeneizar a métrica utilizada para identificar os SBD, embasadas na média espacial do emprego estimada e não nos resíduos estimados. Embora o pesquisador arbitre sobre o quantil escolhido, essa métrica é objetiva, passível de atualização temporal, em função dos dados, e aplicável em qualquer região estudada. Nesse sentido, a metodologia proposta demonstra-se como um avanço à abordagem proposta por Greene (1980), McDonald (1987) e Giuliano e Small (1991).

Sobre a aglomeração do emprego no mercado de trabalho da RMSP, considerando quantil 99, uma cordilheira é formada nas zonas centro, sul e oeste do município de São Paulo. Nos municípios de Barueri e São Caetano do Sul é possível observar SBD tão importantes quanto os elencados em São Paulo. A concentração de emprego se eleva e se propaga às margens dessas regiões a medida que o valor de corte é relaxado, o que demonstra que a expansão do emprego não é desconexa do contexto urbano previamente estabelecido.

Três SBD foram identificados sob um valor de corte mais restrito, a saber SBD-BAR, SBD-SCS e SBD-SAO. Embora pouca seja a heterogeneidade entre os SBD, diferenças intra-SBD podem ser observadas, sobretudo naquelas áreas que compõem o SBD-SAO. Diferenças em

termos salariais, educacionais e concentração de atividades setoriais revelam que o mercado de trabalho da RMSP demanda uma avaliação mais focalizada para maior efetividade de intervenções públicas, sejam elas em transporte, uso do solo, etc.

Quanto ao movimento espacial do emprego, os resultados encontrados apontam para dois movimentos do emprego na RMSP: consolidação do adensamento e descentralização concentrada do emprego, o que significa dizer que as forças de aglomeração são relevantes para as firmas localizadas na RM e justificam, por consequência, a nova posição espacial em aglomerações. Ao longo do período estudado, os empregos se espraíram pela região metropolitana, contudo a expansão se deu próxima das áreas com histórico de elevada concentração de emprego. Todavia, a taxa de crescimento do emprego é concentrada em pontos focais da RM. Nesse sentido, o efeito de deslocamento concentrado domina sobre o efeito de descentralização pulverizado, seja na capital ou nos demais municípios.

Em resumo, o movimento centrífugo do emprego é concentrado. Esse padrão de deslocamento em muito está correlacionado aos ganhos de aglomeração e custo de transporte. A aglomeração de trabalhadores fornece o ambiente para espraçamento de conhecimento, acesso ao mercado de trabalho e redução de custo fixo na oferta de bens públicos, como a oferta de meios de transporte de massa, refletindo sobre a remuneração salarial do trabalhador. Essa hipótese sobre a economia de aglomeração será testada no próximo capítulo, tomando como base uma abordagem econométrica com controles para o efeito da aglomeração sobre os salários dos trabalhadores.

4. EXTENSÃO ESPACIAL DA ECONOMIA DE AGLOMERAÇÃO E ABSORÇÃO DO PRÊMIO SALARIAL INTRA-URBANO

Face à limitada discussão sobre prêmio salarial na escala intra-urbana, esse capítulo estuda o impacto da aglomeração sobre o salário dos indivíduos na Região Metropolitana de São Paulo. Para uma análise mais focalizada, a aglomeração do emprego é agregada em áreas de 1 km². O objetivo principal nesse capítulo é estimar o efeito da aglomeração sobre o salário dos trabalhadores formais. Para tanto, duas estratégias empíricas são implementadas. A primeira abordagem utiliza defasagens espaciais da densidade de emprego e testa a hipótese de atenuação espacial da economia de aglomeração, buscando estimar se quanto mais distante a aglomeração de trabalhadores estiver do posto de trabalho do indivíduo menor é o efeito da economia de aglomeração sobre o salário. A segunda estratégia está embasada na redução e ampliação das células dos grids. Espera-se que quanto mais compacta e densa a área em que o trabalhador está localizado, maior seja a probabilidade do contato F2F (*face to face*) entre os trabalhadores e, então, maior o retorno sobre o salário. A abordagem econométrica utilizada para alcançar tais objetivos conecta o dinamismo da firma, o salário do trabalhador e a dotação local aos microdados da RAIS dispostos longitudinalmente, o que permite controlar a auto seleção dos indivíduos, firmas, *match* trabalhador-firma-célula e dotações locacionais fixas.

Este capítulo está organizado em quatro seções. A próxima seção discute a literatura sobre economias de aglomeração e prêmio salarial urbano, sublinhando os escassos trabalhos que abordam tal relação no espaço intra-urbano. Na seção 4.2 são apresentadas a discussão metodológica, expondo um modelo microfundamentado, a especificação econométrica e a estratégia de identificação e o banco de dados, em que se apresentam as variáveis utilizadas nas estimações. Na seção 4.3 são apresentados os resultados, envolvendo as estatísticas descritivas dos dados em painel, os resultados sobre as hipóteses de atenuação espacial da economia de aglomeração e heterogeneidade. Em seguida, considerações finais são feitas.

4.1. ECONOMIAS DE AGLOMERAÇÃO E O PRÊMIO SALARIAL URBANO: TEORIA E EVIDÊNCIAS EMPÍRICAS

Desde o trabalho canônico de Mincer (1974), a questão do diferencial salarial tem ganho relevância em Economia. Pesquisas no campo da Economia do Trabalho apontavam para a

concentração desse diferencial em grandes centros urbanos, relativamente aos municípios menores, em termos populacionais (COMBES; GOBILLON, 2015). Tais resultados motivaram pesquisas focadas no papel da aglomeração sobre a determinação do diferencial salarial, dando início a uma vasta literatura teórica e empírica que conecta Economia Regional e Urbana e Economia do Trabalho, tornando-se conhecida na literatura como Economia de Aglomeração.

Assim como no contexto regional, o debate do diferencial salarial é relevante numa escala menos agregada, pois a heterogeneidade espacial intra-urbana, inerente às economias, implica em rendimentos não isotrópicos (FUJITA; KRUGMAN; VENABLES, 1999). Combes, Duranton e Gobillon (2008) elencam três fontes da heterogeneidade espacial do diferencial salarial: 1) diferença na composição do mercado de trabalho (que reflete na habilidade dos trabalhadores demandados); 2) diferença na dotação natural local (clima, por exemplo); 3) interações entre trabalhadores ou entre firmas, isto é, economias de aglomeração.

Os *Central Business Districts* e *Subcenter Business Districts* são regiões de elevada concentração de emprego relativamente às demais. Como destacado por White (1999), os principais determinantes dessas regiões são economias de aglomeração e custo de transporte. No contexto de espraiamento do emprego no espaço, demanda e oferta por bens e serviços espacialmente diferenciados daquelas localizadas no CBD, no SBD ou fora dessas localidades explicam a estrutura heterogênea dos salários. Em consonância com a literatura de economia de aglomeração, sejam modelos derivados de equilíbrio geral (NEU) ou de equilíbrios parciais (NEG), o que seus resultados sugerem é que maiores salários/produktividade do trabalhador podem ser observados em áreas mais densas. O equilíbrio com diferencial salarial só é estável na presença de retornos crescentes derivados da economia de aglomeração, seja no espaço intra-urbano com múltiplos centros (JAFJE; TRAJTENBERG; HENDERSON, 1993; HENDERSON, 1991; KIM, 1991) ou no contexto regional (BRAKMAN, et al., 2009) ou na presença de competição imperfeita (COMBES; MAYER; THISSE, 2008).

A literatura empírica conectada à Economia de Aglomeração tem focado em determinantes locais que se relacionam com a produtividade, isto é, acesso ao mercado, diversidade

industrial, ou especialização. Controles para a heterogeneidade setorial abrem espaço para a distinção entre economias de urbanização e economias de localização (COMBES; DURANTON; GOBILLON, 2008; COMBES; GOBILLON, 2014). Economias de urbanização correspondem às externalidades que surgem das características relacionadas à localização, tais como a densidade; economias de localização dizem respeito às externalidades que derivam das características da indústria no espaço em que está localizada. O primeiro tipo são características que não são específicas aos setores, mas à aglomeração como um todo, e caracterizam as economias de urbanização. O segundo tipo é específico do setor e configura economias de localização.

Dois tipos de externalidades desempenham papel relevante no processo de criação de conhecimento e de difusão: externalidades de especialização, que decorrem da indústria, e externalidades de diversidade, que derivam das relações entre setores da economia. Para Marshall (1890), economias de localização relacionam aglomeração de indústrias e especialização urbana como determinantes da economia de aglomeração derivada da externalidade de especialização. Jacobs (1969), por sua vez, relaciona economias de localização à diversidade urbana, a qual permite contatos e transbordamento de ideias que derivam das relações entre setores.

Mais recentemente, Duranton e Puga (2004) microfundamentaram os determinantes da aglomeração e apontaram o compartilhamento (*sharing*), agrupamento (*pooling*) e associação (*matching*) como os canais que ativam as economias de aglomeração. Como a externalidade gerada pelo mercado de trabalho é muito similar à interação entre firmas e transbordamento de conhecimento, a literatura empírica enfrenta dificuldades em decompor tais efeitos (ROSENTHAL; STRANGE, 2004; COMBES; DURANTON; GOBILLON, 2008; GROOT; DE GROOT; SMIT, 2014). Esta dificuldade é conhecida na literatura como equivalência marshalliana, isto é, os microfundamentos são empiricamente equivalentes em termos de resultados ao nível da cidade.

Como a cidade é o *locus* das economias de aglomeração e a conexão entre densidade urbana e retornos é a razão primária para a sua existência (GLAESER; RESSEGER, 2010), esta atua como fator amplificador da taxa de crescimento salarial dos trabalhadores, uma vez que cataliza a acumulação de capital humano (GLASER; MARÉ, 2001), combina a proximidade

física com efeitos de escala, reduzindo custo de transporte sobre o mercado de bens finais e de produção de insumos (FUJITA; OGAWA, 1982; FUJITA, 1988; HELSLEY; SULLIVAN, 1991; HENDERSON; SLADE, 1993; ANAS; KIM, 1996) e reduz o custo das informações incompletas (DURANTON; PUGA, 2004).

Nesse contexto, pode haver auto seleção tanto na órbita das firmas quanto dos trabalhadores para se localizarem nas cidades. Trabalhadores mais habilidosos se selecionam para áreas mais produtivas, bem como firmas se selecionam para áreas com maior dotação não humana (estrutura de transporte, clima, etc.), causando interações relevantes (economias de urbanização e/ou economias de localização) que potencializam a produtividade, tanto do trabalhador quanto da firma (COMBES; DURANTON; GOBILLON, 2008; COMBES; GOBILLON, 2014). Essa circularidade gerada entre retornos salariais e aglomeração pode confundir o efeito da aglomeração sobre os salários, demandando a utilização de correção da endogeneidade nas estimações econométricas.

É recorrente na literatura de economia de aglomeração encontrar efeitos positivos da aglomeração urbana sobre salários, inovação, valor da terra, acessibilidade, acesso a serviços, etc. (AHLFELDT; PIETROSTEFANI, 2017). Focando exclusivamente em salários dos trabalhadores, evidências empíricas apontam prêmios salariais urbanos estimados entre 3% e 8% quando o tamanho da aglomeração dobra (ROSENTHAL; STRANGE, 2004; MELO; GRAHAM; NOLAND, 2009). Justificam-se os diferentes resultados das estimações do prêmio salarial urbano em função do país em análise (desenvolvido ou em desenvolvimento), do setor, especificação (função minceriana ou de produção da firma), identificação da variável que mensura economia de aglomeração, controles utilizados, especialmente quando se controla a heterogeneidade não observada (firma e trabalhador), tipo de dados utilizados (*cross-section* ou longitudinal) e unidade geográfica utilizada (MELO; GRAHAM; NOLAND, 2009).

O capital humano é a variável chave na determinação do nível (ROBACK, 1982; MELO, GRAHAM; NOLAND, 2009; BACOLOD; BLUM; STRANGE, 2010) e no crescimento (GLASER; MARÉ, 2001; YANKOW, 2006; GLAESER; RESSEGER, 2010) do salário do trabalhador. De acordo com Glaser e Maré (2001), trabalhadores mais qualificados se direcionam às cidades por distintas razões: 1) indivíduos com capital humano mais elevado

consideram o acesso à informação importante para seu rendimento e se direcionam aos centros urbanos devido à velocidade de difusão das informações; 2) as cidades são centro de consumo e atraem pessoas mais ricas; 3) o nível de indivíduos com elevado capital humano atuaria atraindo mais capital humano com as mesmas habilidades dos já existentes. Em valendo tais hipóteses, os trabalhadores se selecionando em direção às cidades, o prêmio salarial urbano fica justificado. Outros fatores que explicam esse prêmio são mercado potencial (maior demanda por bens e serviços) e insumos mais baratos no espaço urbano, culminando em salários maiores para os que chegam nos centros urbanos e menores salários para os que deixam as grandes cidades. Alternativamente, as cidades atuam ampliando a acumulação de capital humano e melhorando o *match* no mercado de trabalho, o que, por consequência, acelera intertemporalmente o crescimento salarial nas cidades (GLASER; MARÉ, 2001; HALFDANARSON; HEUERMAN; SUEDEKUM, 2010). Portanto, as cidades facilitam a difusão das informações e esta característica é relevante para aqueles que apresentam maior nível de capital humano (GLAESER; RESSEGER, 2010).

As pesquisas iniciais que estimavam o efeito da aglomeração sobre a produtividade estiveram concentradas em testar a importância relativa das economias de urbanização e localização. Estimções circunscritas aos anos 1970-80, assumindo funções de produção microfundamentadas, utilizavam dados regionais ou setoriais agregados do tipo *cross-section*, devido à dificuldade em obter dados longitudinais. Mais recentemente, dados mais desagregados têm sido utilizados nas estimções empíricas, permitindo melhor representação da otimização das firmas, maior variabilidade dos dados, tratamento do viés de aglomeração (MELO; GRAHAM; NOLAND, 2009) e *sorting* dos trabalhadores e firmas (GLAESER; MARÉ, 2001; ROSENTHAL; STRANGE, 2008).

A disponibilidade de dados longitudinais passou a permitir controles para habilidades observadas e heterogeneidade individual fixas não observada na estimação do prêmio salarial urbano. O controle por heterogeneidade não observada do trabalhador objetiva controlar a auto seleção do trabalhador, que pode confundir o efeito da aglomeração sobre os salários, a heterogeneidade da firma e a auto seleção espacial associada. Ou seja, o efeito positivo da produtividade pode derivar da elevada presença de trabalhadores mais habilidosos nas áreas mais densas, refletindo não o efeito da economia de aglomeração, mas sim o elevado nível

do capital humano local. Todavia, parte significativa do prêmio salarial é acumulada com o tempo de residência na cidade. Portanto, o diferencial salarial urbano não está apenas relacionado à habilidade do trabalhador, mas ao fato de que a cidade torna o trabalhador mais produtivo (GLAESER; MARÉ, 2001). Por outro lado, parte desse *sorting* do trabalhador e localização da firma deriva da própria aglomeração. Por consequência, ao controlar por tais efeitos fixos, a variável que identifica aglomeração captura um efeito puro da aglomeração.

Os trabalhos empíricos, tanto na literatura internacional quanto nacional, têm progredido na identificação do efeito das economias de aglomeração sobre o retorno salarial urbano. Esse progresso deriva da maior disponibilidade de dados, os quais têm permitido não somente controlar por heterogeneidade não observada do trabalhador e da firma, mas também trabalhar com um nível de agregação espacial que permite responder a questões concernentes ao espaço (ROSENTHAL; STRANGE, 2004). Adicionalmente, os trabalhos têm buscado corrigir a endogeneidade que é inerente ao problema do retorno salarial à aglomeração.

Considerando estudos que utilizam dados em painel, pode-se encontrar uma gama de trabalhos mensurando o prêmio salarial urbano na literatura internacional, tais como Glaser e Maré (2001), Yankow (2006), Combes, Duranton e Gobillon (2008), Mion e Naticchioni (2009)³¹.

O estudo publicado por Glaeser e Maré (2001) é um dos primeiros que estimam prêmio salarial urbano com controle para heterogeneidade individual para a economia dos EUA. A economia de aglomeração é identificada via inclusão de *dummy* para capturar o local de trabalho do indivíduo, dentro ou fora da região metropolitana. Utilizando dados em *cross-section*, os autores encontram um diferencial salarial de 33%. Utilizando dados longitudinais de diferentes fontes e controlando para as habilidades individuais não observadas, os autores encontram um prêmio salarial entre 4% e 11%. Yankow (2006) estende as análises feitas por Glaeser e Maré (2001), estimando que trabalhadores de grandes cidades têm em média retorno salarial de 19%, enquanto os de cidades pequenas, 8% (utilizando estimador Mínimos Quadrados Ordinários – OLS). Ao controlar por efeito fixo de indivíduos, os retornos

³¹ Para maior discussão sobre trabalhos que utilizam outros estimadores, ver Melo, Graham e Noland (2009). Os autores fazem uma meta-análise sobre o prêmio salarial derivado das economias de aglomeração, considerando 729 elasticidades estimadas em 34 diferentes estudos.

salariais nas grandes cidades caem para 5% e nas cidades pequenas, para 3,3%. Combes, Duranton e Gobillon (2008) apontam que desconsiderar a auto seleção do trabalhador resulta em um importante problema de variável omitida. O prêmio salarial urbano estimado foi entre 3-5%, representando uma redução do prêmio entre 40-50% após controlar pelas habilidades não observadas dos trabalhadores. Mion e Naticchioni (2009) mostram que habilidades individuais e tamanho da firma estão positivamente correlacionados e explicam grande parte da heterogeneidade espacial dos salários na Itália. Os autores estimaram um retorno salarial médio de 2,2%.

Embora os trabalhos atuais tenham focado em controlar o *sorting* da firma e dos trabalhadores, Groot, de Groot e Smit (2014) apontam à criação de uma caixa preta, decorrente dos efeitos fixos estimados. A inclusão de efeitos fixos pode gerar viés de seleção, dado que estes são identificados por aqueles trabalhadores que migram de região ou setor. Assim, o estimador de dados em painel com efeito fixo estimaria o limite inferior, enquanto OLS, o limite superior do efeito da aglomeração sobre a produtividade.

No Brasil, estudos empíricos com foco na existência de retornos positivos às aglomerações urbanas são recentes e limitados. Entretanto, tais pesquisas apontam evidências de prêmio salarial urbano positivo em áreas aglomeradas. Falcão e Silveira Neto (2007) estudaram a externalidade gerada pela aglomeração (capital humano agregado, em termo de anos de estudo) em 109 municípios brasileiros, todos com mais de 100 mil habitantes, utilizando dados dos censos de 1980, 1991 e 2000. Os resultados derivados do estimador de OLS e 2SLS-IV (mínimos quadrados dois estágios com variável instrumental) indicam prêmio salarial médio entre 0,8% e 8%, sendo trabalhadores com ensino superior completo aqueles que mais se beneficiariam com o acúmulo de capital humano na cidade. Os resultados do estudo de Campos e Silveira Neto (2009) apontam prêmio salarial urbano entre 1,7 e 2%, utilizando estimador 2SLS-IV. Cruz e Naticchioni (2012), utilizando dados da PNAD para os anos 2002 e 2009 e estimador OLS, identificaram um diferencial salarial de 22,8% para os trabalhadores que vivem em regiões metropolitanas no ano de 2002 e de 17,4% em 2009.

Mais recentemente, os estudos que estimam os retornos salariais à aglomeração no Brasil têm prestado atenção às habilidades não observadas do trabalhador (ROCHA; SILVEIRA NETO; GOMES, 2011; SILVA; SANTOS; FREGUGLIA, 2016; BARUFI; HADDAD; NIJKAMP,

2016; SILVA, 2017) e à auto seleção das firmas (SILVA, 2017). Rocha, Silveira Neto e Gomes (2011) apresentam novas evidências sobre as externalidades positivas da aglomeração sobre os salários estimando um painel dinâmico com controle para o *sorting* do trabalhador. Utilizando a Relação Anual de Estatística do Trabalho e Emprego (RAIS-Migra), os autores encontram que parte significativa dos diferenciais salariais entre os trabalhadores de grandes centros e os que trabalham fora dele é removida, isto é, caindo de 76,7% para 9,4%. Silva, Santos e Freguglia (2016) estimam a distribuição espacial dos efeitos da aglomeração sobre os diferenciais salariais para 24 regiões metropolitanas do Brasil com base em um painel de microdados da RAIS-Migra. Os resultados sugerem ganhos de aglomeração de 3% para os trabalhadores formais das RM.

Barufi, Haddad e Nijkamp (2016) avaliam o retorno salarial em diferentes setores da indústria face à economia de aglomeração. Em ambos estudos, os autores utilizam dados da RAIS para os anos 2004, 2008 e 2012 no modelo de efeito fixo. Os resultados vão ao encontro da literatura e apontam números que variam entre 5% e 9%, sugerindo ganhos com economia de urbanização. Os resultados diferem em função da tecnologia da firma. Indústrias com baixa e alta tecnologia apresentaram maiores retornos à aglomeração urbana relativamente às indústrias com tecnologia média. Os modelos estimados com variáveis instrumentais apresentam pouca variabilidade em relação aos resultados do retorno salarial dos trabalhadores, isto é, os resultados variam entre 4% e 9%.

O estudo de Silva (2017) apresenta avanços quanto à abordagem empírica, ao controlar simultaneamente o *sorting* do trabalhador e firma utilizando estimador de efeito fixo, no primeiro estágio do modelo, capturando o efeito sobre o salário individual. Para tanto, os dados da RAIS para os anos entre 2002 e 2014 são utilizados. Os resultados que controlam os efeitos fixos do trabalhador e firma simultaneamente sugerem a inexistência de prêmio salarial urbano. Esse resultado indica que, ao considerar áreas de trabalho no Brasil, o prêmio salarial urbano decorre da auto seleção de trabalhador e firma. Ao avaliar por arranjos populacionais no primeiro estágio da estimação, a autora encontra um padrão de comportamento dos respectivos prêmios salariais, a saber: prêmios derivados dos atributos não observáveis do indivíduo, prêmios derivados dos atributos não observados da firma e prêmios derivados do *match* entre atributos não observáveis dos indivíduos e firmas. Após o

controle simultâneo para os atributos não observáveis dos indivíduos e firmas, além dos controles observáveis, a regressão explicou 93,5% da variação dos salários individuais. A decomposição dos salários locais dos 362 arranjos populacionais considerados apresentou o efeito puro (controladas pelas características observadas e não observadas dos indivíduos e firmas e dos arranjos) da aglomeração urbana sobre o salário médio dos arranjos entre 2,6% e 4,2%. Ao corrigir a endogeneidade existente entre densidade e salários locais, os resultados variam entre 4,7% (desconsiderando componentes não observados) e 4,9% (controlando pelos componentes não observados).

Mesmo que na literatura de economia de aglomeração a relação produtividade-densidade seja bem estabelecida e que a unidade geográfica seja uma potencial fonte das variações dos efeitos estimados (ROSENTHAL; STRANGE, 2004; OVERMAN, 2004; BRIANT; COMBES; LAFOURCADE, 2008; MELO; GRAHAM; NOLAND, 2009), poucos trabalhos se voltaram em explorar a escala geográfica ou estimar o prêmio salarial urbano numa escala intra-urbana. Recentemente, alguns trabalhos têm buscado utilizar unidades espaciais que capturem as características dos mercados de trabalho locais (agrupamento de mercado de trabalho e transbordamento de conhecimento) ao invés de utilizar unidades geográficas pautadas em divisões políticas (municípios, estados, etc.). Por exemplo, Combes, Duranton e Gobillon (2008) utilizam áreas de emprego na França, Barufi, Haddad e Nijkamp (2016) utilizam regiões de influência das cidades (REGIC) e Silva (2017) utiliza arranjos populacionais do IBGE. Todavia, os trabalhadores continuam sendo tratados como aglomerados que compartilham grupos de cidades, desconsiderando qualquer heterogeneidade intra-urbana (ROSENTHAL; STRANGE, 2004).

As externalidades geradas tanto pelo transbordamento de conhecimento (ANSELIN; VARGA; ACS, 1997; AUTANT-BERNARD, 2001; FELDMAN; AUDRETSCH, 1999) quanto pela aglomeração de trabalhadores (ROSENTHAL; STRANGE, 2008; VAN SOEST; GERKING; VAN OORT, 2006) são geograficamente limitadas e apresentam alcances espaciais diferenciados. Assim, a agregação espacial dificulta responder qual a extensão da externalidade derivada da aglomeração de população, de empregados ou capital humano (OVERMAN, 2004). Em outras palavras, limita a estimação da velocidade de atenuação da

economia de aglomeração sobre o rendimento dos trabalhadores ao considerar a heterogeneidade intra-urbana.

Como discutido nos modelos endógenos de subcentralidades, as firmas escolhem sua localização espacial e tal escolha é uma função de como os benefícios da localização são atenuados no espaço. Adicionalmente, as famílias buscam localizar-se minimizando o custo de *commuting*, mas mantendo a renda que maximiza sua utilidade, ou seja, a aglomeração afeta a remuneração do trabalhador tanto pelas economias de urbanização quanto pelas economias de localização. Nesse contexto, destacam Rosenthal e Strange (2004), *attenuation* é um fator importante para compreender o espraiamento urbano do efeito da aglomeração, controlado por características locais, das firmas e dos indivíduos.

Diferentemente das abordagens empíricas que avaliam a densidade sobre os retornos salariais, a literatura empírica que toma como base modelos da NEU estima o gradiente de salários urbanos, ou seja, o efeito do distanciamento do CBD sobre o salário do trabalhador. Nessa abordagem, os trabalhadores vivem em determinada localização e apresentam uma curva de salários que indica o salário mínimo que as famílias representativas estão dispostas a receber para trabalhar naquela região da cidade. O envelope inferior das curvas de salários demandados forma o gradiente de salário de mercado. Na literatura empírica sobre gradientes de salários urbanos, duas hipóteses distintas são testadas. Sob a primeira hipótese, testa-se a relação entre *commuting* e salários, sendo a hipótese subjacente de que os trabalhadores preferem um deslocamento pendular mais elevado para pagar menores alugueis. Quanto à segunda hipótese, testa-se o gradiente de salários, ou seja, a relação entre salários e o distanciamento do CBD, sob a hipótese de que aglomeração importa e, portanto, os salários decrescem quando a distância aumenta (WHITE, 1999).

Em relação à primeira hipótese, White (1999) aponta que a distância de *commuting* não é uma boa medida para avaliar gradiente de salários, pois pode haver uma seleção da amostra por trabalhadores que se desloquem para fora da cidade, ao invés de se deslocar para o *core* da cidade em que moram, ampliando o salário ao se distanciar do CBD e/ou SBD. A direção de tal causação é controversa na literatura da Nova Economia Urbana, abrindo espaço para a discussão sobre a direção da causalidade, i.e., o padrão espacial da população depende da distribuição do emprego ou o padrão espacial do emprego depende da distribuição espacial

da população. A partir dessa discussão, modelos que consideram a localização do CBD ou SBD exógenas ganharam espaço devido à trababilidade dos modelos teóricos. Mesmo em face de tal discussão, Steinnes (1997), Cooke (1978), Mills e Price (1984) e Boarnet (1994) apontam que as firmas seguem a população. Uma vez que os salários estão relacionados à acessibilidade ao centro de emprego, a economia de aglomeração também depende da acessibilidade.

A segunda hipótese também é passível de crítica. Como discutido por Romanos (1979), a consideração apenas do CBD leva a coeficientes estimados de forma enviesada quando existir mais que uma concentração de emprego na região avaliada, ou seja, SBD. As medidas de distância tanto ao CBD quanto aos SBD não capturam os efeitos diretos das densidades sobre os salários, mas sim os gradientes em relação as âncoras de preços (i.e., os CBD e SBD) previamente identificados. Adicionalmente, a limitação dos estudos procedentes dessa linhagem decorre da estimação de gradientes de salários unidimensionais, o que significa assumir implicitamente uma cidade monocêntrica (TOMOTHY; WHEATON, 2001).

Pode-se dividir os trabalhos empíricos que estudam a relação economias de aglomeração e retorno salarial em função da variável de identificação. No primeiro grupo estão os trabalhos que estimam o gradiente de salários (*wage gradients*), tais como Rees e Shultz (1970), Eberts (1981), Ilhanfeldt (1992), McMillen e Singell (1992), Thimoty e Wheaton (2001); Haddad e Barufi (2017). O segundo grupo identifica a relação aglomeração-prêmio salarial considerando a densidade de trabalhadores e estima a atenuação dos salários (*wage attenuation*), tais como van Soest, Gerking, van Oort (2006), Rosenthal e Strange (2008), Andersson, Klaesson, Larsson (2016). Esses trabalhos não impõem qualquer âncora de preços (CBD e/ou SBD). Tal estratégia deriva muito mais do recorte geográfico do que propriamente de uma estratégia de superar a limitação empírica das estimações de gradiente de salários, pois os autores têm como objeto espacial de trabalho o país como um todo.

Poucos trabalhos na literatura internacional se voltaram a estudar *wage gradients*. Isso pode derivar da dificuldade de obtenção de bases de dados com informação sobre o local de residência e de trabalho dos indivíduos, permitindo identificar o *commuting* e calcular distâncias intra-urbanas em relação aos CBD e SBD, e/ou da relativa importância dos

modelos desenvolvidos a partir de Krugman (1991), que atraíram o foco aos modelos da NEG para a escala regional.

Um primeiro esforço foi feito por Rees e Shultz (1970), comparando os salários no CBD de Chicago com as regiões ao norte, sul e oeste do CBD. Ao estimar gradientes de salários para o mercado de trabalho de Chicago utilizando informações sobre características e localização da firma, os resultados apontam para salários mais elevados no CBD em relação às regiões norte e sul. Eberts (1981) estima a elasticidade do gradiente de salários utilizando dados sobre o mercado de trabalho para 100 municípios da região de Chicago. Como resultado, o autor encontra que, em termos absolutos, o salário-hora decai em US\$ 24 para o setor da administração pública, US\$ 10 para o setor eclesiástico e US\$ 12 para polícia quando o indivíduo trabalha fora do CBD. Utilizando informações dos microdados (PUMS – *Public Use Microsample*) do Censo de 1980, Ilhanfeldt (1992) estima o gradiente de salários para as regiões metropolitanas da Filadélfia, Detroit e Boston. Para mensurar a distância de *commuting*, o autor calcula a distância de cada zona (PUMA) ao CBD de cada uma das respectivas RM. Dos 21 gradientes de salários estimados, 18 demonstraram sinais negativos e os gradientes positivos não apresentaram significância estatística. Na média, o salário declina em 1% por milha.

McMillen e Singell (1992) também utilizam PUMS, mas por trabalhador individual ao invés de grandes médias de zonas do Censo, o que representa um avanço nessa literatura de *wage gradient*. A abordagem econométrica de dois estágios desenvolvida pelos autores estima primeiramente a probabilidade de trabalhar no CBD ou no subúrbio; no segundo estágio, após calcular a distância de cada trabalhador ao CBD, estima-se o gradiente de salários. Apenas o gradiente de salários calculado para Detroit é estatisticamente significativo e com sinal esperado, com os demais resultados não sendo estatisticamente diferentes de zero. Timothy e Wheaton (2001) mensuram o gradiente de salários intra-urbano utilizando a base de dados PUMS para o ano de 1990, considerando a divisão espacial PUMAs³². Os autores identificam o efeito da aglomeração considerando *dummies* de área de trabalho (PUMAs),

³² Cabe destacar que as áreas residências e de trabalho que a base de dados PUM fornecem são relativamente grandes, pois consistem de grupos de *counties* com uma população mínima de 100 mil habitantes. Assim, muitas áreas metropolitanas que contêm uma grande área, como Los Angeles, tendem a apresentar uma limitação para trabalhos que estimam gradientes salariais.

isto é, se a pessoa trabalha no CBD ou fora dele, e tempo de *commuting* médio do trabalhador. Os resultados mostram prêmios salariais de 15% para Boston e de 18% para Minneapolis-Saint Paul. Recentemente, Haddad e Barufi (2017) estimaram o gradiente de salários em relação ao CBD para a Região Metropolitana de São Paulo utilizando dados da Pesquisa Origem-Destino (OD), considerando as áreas de ponderação da amostragem da OD. Embora os autores considerem a distância apenas ao CBD, o trabalho demonstra um avanço, ao tratar a endogeneidade causada pela circularidade entre aglomeração e salário.

Resumidamente, em geral, os trabalhos empíricos que estimam *wage gradient* desconsideram a endogeneidade entre aglomeração e produtividade do trabalhador, bem como o *sorting* espacial, tanto das firmas quanto dos trabalhadores. Adicionalmente, assumem, por simplicidade, gradientes de salários unidimensionais e monotônicos, desconsiderando a possibilidade de cidades policêntricas. Tal pressuposto, implicitamente, considera o centro histórico como sendo a única região que concentra economias de aglomeração. Essas estimativas para Chicago, Boston e Minneapolis-Saint Paul desconsideram a literatura de identificação de subcentralidades (MCDONALD, 1987; MCDONALD; PRATHER, 1994; MCMILLEN; MCDONALD, 1998, MCMILLAN, 2001b) que mostram múltiplas centralidades para tais regiões, bem como as estimativas para a RMSP.

A partir de meados dos anos 2000, uma nova onda de trabalhos empíricos, embasados na abordagem da Nova Geografia Econômica, tem estimado a extensão geográfica das economias de aglomeração e calculam a *wage attenuation*. A hipótese implícita desses estudos é que os benefícios das economias de urbanização ou de localização tendem a se atenuar no espaço. Esses trabalhos representam um avanço empírico, pois relaxam a hipótese de gradiente de salários monotônicos e lançam luz sobre a questão da escala da densidade e sua extensão (ANDERSSON; KLAESSON; LARSSON, 2016). Ao considerar áreas de aglomeração pequenas, os resultados podem fornecer indícios sobre alguns dos vários microfundamentos das economias de aglomeração, principalmente *learning*. Efeitos de aprendizado são comumente relacionados à proximidade entre pessoas e derivados de contato face a face (DURANTON; PUGA, 2004). Storper e Venables (2004) apontam para quatro principais características do contato face a face: a) é uma eficiente tecnologia de comunicação; b) ajuda a resolver problemas; c) facilita a socialização e o aprendizado; e, d)

fornece motivação psicológica. Segundo Granovetter (1995), essa relação mais próxima entre trabalhadores facilita o processo de *signaling* e *screening*, bem como a construção de uma rede de contatos³³.

Em geral, os resultados empíricos derivados desses estudos apontam para a queda do efeito da aglomeração em função da distância e que o escopo geográfico dessas externalidades é menor do que o tamanho médio das áreas de trabalho geralmente utilizados. Isso gera dúvida sobre quais dos municípios ou quaisquer outras divisões espaciais que compõem a RMSP, as áreas REGIC ou os arranjos populacionais, por exemplo, apresentam maior parcela do efeito positivo da aglomeração. E tanto para Jacob (1969), que defendia a tese de que as cidades obtinham vantagens produtivas devido sua diversidade social e econômica “embalada” em espaços restritos, facilitando o contato aleatório e casual entre pessoas, quanto para Marshall (1890), que também defendia a hipótese de transbordamento de conhecimento, embora dentro de *clusters* industriais, o nível do aninhamento importa. Quanto mais próximo, mais provável que a aglomeração afete o salário.

Rosenthal e Strange (2003) e Van Soest, Gerkin e van Oort (2006) avaliam o efeito da aglomeração sobre o nascimento e taxa de crescimento do emprego no espaço. Rosenthal e Strange (2003) estimam o efeito da aglomeração sobre o surgimento de empregos e novas firmas. Para identificar a extensão da aglomeração, os autores criaram círculos concêntricos ao redor do centroide de cada zona postal, com raios variando de 1 a 15 milhas. Van Soest, Gerkin e van Oort (2006) utilizam dados do Censo 1988 e 1997 para o sul da Holanda ao nível de CEPs (416 áreas postais contíguas que medem em média 6 Km²). Os autores estimam os efeitos das economias de aglomeração sobre os setores de atividades econômicas (taxa de crescimento do emprego). A identificação do transbordamento da aglomeração deriva da concentração de empregos ponderada pelas respectivas distâncias. Ambos estudos apontam

³³ Stoper e Venables (2004) destacam que a internet reduziu o custo da informação e forças opostas são observadas no mundo atual. A primeira tende a descentralizar as atividades, dada essa redução do custo do espraio da informação e dado custo de transporte do trabalhador. As atividades que adeririam à produção remota e não aglomerada seriam aquelas cuja complexidade de coordenação foi padronizada e rotineirizada, passível de ser cumprida de forma atômica pelo trabalhador. Atividades complexas e atividades recém-criadas, cuja complexidade da coordenação é ainda não familiar, demandariam contatos face a face e, por tanto, proximidade geográfica. Para os autores, a tecnologia facilita a dispersão da produção, mas ao mesmo tempo desestabiliza as atividades, criando incertezas, questões de pesquisa e oportunidades desconhecidas, em um ambiente em que o conhecimento é tácito e cheio de “barulho”.

que aglomerações intra-áreas postais apresentam resultados mais elevados comparativamente aos efeitos da aglomeração identificadas pelas áreas que circundam o local de trabalho do indivíduo.

A preocupação sobre o papel das economias de aglomeração sobre a produtividade do trabalhador associada ao aninhamento espacial é mais recente ainda. Rosenthal e Strange (2008) discutem aglomeração e o papel do capital humano nas relações face a face, explorando a relação desses fatores com o prêmio salarial urbano. Os autores lidam com a questão espacial a partir de círculos concêntricos de diferentes raios (0-5, 5-25, 25-50 e 50-100 milhas), que medem o emprego de trabalhadores educados e pouco educados, sendo a área geográfica o local de trabalho (PUMA) nos Estados Unidos. Outro avanço nessa abordagem empírica está em lidar com a endogeneidade entre aglomeração e salários. Os autores estimam a semi-elasticidade da equação de salários, buscando testar a hipótese de atenuação espacial da aglomeração e encontram resultados estimados de 0,02% no primeiro círculo e queda significativa dos coeficientes para círculos com raios maiores que 5 milhas, sendo os trabalhadores mais educados os que mais se beneficiam da aglomeração. Esses resultados vão ao encontro das teorias e resultados discutidos na literatura sobre a existência do capital humano, em que contatos entre indivíduos mais educados têm efeito sobre renda e que a aglomeração de pessoas resulta em prêmio salarial heterogêneo em função do nível de educação do trabalhador.

Embora Rosenthal e Strange (2008) lidem com a questão espacial, as áreas de ponderação do censo e círculos concêntricos com raios grandes ainda conferem atritos para o entendimento da externalidade gerada pela aglomeração e em que medida tais resultados seriam capazes de serem associados às relações face a face. Andersson, Klaesson e Larsson (2016) buscam superar tal limitação ao criar células quadradas de 1 km², todavia a estratégia é aplicada para toda a Suécia, gerando certos ruídos quanto à real homogeneidade entre municípios que não compartilham semelhanças quanto ao mercado de trabalho, nesse caso o prêmio salarial estimado não é intra-urbano. Apesar da limitação, a abordagem empírica permite lidar com a questão da densidade de emprego por área levantada por Mieszkowski e Smith (1991) e Redfearn (2007), responder qual a extensão das externalidades geradas pelas economias de aglomeração e qual o papel do capital humano na apropriação dos efeitos gerados pela

aglomeração. Considerando o total de empregos em cada célula, em seus vizinhos de primeira e segunda ordem, os autores identificaram o efeito direto e indireto da aglomeração sobre os salários dos trabalhadores, sendo que o efeito direto é maior que os demais e coeficientes que capturam o efeito de primeira-ordem são maiores do que os que capturam o efeito de segunda ordem.

Em termos de abordagem empírica, os trabalhos com foco nos efeitos da aglomeração no espaço avançaram em relação aos estudos dos anos 1990, pois lidaram com a não monotonicidade espacial dos salários e com a endogeneidade gerada por causalção circular. Todavia, não introduziram na discussão a questão da auto-seleção espacial dos trabalhadores nem das firmas. Ademais, em todos os trabalhos desse grupo, a discussão se dá considerando todo o país e não foca a questão do espaço intra-urbano. Nesse sentido, esses trabalhos deixam em aberto o papel das características não observadas fixas do trabalhador e da firma, bem como o papel do *match* trabalhador-firma e não trazem ao debate o papel da aglomeração no espaço intra-urbano. Assim, o esforço desta pesquisa está em preencher essas lacunas.

4.2. METODOLOGIA

Esta sub-seção está reservada à exposição e discussão do modelo teórico que motiva esta etapa do trabalho. Em seguida são abordadas a especificação econométrica e questões relacionadas à identificação e estimação.

4.2.1. Modelo de Diferencial Salarial: *sorting* e *match* importam

Esta subseção está embasada nos trabalhos de Combes, Duranton e Gobillon (2008) – doravante, CDG – e de Rosenthal e Strange (2008).

A heterogeneidade espacial do salário não deriva apenas das distribuições não homogêneas dos efeitos da aglomeração no espaço, mas também decorre da heterogeneidade dos trabalhadores no que tange às habilidades, das amenidades espaciais e da interação produtiva entre firmas, entre firmas e trabalhadores, e entre esses agentes e o espaço (ROBACK, 1982; CDG, 2008).

De acordo com CDB (2008), três conjuntos de fatores explicam a não isotropia do salário entre áreas. O primeiro fator decorre da heterogeneidade espacial nas habilidades dos trabalhadores. Os trabalhadores podem utilizar de tomadas de decisão estratégicas, auto

selecionando-se para áreas que reforcem a produtividade do seu trabalho. O segundo fator associa tal diferença às dotações não humanas, isto é, amenidades espaciais (características geográficas, por exemplo). O último conjunto de fatores correlaciona heterogeneidade espacial dos salários às interações entre trabalhadores e entre firmas, as quais ocorrem em localizações específicas e elevam a produtividade.

No modelo CDG, o salário do trabalhador i , w_i , não depende da localização em que esse trabalhador está localizado, mas segue a seguinte relação:

$$w_i = A s_i \quad (4.1)$$

Onde, s_i captura a habilidade do trabalhador e A , a produtividade. A principal questão colocada é que a produtividade do trabalho depende de dotações locais E_c localizadas na área/célula c . Nesse contexto, o salário local é dado como $w_c = A(E_c)$. A partir do modelo, tem-se, em equilíbrio, que o salário do trabalhador i na área $c(i)$ é dado por

$$w_i = A(E_{c(i)}, I_{c(i)}, I_{c(i),k(i)}) s_i \quad (4.2)$$

Onde, $I_{c(i)}$ captura as economias de de urbanização e $I_{c(i),k(i)}$ captura as economias de localização, discutidas acima.

Assume-se a hipótese de uma firma representativa competitiva que opera com retornos constantes a escala ($0 < b \leq 1$), que utiliza como insumos habilidade e horas de trabalho dos trabalhadores $l_{i,t}$, outros fatores $z_{a,k,t}$ (terra, por exemplo) e observa um fator de produtividade $A_{c,k,t}$, o quais são heterogêneos em função da localização c e do setor k no período t . A função de maximização de lucro da firma segue

$$\pi_{c,k,t} = p_{c,k,t} \left[A_{c,k,t} (\sum_{i \in (c,k,t)} s_{i,t} l_{i,t})^b (z_{c,k,t})^{1-b} \right] - \sum_{i \in (c,k,t)} w_{i,t} l_{i,t} - r_{c,k,t} z_{c,k,t} \quad (4.3)$$

Em equilíbrio, a função de compensação do trabalho do trabalhador é

$$w_{i,t}^* = \frac{\partial \pi_{c,k,t}}{\partial l_{i,t}} = b p_{c(i,t),k(i,t),t} A_{c(i,t),k(i,t),t} \left(\frac{z_{c(i,t),k(i,t),t}}{\sum_{i \in (c,k,t)} s_{i,t} l_{i,t}} \right)^{1-b} s_{i,t} \quad (4.4)$$

Derivando a função de lucro da firma em relação a $z_{a,k,t}$ e combinando com a Equação 4.4, tem-se

$$w_{i,t}^* = b(1-b)^{\left(\frac{1-b}{b}\right)} (p_{c(i,t),k(i,t),t} \left(\frac{A_{c(i,t),k(i,t),t}}{(r_{c(i,t),k(i,t),t})^{1-b}} \right)^{\frac{1}{b}} s_{i,t} = B_{c(i,t),k(i,t),t} s_{i,t} \quad (4.5)$$

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

A partir da equação 4.5 nota-se que o diferencial salarial espacial está relacionado à heterogeneidade das habilidades individuais $s_{i,t}$, ao fator de produtividade total $A_{c,k,t}$, e/ou preço dos produtos $p_{c,k,t}$ e/ou insumos não relacionados ao fator trabalho $r_{c,k,t}$. Esses mecanismos que afetam os salários são dissociativos. Embora os determinantes das desigualdades salariais possam ser identificados (dotação local, interações e habilidades), não é possível rastrear por qual destes o efeito está passando. Sob essa impossibilidade, uma localização espacial pode enfrentar efeitos aglomerativos positivos e negativos simultaneamente, tornando difícil a decomposição de tais efeitos. Portanto, resultados derivados de modelos econométricos empíricos apenas identificam o efeito líquido da aglomeração, como no modelo CDG (2008).

4.2.2. Especificação Econométrica e Estratégia de Identificação

Para a tratabilidade empírica da Equação 4.5, segue-se a proposta elaborada em CDG (2008), mas estendendo a especificação em face à discussão empírica recente do campo da Economia do Trabalho. Considere-se que a função de *skill* dos trabalhadores e o fator de produtividade seguem as seguintes relações, respectivamente:

$$\log s_{i,t} = X_{i,t} \varphi + \delta_i + \epsilon_{i,t} \quad (4.6)$$

$$B_{c,k,t} = \beta_{c,t} + \mu_{k,t} + A_{c,t} \rho + I_{a,k,t} \gamma_k \quad (4.7)$$

onde $X_{i,t}$ denota a matrix de covariadas referente às características do trabalhador, que variam no tempo, φ é o vetor de coeficientes a ser estimado, δ_i corresponde aos efeitos fixos do trabalhador (efeito puro do trabalhador), $\epsilon_{i,t}$ é o termo de erro assumido i.i.d, $\beta_{c,t}$ denota o efeito fixo área-ano (efeito puro de área), $\mu_{k,t}$ denota o efeito fixo indústria-ano, $A_{c,t}$ é uma parcela observada da dotação local (aqui identificada pela densidade do emprego, também referida por aglomeração), ρ é o parâmetro associado a $A_{c,t}$, $I_{a,k,t}$ representa a interação entre

área e indústria no tempo e γ_k é um vetor de coeficientes associados. Combinando as duas últimas equações, tem-se:

$$\log w_{i,t} = A_{c,t}\rho + \beta_{c,t} + \mu_{k,t} + I_{a,k,t}\gamma_k + X_{i,t}\varphi + \delta_i + \epsilon_{i,t} \quad (4.8)$$

Devido à tratabilidade e focando em uma especificação econométrica mais parcimoniosa, assume-se que a tendência temporal é a mesma para todas as indústrias e células do grid, resultando em menos coeficientes a serem estimados. Dessa forma, a especificação econométrica apresenta efeitos fixos de célula, indústria e tempo.

A especificação econométrica proposta em CDG (2008) apresenta duas limitações principais. Primeiro, especificações econométricas que controlam apenas por efeito fixo de indústria podem enfrentar viés de estimação devido à correlação entre os efeitos fixos de firmas e indústria (Abowd, Kramarz e Margolis, 1999). Adicionalmente, como características espaciais são relevantes para a escolha locacional da firma, a omissão de efeitos fixos da localização espacial forneceria estimações enviesadas dos coeficientes. Segundo, tanto os modelos teóricos quanto empíricos no campo da Economia do Trabalho discutem sobre o *match* trabalhador-firma como um acelerador da produtividade do trabalhador, caso esse *match* associe boas firmas a bons trabalhadores. No caso contrário, um *match* ruim, i.e., bons trabalhadores em firmas ruins ou vice-versa, o efeito sobre a produtividade é inferior ao caso de bons *matches* (WOODCOCK, 2008; 2015; LAVETI; SCHUMUTT, 2016). Woodcock (2008; 2015) apontam que trabalhadores em *matches* melhores apresentam menores probabilidade de mudança de empregos do que trabalhadores em *matches* piores. A omissão de controle do *match* trabalhador-firma também envia os resultados das estimações (WOODCOCK, 2008; 2015; LAVETI; SCHUMUTT, 2016). Entretanto, firmas e indivíduos podem beneficiar-se diferentemente da característica espacial do mercado de trabalho, dependendo das habilidades do trabalhador e tecnologias da firma. Buscando mitigar o viés de variável omitida, a especificação econométrica do modelo CDG é expandida acrescentando controles para efeito fixo da firma ψ_j e para efeitos de *match* trabalhador-firma-área, $\phi_{i,j,c}$.

A estimação do modelo econométrico é feita em um único estágio, similar a Rosenthal e Strange (2008). A dotação local é decomposta nos seguintes componentes: a) densidade do

emprego (economia de urbanização); b) economias de localização (nível industrial - $S_{c,t}$); c) efeito fixo da área (células do grid); d) interação setor-área; e e) *match* trabalhador-firma-área. Os três últimos componentes capturam características fixas não observáveis. Assim, a especificação econométrica final é

$$\log w_{i,t} = A_{c,t}\rho + S_{c,t}\eta + \beta_c + \mu_{k,t} + I_{a,k,t}\gamma_k + X_{i,t}\varphi + \delta_i + \psi_j + \phi_{i,j,c} + \epsilon_{i,t} \quad (4.9)$$

Sob a hipótese de que $A_{c,t}$ pode ser decomposta espacialmente e representada por $A_{1,t}$, $A_{2,t}$ e $A_{3,t}$, onde $A_{2,t}$ e $A_{3,t}$ representam os vizinhos mais próximos e mais distantes de $A_{1,t}$, respectivamente, tem-se

$$\log w_{i,t} = A_{1,t}\rho_1 + A_{2,t}\rho_2 + A_{3,t}\rho_3 + S_{c,t}\eta + \beta_c + \mu_{k,t} + I_{a,k,t}\gamma_k + X_{i,t}\varphi + \delta_i + \psi_j + \phi_{i,j,c} + \epsilon_{i,t} \quad (4.10)$$

Mesmo em face de um modelo que controla por características observáveis do trabalhador e da firma e pelos diversos efeitos fixos supracitados, parte da habilidade do trabalhador ω_{i_s} pode estar omitida e compor o termo de erro, tal que

$$\epsilon_{i,t} = \omega_{i_s} + \varepsilon_{i,t} \quad (4.11)$$

Tal omissão pode estar correlacionada com a aglomeração do emprego pelo canal da auto seleção (*sorting*) do trabalhador. Nesse sentido, trabalhadores com determinadas habilidades que variam no tempo e que não são controladas no modelo podem explicar o *sorting* e, por sua vez, aglomeração. Seguindo a abordagem proposta por Rosenthal e Strange (2008), assume-se que $\omega_{i_s,t}$ é linear e relacionada com $A_{1,t}$, $A_{2,t}$ e $A_{3,t}$ tal que

$$\omega_{i_s,t} = a_1A_{1,t} + a_2A_{2,t} + a_3A_{3,t} + \delta_{i_s} \quad (4.12)$$

Onde δ_{i_s} captura características não observadas constantes no tempo do trabalhador e não está correlacionado com a densidade do emprego. Substituindo a Equação 4.12 na Equação 4.11, tem-se a Equação 4.13.

$$\log w_{i,t} = A_{1,t}(\rho_1 + a_1) + A_{2,t}(\rho_2 + a_2) + A_{3,t}(\rho_3 + a_3) + S_{c,t}\eta + \beta_c + \mu_{k,t} + I_{a,k}\gamma_k + X_{i,t}\varphi + \delta_i + \psi_j + \phi_{i,j,c} + \epsilon_{i,t} \quad (4.13)$$

O termo δ_{i_s} é excluído da Equação 4.13 por estar aninhado no efeito fixo do trabalhador δ_i .

Assumindo que existe atenuação espacial da aglomeração, dois principais cenários podem ser observados. Primeiro, sob a hipótese de que $a_1 = a_2 = a_3 = 0$, a taxa de atenuação é dada apenas por $d = \rho_1 - \rho_2 - \rho_3$, onde $\rho_1 > \rho_2 > \rho_3$. Nesse cenário, o estimador de Efeitos Fixos forneceria coeficientes estimados não enviesados. Segundo, sob a hipótese de que $a_1 > a_2 > a_3$ vale, a taxa de atenuação é tal que $d = (\rho_1 - \rho_2 - \rho_3) + (a_1 - a_2 - a_3)$. Se o segundo cenário vale, então a estimação econométrica fica comprometida devido à clara endogeneidade existente. Assim, variáveis *proxies* ou instrumentais são demandadas para lidar com tal viés.

Outra fonte de endogeneidade pode derivar do *sorting* da firma. Firms se auto selecionam para lugares em que sua produtividade possa ser otimizada. Áreas que provêm melhores condições geográficas ou sejam mais próximas de um *pool* de trabalhadores com habilidades adequadas às atividades, por exemplo, são escolhidas pelos empresários. Tal estratégia eleva a densidade de trabalhador por área e retroalimenta a produtividade do trabalhador e o efeito da aglomeração (causação reversa). Assumindo que existam características não observadas variantes no tempo da firma $v_{j,t}$ que expliquem sua tecnologia T , este fator compõe o termo de erro, tal que

$$\epsilon_{i,t} = v_{j,t} + \varepsilon_{i,t} \quad (4.14)$$

Supondo a relação linear entre $v_{j,t}$ e as variáveis de aglomeração, tem-se:

$$v_{j,t} = b_1 A_{1,t} + b_2 A_{2,t} + b_3 A_{3,t} + \psi_{jT} \quad (4.15)$$

Onde ψ_{jT} captura o efeito fixo da firma j em relação à tecnologia T . Substituindo a equação 4.15 em 4.14 e assumindo que a hipótese de atenuação espacial da aglomeração valha ($b_1 > b_2 > b_3$), a taxa de atenuação é similar ao caso discutido acima.

É possível que ambas fontes de endogeneidade ocorram simultaneamente. Nesse contexto, os parâmetros que capturam o efeito da aglomeração seriam compostos por três elementos, tais como

$$\log w_{i,t} = A_{1,t}(\rho_1 + a_1 + b_1) + A_{2,t}(\rho_2 + a_2 + b_2) + A_{3,t}(\rho_3 + a_3 + b_3) + S_{c,t}\eta + \beta_c + \mu_{k,t} + I_{a,k}\gamma_k + X_{i,t}\varphi + \delta_i + \psi_j + \phi_{i,j,c} + \epsilon_{i,t} \quad (4.16)$$

Valendo a hipótese de atenuação para ambos os casos, a taxa de atenuação da aglomeração é medida por

$$d = (\rho_1 - \rho_2 - \rho_3) + (a_1 - a_2 - a_3) + (b_1 - b_2 - b_3) \quad (4.17)$$

Assumindo que valha a hipótese de atenuação e que os respectivos parâmetros sejam estritamente positivos, observa-se correlação positiva entre as variáveis omitidas e aglomeração. A omissão de características relevantes dos indivíduos e/ou firmas resultaria em subestimação dos coeficientes, *lower bound*. Todavia, assumindo que a hipótese de atenuação ainda é válida e que os parâmetros a e b sejam estritamente negativos, então os coeficientes estimados seriam enviesados para cima, sobrestimados.

A questão de pesquisa que se coloca nesse capítulo claramente se depara com as duas fontes de endogeneidade devido aos seguintes mecanismos: a) potencial circularidade entre produtividade e aglomeração e b) omissão de variável relevante. A circularidade entre a aglomeração e a remuneração do trabalhador pode ser facilmente explicada, pois elevados salários em determinadas áreas tendem a atrair mais trabalhadores. Quanto maior a aglomeração, maior o retorno salarial, considerando os resultados positivos da aglomeração sobre a renda, como discutido na sub-seção 4.1. Ademais, é possível que existam variáveis omitidas que justifiquem a auto seleção do trabalhador, tais como habilidades que variam no tempo e não são controladas no modelo.

Para mitigar tal preocupação quanto à provável existência de viés, variáveis instrumentais ou *proxies* são requeridas. O problema de variáveis omitidas e dependência circular entre covariadas e variável dependente podem ser resolvidos com o uso de variáveis instrumentais (IV – *Instrumental Variables*). Sabe-se que um bom instrumento é aquele correlacionado apenas com a variável endógena e não se relaciona diretamente com a variável explicada.

$$A_{c,t} = Z_{1,T-t}\Phi_1 + Z_{2,T-t}\Phi_2 + Z_{3,T-t}\Phi_2 + P\kappa + \zeta_{c,t} \quad (4.18)$$

Onde, $A_{c,t}$ é a decomposição espacial da aglomeração na área/célula c e $c = 1, 2, 3$ e tempo t , para $t = 2002, \dots, T$, $Z_{1,T-t}$, $Z_{2,T-t}$, $Z_{3,T-t}$ são as variáveis instrumentais defasadas em t períodos, P é a matriz que contém as variáveis de controle variantes no tempo e os efeitos fixos $P = (S_{c,t}, \beta_c, \mu_{k,t}, I_{a,k}, X_{i,t}, \delta_i, \psi_j, \phi_{i,j,c})$ e κ os respectivos coeficientes. Dada a existência de três variáveis exógenas, assumindo a defasagem espacial em duas ordens, a

especificação representada na equação 4.18 será estimada para cada decomposição espacial da aglomeração e, portanto, haverá três primeiros estágios. Os valores preditos $\widehat{Z_{1,t-T}}$, $\widehat{Z_{2,t-T}}$ e $\widehat{Z_{3,t-T}}$ no primeiro estágio são acoplados à equação 4.13 ou equação 4.16, a depender da hipótese de endogeneidade assumida. A equação do segundo estágio a ser estimada é:

$$\log w_{i,t} = \widehat{Z_{1,t-T}} \omega_1 + \widehat{Z_{2,t-T}} \omega_2 + \widehat{Z_{3,t-T}} \omega_3 + P\kappa + \epsilon_{i,t} \quad (4.19)$$

A literatura empírica tem utilizado um conjunto de instrumentos para corrigir a endogeneidade presente nas economias de aglomeração. As estratégias mais comuns consistem em utilizar a defasagem da densidade do emprego e/ou população (CICCONE, 1996; MION; NATICCHIONI, 2005; COMBES, DURANTON, GOBILLON, 2008), variáveis geográficas (ROSENTHAL; STRANGE, 2008; COMBES, DURANTON, GOBILLON, 2008; BARUFI; HADDAD; NEJKAMP, 2016; HADDAD; BARUFI, 2017) e *Bartik Instrumental Variables* (SILVA, 2017) para instrumentalizar as economias de aglomeração correntes.

A estratégia de defasagem da concentração da população, ou do total de emprego, tem por hipótese subjacente que a densidade contemporânea é determinada pela concentração passada. Os trabalhos que consideram as variáveis geográficas utilizam informações sobre a condição do solo (terreno sedimentar, risco de abalo sísmico e risco de deslizamento). Autores que empregam tal estratégia assumem que terrenos sedimentares e áreas com alto risco sísmico ou de deslizamento apresentam menor probabilidade de aglomeração. Distância ao centro histórico da cidade ou à região costeira do país também são recorrentes, visando explorar a concentração de aglomeração em áreas com melhor acesso a meios de transporte, seja intra-urbano (no caso do centro histórico, como em Haddad e Barufi, 2017) ou relativos ao escoamento de produtos na escala nacional ou internacional (como em Barufi, Haddad e Nijkamp, 2016). O instrumento Bartik é recorrente na literatura de Economia do Trabalho. Em linhas gerais, o instrumento busca isolar os choques de demanda local por trabalho dos choques de oferta. O principal canal do instrumento é ativar a demanda por emprego na mesma magnitude do crescimento nacional. O choque de crescimento exclui a área em análise e tem como base uma defasagem longa.

Dois desafios extras devem ser superados para encontrar um bom instrumento para a aglomeração nesse trabalho, além de ser exógeno e parcialmente correlacionado com a variável endógena. Primeiro, a variável deve ser suficientemente granular a tal ponto de adequá-la às células do grid. Nesse contexto, poucas variáveis apresentam tal característica³⁴. Adicionalmente, devido ao fato da especificação econométrica considerar efeitos fixos da célula, nenhum instrumento fixo no tempo (os geográficos, por exemplo) é candidato a resolver a endogeneidade existente. A variável instrumental Bartik, como utilizada em Silva (2017), poderia ser uma solução, todavia a base de dados da RAIS não contém endereços para as observações anteriores a 2002, o que impossibilita a geocodificação dos dados.

Neste trabalho utiliza-se a área construída na RMSP no passado para solucionar o problema da endogeneidade. O instrumento proposto assume que o total de área construída no passado para fins comerciais e habitacionais é uma boa medida de aglomeração. Considerando que as áreas ofertadas pelas construtoras capturam as demandas passadas, contemporâneas e, em alguma medida, a expectativa por espaço, essa estratégia é capaz de relacionar a aglomeração ao total de área construída por km². Pelo lado da demanda, a aglomeração de pessoas está restrita pela oferta, nesse caso. Uma vez que a base de dados é restrita apenas aos trabalhadores cuja atividade laboral não apresenta alta mobilidade espacial, o aumento da aglomeração apresenta um limite superior, que é determinado pela oferta de áreas para alocação desses novos trabalhadores.

Entre o processo de lançamento e apropriação das áreas existe uma defasagem temporal. A fim de adequar tal comportamento do mercado imobiliário, utilizam-se valores temporalmente defasados das construções. Essa estratégia evita que qualquer pressão que possa existir sobre os salários contemporâneos dos profissionais envolvidos direta e indiretamente com o setor da construção civil³⁵ seja pressionado pela demanda

³⁴ Em Andersson, Klaesson, Larsson (2016), os quais utilizam áreas de 1 km², os autores não controlam a potencial endogeneidade constantemente retratada na literatura. Possivelmente, esse resultado derive da dificuldade em obter um instrumento numa escala tão restrita quando se controla pelo efeito fixo da célula.

³⁵ O setor da construção civil é excluído da amostra devido a suas características, como expiado na Seção 3.2.2. Todavia, trabalhadores ligados às atividades com fortes encadeamentos com o setor, tais como engenharia e arquitetura, podem observar pressões salariais em função da demanda. Nesse sentido, a hipótese que se assume é que a defasagem temporal garante que oferta de áreas construídas em períodos anteriores não pressionam o salário contemporâneo das atividades altamente correlacionadas ao setor da construção civil. Nesse caso, o possível problema de causalidade reversa pode ser resolvido.

contemporânea por novas áreas construídas. Nesse sentido, a oferta de áreas construídas locais afeta diretamente e explica uma parcela da aglomeração espacial de emprego. Assim, ao instrumentalizar a densidade interna e as aglomerações de primeira e segunda ordens, por exemplo, no período que abrange o painel, corrige-se tanto a simultaneidade quanto a possível variável omitida.

Uma limitação dessa hipótese está em desconsiderar reformas que ampliem áreas construídas já existentes, reversão de finalidade de áreas residenciais para fins comerciais e a possibilidade de dupla finalidade da mesma área construída, isto é, o trabalhador residindo e trabalhando na mesma área. Essa limitação é de difícil superação pela inexistência de bases de dados que abranjam tais especificidades. Contudo, acredita-se que a oferta de empreendimentos de áreas novas esteja capturando de forma efetiva a demanda por determinada localização espacial e, conseqüentemente, captura a aglomeração.

4.2.3. Base de Dados

Esta sub-seção discute as bases de dados utilizadas nesse capítulo. Primeiramente, discutem-se desde os procedimentos utilizados de limpeza dos microdados da RAIS até a formatação longitudinal balanceada da base de dados. Em seguida são discutidas a origem, limpeza e criação das variáveis instrumentais.

4.2.3.1. Dados em Painel

O modelo de diferencial salarial discutido acima considera *sorting* e *matching*, o que demanda uma estrutura de base de dados que permita controlar a auto-seleção e os retornos decorrentes dos possíveis *matches* entre trabalhador, firma e espaço. Todavia, a base de dados da RAIS contém informações para cada vínculo formal do trabalhador, o que não impede que os trabalhadores apareçam mais do que uma vez na base de dados caso tenham mais que um emprego formal.

A adequação da base de dados ao modelo proposto por Combes, Duranton e Gobillon (2008) requer o ajuste dos microdados da RAIS. Assim, para explorar a estrutura longitudinal dos microdados, algumas etapas de tratamento na base de dados, adicionais às descritas na sub-seção 3.3, são essenciais para assegurar consistência e confiabilidade dos resultados, bem como a formatação em painel dos dados. Os critérios de tratamento aplicados seguem a

abordagem proposta pelo Ministério do Trabalho para o arranjo da RAIS-Migra³⁶ e são utilizados em Barufi (2015) e Silva (2017).

Primeiramente, indivíduos cujo contrato de trabalho é impossível de ser identificado (*missing Ids*) (Etapa I), tenham remuneração mensal³⁷ igual a zero (Etapa II) ou que despendam menos que 20 horas semanais à atividade laboral³⁸ (Etapa III) são excluídos da amostra. Mantêm-se apenas trabalhadores com até cinco tipos de contratos diferentes (Etapa IV). Buscou-se também corrigir possíveis inconsistências quanto ao gênero do trabalhador, excluindo vínculos cujo gênero difere da média do reportado em cada contrato de trabalho (Etapa V). Dentre os vínculos do trabalhador, foram excluídos todos os que apresentavam horas trabalhadas inferior à média (Etapa VI). Visando manter apenas um contrato de trabalho por indivíduo, mantiveram-se os vínculos ativos mais recentes, excluindo os contratos de trabalho que estão abaixo da média em termos de anos, contados a partir da admissão (Etapa VII). Para manter somente um único par trabalhador-firma, aleatorizaram-se os múltiplos contratos e apenas um foi selecionado (Etapa VIII). Finalmente, somente trabalhadores observados ao longo do período 2002-2014 foram mantidos na base de dados utilizada em todas as regressões desse capítulo. Considerando o microdados brutos, excluídos apenas firmas sem trabalhadores, na média, após a aplicação de toda as etapas de ajuste (incluindo as etapas referentes ao georreferenciamento), restaram 64% das observações. Comparando o total de observações resultante dessa etapa de ajuste ao total de observações da base pós-ajuste de geocodificação, em média, houve uma redução adicional de 4%. Em média, cada ano contém 4.132.848 trabalhadores. O Apêndice F traz as informações relativas ao total de trabalhadores que permanecem na base de dados a cada etapa de tratamento aplicada.

Com base nos ajustes iniciais, os microdados anuais foram agregados. A vinculação dos trabalhadores intertemporalmente se deu pela variável-chave PIS³⁹, variável identificadora

³⁶ Trata-se de uma base de dados extraída da RAIS, que permite o acompanhamento da trajetória do trabalhador ao longo do tempo. A base RAIS-Migra não é disponibilizada pelo Ministério do Trabalho desde 2012.

³⁷ Na remuneração considerada pela RAIS incluem salários, ordenados, vencimentos, honorários, vantagens, adicionais e gratificações, excluindo o 13º salário (Ministério do Trabalho, 2015)

³⁸ Considera-se que para os trabalhadores nesse contexto de contrato empregatício difere do regime contratual de trabalho superior a 20 horas.

³⁹ A Lei Complementar nº 7/1970 criou o Programa de Integração Social (PIS). Trata-se de um registro do Governo Federal destinado ao trabalhador privado e operacionalizado pela Caixa Econômica Federal cujo

do trabalho⁴⁰. Em função do objetivo de obter um painel balanceado, foram mantidos apenas trabalhadores presentes ao longo de todos os anos. Adicionalmente, essa estratégia é de substancial importância, pois permite a agregação em um único estágio, dado o elevado consumo de memória que os dados da RAIS apresentam. Como resultado dessas agregações, o painel inicial totaliza 4.935.658 observações, correspondendo a 379.666 trabalhadores e 120.497 firmas.

Após a criação do painel balanceado, inconsistências internas foram observadas e tratadas. Como discutido na sub-seção 3.2.1, a RAIS identifica a firma com base no Cadastro Nacional de Pessoas Jurídicas (CNPJ) ou no Cadastro Específico do Instituto Nacional de Seguro Social (INSS) – CEI. Todavia, erros de preenchimento da declaração ou utilização de mesmo número para identificar matriz e filial dificultam a identificação dos estabelecimentos. Nesse contexto, utilizou-se, após o tratamento de tais inconsistências⁴¹, a composição dos dois identificadores para formar o identificador de firmas. Foram excluídas da base de dados em painel as observações que apresentam variância no gênero do indivíduo, bem como áreas que não continham informações sobre o município. Foram mantidos na amostra apenas trabalhadores entre 18 e 65 anos.

Finalmente, o painel balanceado é composto por 4.573.205 observações, correspondendo a 351.785 trabalhadores e 85.072 firmas no período de 13 anos. Do painel balanceado inicial ao painel balanceado final, 362.453 observações foram excluídas por inconsistências inerentes à própria base de dados da RAIS.

Sobre as covariadas presentes no painel, tem-se características do trabalhador, firma e, especialmente, da célula do grid. Primeiramente, as características do trabalhador presentes na base de dados são escolaridade, experiência, idade (restrita ao intervalo entre 18 e 65

objetivo é permitir o acesso do trabalhador aos benefícios sociais garantidos por lei, tais como seguro-desemprego e Fundo de Garantia do Tempo de Serviço (FGTS) (CAIXA, 2018).

⁴⁰ A escolha pela PIS em detrimento do Cadastro de Pessoa Física (CPF) deriva do fato de que esse apresenta maior número de observações faltantes relativamente às informações sobre o PIS.

⁴¹ Como descrito na Seção 3.2.1, foram excluídas firmas que apresentam similaridades de CNPJ e CEI e firmas que não há informação desses identificadores devido à impossibilidade de identificar os diferentes estabelecimentos ou informações não consistentes. Ver Apêndice B para mais informações quanto à essa etapa de limpeza da base de dados.

anos), gênero e salário-hora. A variável de educação foi dividida em quatro ciclos⁴²: 1) até ensino médio incompleto; 2) ensino médio completo; 3) ensino superior incompleto; e, 4) ensino superior completo e mestrado e/ou doutorado. Neste trabalho considera-se experiência do trabalhador os anos dedicados ao mesmo vínculo, medidos em meses. Buscando compatibilidade com a literatura empírica, construiu-se a variável salário-hora, dividindo o salário total por 30 (total de dias no mês) e pelo total de horas trabalhadas. Para manter a comparabilidade intertemporal do salário-hora, utilizou-se o Índice de Preços ao Consumidor medido pela Fundação Instituto de Pesquisa Econômicas (IPC-Fipe) para a RMSP, assumindo-se 2002 como ano base. Quanto às características da firma, constam informações sobre o tamanho do estabelecimento, setor de atividade da CNAE 1.0, especialização setorial e localizações no município e na célula. O tamanho da firma foi categorizado em nove faixas⁴³ de números de trabalhadores: 1) menor que 4; 2) entre 5 – 9; 3) entre 9 – 20; 4) entre 20 – 50; 5) entre 50 – 100; 6) entre 100 – 250; 7) entre 250 – 500; 8) entre 500 – 1000; e, maior que 1000. Os setores de atividade são os mesmos destacados na subseção 3.2.2, excluindo o setor público por alocar os funcionários na sede do governo (estadual e municipal) e setores cujo desenvolvimento da atividade laboral envolva elevada mobilidade espacial, tais como motoristas de ônibus, táxi, mestres de obras, etc. Criou-se um índice de especialização da atividade, buscando controlar o grau de especialização do setor na economia da Região Metropolitana. O índice nada mais é do que o total de trabalhador do setor s sobre o total de emprego médio na RMSP.

A variável que captura a aglomeração está embasada no grid criado para a RMSP, onde cada célula tem 1 km² de área representa a medida mais fina para aglomeração. Para criação dessa variável, todas as firmas foram sobrepostas ao grid da RMSP e foram somados o total de trabalhadores por célula, para cada ano. A aglomeração de trabalhadores na célula onde o trabalhador i está localizado é classificada como densidade interna (DI). A identificação do

⁴² O primeiro ciclo agrega trabalhadores que está em um desses seguintes ciclos educacionais: analfabetismo, até o 5º ano do ensino fundamental incompleto, 5º ano do ensino fundamental completo, 6º a 9º ano do ensino fundamental incompleto. O último ciclo proposto agrega trabalhadores com níveis educacionais heterogêneos devido a ausência de informação sobre mestrado e doutorado até o ano de 2006.

⁴³ Essa divisão segue o modelo proposto pelo Ministério do Trabalho na base de dados da RAIS que agrega informação da firma.

efeito de atenuação se dá a partir da construção de aglomeração de primeira e de segunda ordem, ou mais. A aglomeração de primeira ordem (APO) é a média espacial do total de trabalhadores em células que dividem fronteiras com a área denominada de densidade interna. A aglomeração de segunda ordem (ASO) é a média espacial do total de trabalhadores nas células que dividem fronteiras com as áreas denominadas APO. A Figura 4.1 demonstra a formatação espacial das variáveis criadas.

Figura 4. 1 – Células do grid e vizinhanças de aglomeração

Fonte: elaboração própria

Na literatura econômica é recorrente o pressuposto que os efeitos do aprendizado ocorrem nas microlocalizações (DURANTON; PUGA, 2004; ANDERSSON, KLAESSON; LARSSON, 2016). O mecanismo subjacente é que trabalhadores com diferentes conjuntos de habilidades apresentam maior probabilidade de interação em áreas espaciais menores *vis-à-vis* àqueles que atuam em áreas espaciais maiores. Nesse sentido, interações entre trabalhadores são mais prováveis de ocorrer numa área de 1 km², podendo promover melhoria nas habilidades dos trabalhadores que respondem diretamente sobre o salário. Segundo Glaeser e Scheinkman (2000), as interações podem ser vistas como trocas que ocorrem em um ambiente sem mercado. Essas trocas são essenciais para o entendimento do benefício da proximidade e se dão pelo que se vê, sente e fala (GLAESER; SCHEINKMAN, 2002), podendo estimular o espriamento da informação boca-a-boca e, por consequência, beneficiar os trabalhadores. Durlouf (2003) aponta, nesse sentido, para a maior relevância do efeito da vizinhança relativamente ao resto da cidade.

Resumidamente, dado que as interações tendem a acontecer com maior probabilidade em áreas menores, espriam-se entre vizinhos e tendem a ter o efeito suavizado no espaço, existiria uma redução de importância do efeito da aglomeração quanto mais distante estiver o local de emprego do trabalhador dessa aglomeração. Portanto, o georrenciamento dos dados

da RAIS e a agregação em áreas pequenas cooperam com a identificação do efeito das economias de aglomeração, possivelmente ativada por contatos face a face entre os trabalhadores, e da atenuação espacial desse efeito no mercado de trabalho formal da RMSP.

4.2.3.2. Variáveis Instrumentais

As bases de dados utilizadas para criação das variáveis instrumentais que medem a urbanização da Região Metropolitana de São Paulo são provenientes da Empresa Brasileira de Estudos do Patrimônio (Embraesp) e aperfeiçoadas pelo Centro de Estudos da Metrópole (CEM)⁴⁴. Essa base de dados é amplamente utilizada em Economia Urbana focada em mercado imobiliário (BIDERMAN, 2001; MACIEL; BIDERMAN, 2013; NADALIN, 2010; CAMPOS, 2016; CAMPOS; ALMEIDA, 2018) devido à sua extensão, abrangendo os 39 municípios da RMSP, qualidade e abrangência da informação quanto à estrutura do imóvel (área útil, área total, quantidade de vagas de garagem, dormitórios e banheiros, quantidade de escritório, número de andares, área total do empreendimento, preço do m², etc.). Outra característica importante da base de dados é trabalhar com imóveis novos. Devido à informação da data de lançamento, é possível considerar a área construída períodos antes de ser entregue. É importante notar que constam na base de dados apenas empreendimentos que foram efetivamente construídos, e não apenas lançados.

Biderman (2001) faz algumas ressalvas sobre a base de dados da Embraesp. Primeiramente, os dados não têm uma característica amostral, ou seja, não existe um plano de amostragem subjacente para a obtenção dos dados. Quanto aos empreendimentos habitacionais (tipologias verticais e horizontais – casas), tratam-se de imóveis disponíveis às famílias com renda mensal superior a R\$ 800,00, a preços de 2000. Isso significa que famílias que vivem em setores subnormais não estão contempladas nessa amostra, pois, considerando o Censo 2000, os moradores dessas regiões possuem renda de até R\$ 755,00 (Nadalin, 2010). Essa limitação da base de dados da Embraesp impõe pouco atrito ao objetivo aqui proposto, pois as firmas

⁴⁴ Os aperfeiçoamentos derivados da intervenção dos pesquisadores do CEM são georreferenciamento da base de dados, correção e atualização monetária, padronização dos nomes das empresas incorporadoras de cada empreendimento (CEM, 2018).

nessas regiões geralmente não firmam contratos formais com o trabalhador. Logo, não são aderentes aos dados da RAIS.

Os dados representam apenas uma parcela dos imóveis lançados na RMSP e abrangem apenas o mercado formal, o que, em alguma medida, limita a captura perfeita da aglomeração por via de área construída. Contudo, dado o fato de considerarmos apenas o emprego formal, é possível que o efeito pernicioso de baixa aderência seja suavizado ou reduzido. Outra limitação a destacar é que os dados fornecidos pela Embraesp capturam o fluxo de imóveis e não o estoque. Tendo em vista essa limitação, a variável de estoque de área construída foi criada a partir da compilação de empreendimentos lançados considerando sempre o período de 10 anos. Essa estratégia, implicitamente, considera uma depreciação de 10% a cada década.

É de suma importância em um contexto de dados espaciais agregados numa escala ultrafina, como o utilizado nesta tese, a obtenção de variáveis-chave que contenham informações capazes de associar os dados às escalas espaciais criadas. Nesse sentido, as variáveis latitude e longitude que constam na base de dados ao longo dos anos são essenciais para o desenvolvimento do instrumento, permitindo qualquer agregação espacial dos dados. O georreferenciamento dos empreendimentos foi dirigido pelo CEM. Os pesquisadores utilizaram o *software* Map após a normalização dos endereços. A estrutura do *address locator* proposta considera, simultaneamente, CEP de quatro dígitos, logradouro e número. Como discutido acima, essa estrutura de geocodificação evita erro de medida, isto é, alocar espacialmente de forma errônea o empreendimento.

Para o cálculo da densidade de área construída, plotaram-se os dados sobre as células do grid com área de 1 km² e somaram-se as áreas totais construídas dos empreendimentos. Para a construção das áreas vizinhas de primeira ordem, somam-se todas as áreas totais construídas dividindo pelo total de células que compõem a vizinhança. O mesmo procedimento é aplicado para os vizinhos de segunda ordem.

Os dados limitam-se ao período de 1985-2007. Para instrumentalizar a aglomeração do ano de 2002 foram utilizadas as áreas construídas entre 1985-1995, enquanto, para instrumentalizar a aglomeração em 2003, utilizaram-se as áreas construídas de 1986-1996.

Para os demais anos, segue-se a mesma lógica. Tal estratégia de instrumentalização assume uma depreciação de 10% do estoque de área construída na RMSP e elimina a possível correlação contemporânea entre os salários de trabalhadores cuja atividade esteja vinculada com o setor da construção civil.

Em função da possibilidade de dupla finalidade das áreas residenciais construídas (fins residenciais e/ou comerciais), utilizam-se as informações de áreas construídas para fins residenciais para instrumentalizar a aglomeração de trabalhadores. Ademais, sob essa estratégia, assume-se uma métrica de aglomeração mais ampla, isto é, aglomeração da população inclusive. A estratégia empírica considera no primeiro estágio do estimador 2SLS-IV, o total de área construída para fins comerciais e residências, separadamente, tornando o modelo sobreidentificado. A base de dados apresenta um total de 12.159 empreendimentos lançados, sendo 11.520 residenciais e 639 comerciais, com média anual de 935 empreendimentos lançados.

4.2.4. Estratégia de Identificação e Métodos de Estimação

Assuma-se que N^* , I , J , C e S representam o total de observações, trabalhadores, firmas, células dos grids e indústrias na base dados em painel, respectivamente. $Q \leq SC$ é o total de *match* indústria-célula e tem dimensão menor ou igual que a interação indústria-célula; $M \leq NJC$ é o número total de *matches* trabalhador-firma-célula; k representa o total de variáveis N_i , N_j e N_m representa o total de trabalhador i , firma j e *match* trabalhador-firma-célula m , respectivamente. Reescrevendo a Equação 4.17 em notação matricial

$$y = a + B\beta + C\theta + F\psi + H\tau + L\zeta + G\phi + \varepsilon \quad (4.20)$$

Onde $y = \ln w_{it}$ é um vetor $N^* \times 1$ do log do salário-hora; B é uma matriz $N^* \times k$ contendo covariadas (características do trabalhador, variáveis de aglomeração e especialização, *dummy* de tempo e de municípios); β é um vetor de parâmetros a serem estimados, com dimensão $k \times 1$; C é uma matriz $N^* \times N$ contendo variáveis indicadoras dos trabalhadores (captura *quem ele é*); θ é um vetor $N \times 1$ de efeitos fixos dos trabalhadores; F é uma matriz $N^* \times J$ contendo variáveis indicadoras da firma (captura *onde ele trabalha*); ψ é um vetor de dimensão $J \times 1$ de efeitos fixos da firma; H é uma matriz $N^* \times C$ matrix contendo variáveis indicadoras de célula (captura *onde está está espacialmente localizado*); τ é um vetor $C \times 1$ de efeitos fixos de

célula; L é uma matriz de dimensão $N \times SC$ contendo variáveis indicadores da interação indústria-célula (captura o *setor e a localização espacial em que ele está*); ζ é um vetor de dimensão $SC \times 1$ de efeitos fixos da interação indústria-célula; G é uma matriz $N \times M$ contendo variáveis indicadoras do *match* trabalhador-firma-célula (captura *quem ele é, onde trabalha e onde está espacialmente localizado*); ϕ é um vetor $M \times 1$ de efeitos fixos do *match*; e ε é um vetor $N \times 1$ do termo de erro.

A base de dados em painel utilizada nesse trabalho permite a utilização de um conjunto de diferentes estimadores econométricos, como Mínimos Quadrados Ordinário Empilhados (POLS – *Pooled Ordinary Least Squares*), Efeitos Aleatórios (RE – *Random Effects*) e Efeitos Fixos (FE – *Fixed Effects*).

O estimador POLS não controla por características não observadas constantes no tempo. Na ausência de controles de efeitos fixos dos indivíduos, o viés está presente e aninhado no tempo de erro, ε^{POLS} . A partir das equações 4.19 e 4.20, nota-se que os coeficientes estimados serão enviesados, dado que a condição de ortogonalidade entre o termo de erro e as covariadas não vale.

$$E(\varepsilon^{POLS} | C, F, H, L, G, B) \neq 0 \quad (4.21)$$

$$E(BCFHLG) \neq 0 \quad (4.22)$$

De acordo com Wooldridge (2010), o estimador RE é consistente se valem as seguintes condições: $(\varepsilon^{RE} | B, C, F, H, L, G) = 0$ e $E(BCFHLG) = 0$. A principal questão é que a auto seleção do trabalhador e da firma são características não observadas e são correlacionadas com a concentração espacial de trabalhador, a medida de aglomeração nesse trabalho. Logo, tal estimador não é recomendável para essa análise em específico.

Ao relaxar a hipótese de exogeneidade estrita, isto é, $E(C, F, H, L, G | B) \neq 0$, o modelo passa a ser estimado por FE. A principal hipótese é a expectativa condicional do termo de erro nulo, tal como $(\varepsilon^{FE} | C, F, H, L, G, X) = 0$. Tal pressuposto é equivalente a assumir que a mobilidade dos trabalhadores é condicionalmente nula. Em outras palavras, a mobilidade do trabalhador depende de características observáveis dos trabalhadores, firmas e efeitos fixos mencionados acima, mas não do termo de erro.

Para a identificação dos efeitos fixos do trabalhador é requerida covariação condicional entre os salários dos trabalhadores em diferentes firmas. Da mesma forma, a identificação dos efeitos fixos de firma está embasada no componente comum dos salários entre diferentes trabalhadores das firmas, enquanto o efeito fixo da célula depende da mobilidade dos trabalhadores entre áreas. Os modelos econométricos que seguem a tradição Abowd, Kramarz e Margolis (1999) não se preocupam quanto ao papel do efeito fixo do match trabalhador-firma. De acordo com Woodcock (2015), a omissão do efeito do match enviesaria o estimador de mínimos quadrados.

A especificação econométrica utilizada expande a versão proposta por Abowd, Kramarz e Margolis (1999) ao adicionar os efeitos fixos dos *matches* indústria-área trabalhador-firma-área. O primeiro grupo de efeitos fixos busca controlar as economias de especialização e o segundo controla pelos choques derivados de melhores associações entre trabalhadores, firmas e localização espacial ou más associações, como maus trabalhadores trabalhando em ótimas firmas, por exemplo. A identificação do primeiro *match* requer mobilidade do trabalhador entre a associação indústria-área, enquanto a identificação do segundo *match* depende do componente comum dos salários entre a tripla interação.

Pressupostos adicionais são necessários para identificar os efeitos da tripla associação. Quando estimado em um modelo econométrico de um único estágio, o *match* é identificado desde que a mobilidade dos trabalhadores não dependa de heterogeneidade não observada do *match* (WOODCOCK, 2015). Utilizando a abordagem de FE, a hipótese adicional é que os efeitos fixos da tripla interação sejam ortogonais aos efeitos fixos de trabalhador, firma, área e indústria-área, $E[\phi_{ijc} | C_i, F_j, H_c, L_{sc}] = 0$ – onde, C_i é o indicador para o trabalhador i , F_j é o indicador para firma j , H_c é o indicador para a área c e L_{sc} é o indicador para a interação da indústria-área sc . O efeito do *match* é sempre identificado uma vez que os correspondentes efeitos de trabalhador, firma e área sejam identificados em um modelo sem a tripla interação (WOODCOCK, 2015). Embora a hipótese discutida seja forte, a omissão do *match* enviesará a estimação dos coeficientes $\beta, \theta, \psi, \tau$ e ζ .

Considerando a abordagem FE e assumindo que as condições expostas na Equação 4.12 e Equação 4.14 não valham, o estimador de Mínimos Quadrados Generalizados Factíveis

(FGLS – *Feaseble Generalized Least Square*) é o estimador utilizado para a estimar a especificação econométrica exposta na Equação 4.10.

Considerando a endogeneidade inerente à aglomeração de empregos, os coeficientes que capturam o prêmio salarial intra-urbano e sua atenuação são enviesados. Logo, as equações 4.12 e 4.14 valem, então a hipótese de exogeneidade não é observada e $[\varepsilon^{FE} | D, F, H, L, G, X] \neq 0$. Para lidar com os vieses, variáveis instrumentais que sejam correlacionadas com aglomeração, mas não relacionadas com o salário, são necessárias. Nesse contexto, as condições padrão para obter a estimação consistente do vetor $\Phi = (\Phi_1, \Phi_2, \Phi_3)$, utilizando o estimador de Mínimos Quadrados de 2-Estágios (2SLS – *2 Stages Least Squares*), são $\Phi \neq 0$ no primeiro estágio e $E[Z_{1,t-T} \cdot \zeta_{c,t} | P] = 0$. Na Subseção 4.3.1 demonstra-se que a primeira condição é facilmente observada a partir das regressões de primeiro estágio. A segunda condição requer que o termo de erro da regressão da variável instrumental (primeiro estágio) seja não correlacionado com os instrumentos utilizados. Essa condição seria violada se de alguma forma a área construída defasada em sete anos fosse correlacionada com os salários contemporâneos dos trabalhadores.

A segunda condição não é explicitamente testável, mas a base de dados e o desenho empírico proposto permitem fazer alguns testes indiretos. A primeira estratégia consiste em demonstrar que a aglomeração contemporânea é explicada pela defasagem temporal das áreas construídas, mas o inverso não é verdadeiro. Outra estratégia aplicada consiste em instrumentalizar diretamente a variável endógena no modelo *baseline*, apontando para o baixo impacto sobre os salários contemporâneos.

O primeiro estágio é estimado seguindo a especificação destacada na Equação 4.20. Utilizam-se áreas construídas com defasagem temporal para o nível de aglomeração de emprego contemporâneo. Nesse estágio, estimam-se as parcelas da aglomeração de emprego previstas pelo instrumento. No segundo estágio, como demonstrado na Equação 4.21, utilizam-se os valores preditos do primeiro estágio para estimar o impacto da aglomeração sobre a taxa de remuneração dos trabalhadores.

A abordagem de FE é aplicada seguindo a literatura empírica em Economia de Aglomeração e Economia do Trabalho. Assim, variáveis de controle constantes no tempo são excluídas da especificação econométrica, tais como gênero, idade (WOOLDRIDGE, 2010). A principal

contribuição do modelo econométrico proposto está relacionada ao conjunto de controles de efeitos fixos.

Modelos de FE podem ser estimados considerando duas abordagens. A primeira abordagem assume uma transformação *within* intra grupo (por trabalhador), eliminando características não observadas constantes no tempo (WOOLDRIDGE, 2010). Abowd, Kramarz e Magnolis (1999) abordaram o problema relacionado à regressão linear de efeitos fixos com elevada dimensão. A segunda abordagem considera a expansão do modelo de Abowd, Kramarz e Magnolis (1999) utilizando as interações do ponto fixo, enquanto controla por um conjunto de efeitos fixos. Na prática, este procedimento considera toda variabilidade entre grupos ao adicionar ao conjunto de regressores algumas variáveis *dummies* que observam a heterogeneidade específica do grupo (GUIMARÃES; PORTUGAL, 2010). Devido à quantidade de efeitos fixos para serem estimados nesse trabalho, a segunda abordagem é preferida.

Na abordagem iterativa dos efeitos fixos, ao invés de adicionar os efeitos fixos explicitamente e covariadas em uma única matriz, Guimarães e Portugal (2010) propõem uma partição em duas submatrizes que capturam as covariadas em B e variáveis indicadores. Tal procedimento generaliza a abordagem *within* utilizando a aplicação iterada do Teorema Frishch-Waugh-Lovell (FWL). Correia (2014) desenvolveu um algoritmo no Stata seguindo Guimarães e Portugal (2010). Nesse contexto, a solução do problema dessa tese envolve estimações de equações lineares tais como

$$\beta = (B'B)^{-1}(B'(y - C\theta - F\psi - H\tau - L\zeta - G\phi)) \quad (4.23)$$

$$\theta = (C'C)^{-1}(C'(y - F\psi - H\tau - L\zeta - G\phi)) \quad (4.24)$$

$$\psi = (F'F)^{-1}(F'(y - C\theta - H\tau - L\zeta - G\phi)) \quad (4.25)$$

$$\tau = (H'H)^{-1}(H'(y - C\theta - F\psi - L\zeta - G\phi)) \quad (4.26)$$

$$\zeta = (H'H)^{-1}(H'(y - C\theta - F\psi - H\tau - G\phi)) \quad (4.27)$$

$$\phi = (H'H)^{-1}(H'(y - C\theta - F\psi - H\tau - L\zeta)) \quad (4.28)$$

Segundo Guimarães e Portugal (2010), a interação entre esse conjunto de equações fornece a solução exata utilizando OLS. Sob suspeita de heterocedasticidade, o modelo pode ser estimado considerando o estimador FGLS.

4.3. RESULTADOS

Essa subseção apresenta as características do painel balanceado criado a partir dos microdados da RAIS. Mais adiante, a partir da estimação econométrica na forma reduzida e com variáveis instrumentais, testa-se a hipótese inicialmente levantada de atenuação espacial das economias de aglomeração. Devido à possibilidade de que contatos face a face ocorram em áreas menores, conduziram-se testes de sensibilidade ao estimar o modelo *baseline* considerando células menores e maiores. Em seguida, avaliam-se as diferentes habilidades de absorção do efeito gerado pela economia de aglomeração em função do nível educacional do trabalhador. Outro ponto também discutido é se a hipótese de atenuação é observada nos subcentros identificados no terceiro capítulo desta tese. Para tanto, testou-se o modelo *baseline* para os SBD^{Q99} , SBD^{Q95} , SBD^{Q90} e para locais de emprego fora das subcentralidades (*NSBD*).

4.3.1. Características do Painel e das Variáveis Instrumentais

Nesta subseção, as estatísticas descritivas e estimações econométricas derivam da base de microdados da RAIS (2002-2014) com os ajustes discutidos na subseção 4.2.3. A Tabela 4.1 apresenta o primeiro e o segundo momentos da distribuição das variáveis não categóricas, bem como máximos e mínimos. Na Tabela 4.3 são apresentadas frequências e totais de observações para as variáveis categóricas. É importante notar que a presente análise se refere a um banco de dados longitudinal balanceado.

Para capturar o efeito real do diferencial salarial utilizou-se a variável salário-hora deflacionada pelo Índice de Preços ao Consumidor (IPC-Fipe) calculado pela Fundação Instituto de Pesquisas Econômicas (Fipe), a preços de 2014, referente à RMSP. Assim, o salário-hora médio do trabalhador formal da amostra é de R\$ R\$ 14,69. A idade média do trabalhador é de 39 anos, dada a pré-seleção de trabalhadores entre 18 e 65 anos. O período de vinculação à mesma firma (experiência) é de 111 meses, com elevada dispersão (desvio-padrão igual a 83 meses).

Tabela 4. 1 – Estatística descritiva das variáveis

Variável	Média	Desvio	Min.	Max.
Ln(salário-hora)	1,97	1,198	-2,85	6,74
Salário-hora	14,69	24,262	0,058	845
Densidade interna (Emp)	10.671	14.104	1	115.546
Densidade 1ª ordem (WEmp)	7.235	7.815	0	44.642,38
Densidade 2ª ordem (W2Emp)	6.112	5.954	0	26.821,5
Experiência	111,24	83,26	0	598,9
Experiência ²	19.307	26.516	0	358.681,2
Idade	39,10	9,05	18,00	65,00
Idade ²	1.610	731	324	4.225
Especialização	1	1,1383	0,00013	14

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

A média de densidade interna é de 10.671 trabalhadores por célula. Esse valor decai ao comparar a aglomeração média dos vizinhos de primeira (*WEmp*) e segunda ordem (*W2Emp*), os quais apresentam totais de trabalhadores 68% e 57% menores relativamente à variável *Emp*, respectivamente. A Tabela 4.2 relaciona a média salarial da célula com a aglomeração interna e aglomerações localizadas nos vizinhos, demonstrando uma correlação positiva entre as variáveis, embora declinante. Como se nota, quanto maior o distanciamento espacial da aglomeração em relação ao posto de trabalho do indivíduo, menor é a relação entre aglomeração e salário.

Tabela 4.2 – Matriz de correlação do salário real e densidade de emprego

	Salário	Em	WEmp	W2Emp	ACT	WACT	W2ACT	ACT	WACT	W2ACT
Salário	1									
Emp	0,15	1								
WEmp	0,14	0,74	1							
W2Emp	0,13	0,68	0,89	1						
ACTC	0,09	0,39	0,44	0,40	1					
WACTC	0,11	0,48	0,65	0,63	0,74	1				
W2ACT	0,08	0,48	0,65	0,73	0,44	0,70	1			
ACTH	0,11	0,32	0,44	0,49	0,47	0,55	0,48	1		
WACTH	0,17	0,40	0,59	0,63	0,50	0,68	0,63	0,73	1	
W2ACTH	0,21	0,48	0,62	0,71	0,43	0,62	0,72	0,59	0,82	1

Fonte: Elaboração própria a partir dos dados da RAIS (2002-2014) e Embraesp (1985-2007). Nota: ACTC e ACTH refere-se à área construída total para fins comerciais e para fins habitacionais, respectivamente. W refere-se à defasagem de primeira ordem e W², à defasagem de segunda ordem.

Focando em análise da estrutura salarial e considerando categorias de referências, é possível examinar os rendimentos dos trabalhadores no mercado formal de trabalho na RMSP. A Tabela 4.3 mostra que os trabalhadores do gênero masculino são a maioria na base de dados, alcançando a parcela de 63% do total de indivíduos e auferem 12% a mais quando comparados à média recebida pelas mulheres (R\$ 13,62). Observa-se um padrão de crescimento do salário-hora médio em função do grau de escolaridade dos indivíduos e que

os saltos salariais são cada vez mais pronunciados para cada mudança de nível de grau de instrução. Trabalhadores que compõem a categoria *educ1* representam 40% da base de dados e apresentam a menor média do salário-hora, em torno de R\$ 7,26. Indivíduos com ao menos ensino superior completo, categoria *educ4* (23% dos trabalhadores da base de dados), auferem salário-hora médio de R\$ 31,94. Na categoria *educ2* (31% dos trabalhadores da base de dados), o salário-hora médio é de R\$ 10,87, e o salário-hora médio na categoria *educ3* (6% dos trabalhadores da base de dados) é de R\$ 16,76. O salário médio cresce em função da idade dos trabalhadores, todavia o crescimento é mais acentuado a cada mudança de faixa etária que compõe as categorias iniciais, passando a ter crescimento menor para indivíduos que compõem as faixas etárias mais elevadas (*idad4* e *idad5*). A experiência é outra característica relevante do trabalhador no mercado de trabalho, todavia apresenta uma distribuição pouco óbvia em relação ao salário médio, como se observa na Tabela 4.3.

Em se considerando a classificação criada no capítulo anterior, a Tabela 4.3 mostra que 33% dos indivíduos estão concentrados em células classificadas como SBD^{Q99} (i.e., células 1% mais concentradoras de trabalhadores), enquanto 37% e 17% dos indivíduos estão concentrados em células do tipo SBD^{Q95} e SBD^{Q90} , respectivamente. O salário-hora médio dos trabalhadores na região com maior concentração de trabalhadores é R\$ 17,39, enquanto indivíduos alocados em áreas com menor aglomeração apresentam médias salariais menores. Trabalhadores fora das aglomerações predefinidas apresentam média de salário hora 62% menor relativamente aos indivíduos do SBD^{Q99} .

Ao analisar as características das firmas e sua relação na média salarial dos indivíduos, observa-se que quanto maior as firmas, maior é a média salário-hora pago aos trabalhadores. Enquanto firmas com até 4 funcionários apresentam média salário-hora igual a R\$ 6,67, firmas com mais de 1000 funcionários pagam 3 vezes mais. O setor de comércio e administração de imóveis apresenta maior concentração de trabalhadores no painel utilizado (19% do total das observações), seguido pelo comércio varejista (12%) e serviços médicos, odontológicos e veterinários (9%). O maior salário-hora médio real está associado às atividades desenvolvidas no setor de instituições financeiras, cerca de R\$ 26,00.

Tabela 4. 3 – Renda e distribuição dos trabalhadores do mercado formal por características

Variável	Média	Desvio	continua	
			Obs.	%
<i>Subcentros</i>				
sbd99	17,39	29,591	1.530.92	33%
sbd95	14,61	22,816	1.708.34	37%
sbd90	12,34	19,316	777.266	17%
Fora de qualquer SBD	10,83	16,33	556.667	12%
<i>Gênero</i>				
Homem	15,324	25,5681	2.876.86	63%
Mulher	13,623	21,8282	1.696.35	37%
<i>Educação</i>				
educ1	7,24	9,198	1.813.48	40%
educ2	10,87	14,086	1.432.22	31%
educ3	16,76	23,609	257.504	6%
educ4	31,94	39,503	1.069.98	23%
<i>Idade</i>				
idad1	3,34	4,402	194.613	4%
idad2	8,29	12,020	514.940	11%
idad3	13,68	21,001	1.735.55	38%
idad4	17,36	27,636	1.481.90	32%
idad5	19,80	31,501	646.226	14%
<i>Experiência</i>				
exp1	13,27	23,42	368.572	8%
exp2	12,13	22,87	304.921	7%
exp3	11,08	21,13	295.761	6%
exp4	11,43	21,26	279.991	6%
exp5	10,84	20,13	267.474	6%
exp6	10,53	18,91	259.318	6%
exp7	12,32	21,57	250.057	5%
exp8	13,94	23,27	244.406	5%
exp9	17,38	26,24	2.302.71	50%
<i>Tamanho da Firma</i>				
1-5 trabalhadores	6,63	10,25	232.441	12%
5-10 trabalhadores	7,858	11,84	411.692	8%
10-20 trabalhadores	10,07	16,84	410.991	8%
20-50 trabalhadores	11,76	19,77	631.733	13%
50-100 trabalhadores	13,62	22,79	498.796	10%
100-250 trabalhadores	16,38	26,11	622.824	13%
250-500 trabalhadores	18,28	27,77	549.298	11%
500-1000 trabalhadores	18,5	27,98	437.052	9%
≥ 1000 trabalhadores	20,18	30,19	778.378	16%
<i>Setor de Atividade Econômica</i>				
Comércio, administração de imóveis, valores	12,94	23,88	884.546	19%
Agricultura, silvicultura, extrativista	15,41	31,56	4.511	0%
Alimentação e Bebidas	9,59	19,08	223.497	5%
Alojamento e Alimentação	14,5	24,67	243.750	5%
Borracha Fumo e Couro	10,71	18,92	63.291	1%
Comércio atacadista	16,53	28,28	231.873	5%
Comércio varejista	9,58	17,88	558.412	12%
Eletrônicos e Comunicação	14,86	23,12	78.604	2%
Ensino	20,32	28,00	287.380	6%
Extrativa mineral	22,33	40,42	4.772	0%
Indústria calçadista	6,96	10,17	5.757	0%
Indústria mecânica	17,42	24,30	147.347	3%

Variável	Média	Desvio	conclusão	
			Obs.	%
Indústria metalúrgica	11,69	15,96	189.485	4%
Indústria química	17,3	26,92	311.788	7%
Indústria têxtil	8,6	14,58	131.829	3%
Instituições financeiras	26,52	36,30	255.661	6%
Madeira e Mobiliário	13,49	25,61	6.192	0%
Indústria do material de transporte	22,3	27,00	236.191	5%
Médicos, Odontol. e Veterinária	12,57	18,70	428.759	9%
Indústria minerais não metálicos	11,06	15,16	36.101	1%
Papel e Gráfica	17,16	25,71	135.314	3%
Transportes e Comunicações	13,25	20,21	108.145	2%
Observações (N x T)			4.573.20	100

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

As figuras 4.2 e 4.3 fornecem uma descrição básica da relação entre salários médios e densidade interna e dos vizinhos de primeira e segunda ordem. A partir da Figura 4.2 observa-se uma relação positiva entre aglomeração interna e a média salarial, para um ajuste linear. Em outras palavras, o salário do trabalhador cresce com a aglomeração. A parcela do salário explicada pela concentração interna é de 4,8%. Na Figura 4.3, os painéis 1 e 2 apresentam associações positivas da densidade dos vizinhos de primeira e segunda ordem com o log do salário real do trabalhador. Considerando o R^2 ajustado, nota-se uma redução da parcela do salário que é explicada pela densidade interna e dos vizinhos de primeira e segunda ordem.

A semi-elasticidade derivada dessa simples regressão linear é de 1,51%, considerando um aumento de 100.000 trabalhadores na área em que se localiza o posto de trabalho do indivíduo. As curvas estimadas para as aglomerações vizinhas são mais inclinadas, sugerindo uma tendência de efeito positivo e crescente no espaço; isto é, semi-elasticidades de 2,59% e 3,18% (Painéis 1 e 2, Figura 4.3).

Figura 4. 2 – Salário médio real e densidade de emprego nas células (2002-2014)

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

Nota: Salário deflacionado pelo IPC-Fipe

Figura 4. 3 – Salário médio real e densidade da defasagem de emprego nas células (2002-2014)

Painel 1: Densidade (1ª ordem – PO)

Painel 2: Densidade (2ª ordem – SO)

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

Nota: Salário deflacionado pelo IPC-Fipe

As relações de aglomeração atuam de forma simultânea. Assim, para capturar o efeito da aglomeração “limpo” das correlações entre densidade do emprego interna, vizinhos de primeira e segunda ordem sobre os salários, estimaram-se duas regressões lineares (OLS). A Figura 4.4 mostra os efeitos marginais da expansão da aglomeração (normalizada para 100.000 trabalhadores – Painel 1) e para trabalhadores localizados nos diferentes SBD. A relação da aglomeração no espaço sugere uma relação de atenuação do salário, não

importando a forma de mensurar o efeito da aglomeração. Em termos de magnitude, os coeficientes estimados são menores relativamente à especificação estimada acima.

Figura 4. 4 – Regressão linear do salário-hora e densidade de emprego
Painel 1: Regressão Linear 1 **Painel 2: Regressão Linear 2**

Fonte: Elaboração própria a partir dos microdados da RAIS. Nota: coeficientes derivados de estimador OLS.

A especificação econométrica completa exige o controle por múltiplos efeitos fixos. A identificação dos efeitos fixos depende da mobilidade de indivíduos e firmas entre diferentes grupos de aninhamento. Para identificar efeito fixo de indivíduo uma parcela desses trabalhadores deve se mover entre firmas e células, enquanto para identificar características não observadas fixas das firmas é necessário que uma parcela das firmas se mova entre as células. A Tabela 4.4 traz informação sobre essas mobilidades. A mobilidade do trabalhador entre firmas representa 15% do total de trabalhadores da base e entre células, 17% do total. A mobilidade das firmas é menor, atingindo 3% do total de trabalhadores. Essas parcelas informam que, ao longo do período 2002-2014, trabalhadores mudaram de firmas e células e que firmas se deslocaram sobre a área metropolitana de São Paulo. Sobre os trabalhadores que se deslocaram, a média salário-hora é relativamente menor que a média salarial da base de dados completa.

Tabela 4. 4 – Mobilidade da firma e trabalhador

	Salário-Hora	Salário-Hora	Observações	Participação
Mobilidade da Firma				
entre células	6,24	14,445	11.345	10
Total de Firmas (ano)			112.400	100
Mobilidade do Trabalhador				
entre firmas	9,07	19,870	53.265	11
entre células	9,36	19,619	60.276	13
Total de Trabalhadores (ano)			465.261	100

Fonte: Elaboração própria, a partir dos dados da RAIS (2002-2014). Nota: Total de observações é a média anual ao longo do painel.

O controle de características não observadas não é suficiente para lidar com a questão da endogeneidade decorrente da causalidade reversa entre aglomeração e salários. Choques locais positivos podem atrair mais trabalhadores para determinada área da RMSP e, por conseguinte, ampliar o total de trabalhadores aglomerados. Assim, desconsiderar a endogeneidade incorre em má identificação do real efeito da aglomeração sobre o salário-hora médio do trabalhador, culminando em estimativas inconsistentes.

Como discutido na subseção acima, para a criação de variáveis instrumentais recorre-se à base de dados da Embraesp (1985-2007), agregando o total de área construída por célula do grid da RMSP. Duas medidas de áreas construídas foram criadas. Uma agrega o total de área construída para fins comerciais e a outra para fins residenciais.

A Tabela 4.5 apresenta o total de área construída ao longo de 22 anos, considerando o valor agregado de toda a produção. Nesse período foram construídas 111,3 mil m² em toda a Região Metropolitana. Desse total, 5,9 mil m² para fins comerciais e 105,3 mil m² para fins residenciais. Embora o agregado da área total construída na RMSP com a função inicial de moradia seja superior à área total construída para fins comerciais, a média da área construída para fins comerciais por célula é superior. O total de área construída para ambas finalidades corresponde ao total de 12.159 empreendimentos lançados na RMSP.

Tabela 4. 5 – Estatística de Descritiva da Área Total Construída (m²)

Tipo	Área	Empreendimentos	Média ¹	Desvio ¹	Min ¹	Max ¹
Comercial	5.935.943	639	9.289	10.317	174	124.318
Residencial	105.328.677	11.520	9.143	11.432	105	438.741

Fonte: elaboração própria a partir dos dados da Embraesp (1985-2007). Nota: ¹ estatísticas derivadas dos valores agregados por célula.

Para compatibilização dos dados ao nível espacial das células do grid foram construídas densidades de área total de construção, agregando as variáveis e pareando suas respectivas

células à base de dados longitudinal da RAIS (2002-2014). A Tabela 4.6 apresenta as estatísticas descritivas das variáveis instrumentais. As variáveis que medem o total de área construída e suas respectivas defasagens espaciais apresentam um comportamento decrescente do valor médio à medida em que a defasagem espacial aumenta, seguindo o mesmo comportamento da densidade de emprego. Na Tabela 4.2 também são apresentadas as correlações das variáveis instrumentais e emprego. Dessa relação não é possível encontrar um padrão claro entre a densidade de emprego interna e suas médias espaciais com a densidade interna das áreas construídas e suas respectivas áreas construídas. Todavia, a correlação entre essas variáveis é no mínimo de 40%. Como se espera, a correlação entre as áreas construídas (densidade interna e médias espaciais) para fins comerciais e residências é elevada, sendo a menor correlação de 43%.

Tabela 4. 6 – Estatística descritiva das Variáveis Instrumentais

Variáveis	Média	Desvio	Min	Max
ACTC	12.113	29.789	0	235.854
WACTC	8.618	14.690	0	89.116
W2ACTC	7.034	10.213	0	55.387
ACTH	65.275	102.180	0	1.253.178
WACTH	58.383	65.997	0	444.162
W2ACTH	55.510	54.607	0	304.023

Elaboração própria a partir dos dados Embraesp (1985-2007). Nota: ACTC e ACTH referem-se às áreas construídas para fins comerciais e habitacionais, respectivamente. W refere-se à defasagem de primeira ordem e W2, à defasagem de segunda ordem.

A Figura 4.4 traz a disposição espacial das áreas construídas para fins habitacionais (Painel 1) e comerciais (Painel 2), isto é, ACTH e ACTC, respectivamente. Considerando o Painel 1, o valor máximo da área construída por km² (1.253.178 m²) está localizado no distrito de Vila Andrade. Esse distrito é fronteiro do distrito de Itaim Bibi, onde se localiza o valor máximo de área construída para fins comerciais (235.854 m²), como demonstra no Painel 2.

Figura 4. 5 – Distribuição espacial da área construída na RMSP

Painel 1: Área Construída para Fins Habitacionais (1985-2007)

Painel 2: Área Construída para Fins Comerciais (1985-2007)

Fonte: Elaboração própria a partir dos microdados da Embraesp para habitação (Painel 1) e comercial (Painel 2).

O modelo a ser estimado apresenta três variáveis endógenas (*Emp*, *WEmp* e *W2Emp*), que serão instrumentalizadas pelas defasagens temporais das áreas construídas para fins habitacionais (*ACTH*, *WACTH*, *W2ACTH*) e para fins comerciais (*ACTC*, *WACTC*, *W2ACTC*). As seis variáveis instrumentais estão positivamente correlacionadas com cada variável endógena, como apresenta a Tabela 4.2.

A correlação média calculada entre *Emp* e as áreas construídas para fins comerciais é de 0,45 enquanto para fins habitacionais é de 0,40. O comportamento de cada variável instrumental com *Emp* não é conclusivo, e a correlação é ampliada à medida que a distância aumenta. As correlações médias entre *WEmp* e as variáveis instrumentais para fins comerciais e habitacionais são 0,58 e 0,55, respectivamente. A correlação entre *WEmp* e *WACTC* é maior quando comparada com as demais e a correlação entre *WEmp* e áreas construídas para fins habitacionais segue uma tendência crescente com as defasagens espaciais. As variáveis de área construída e *W2Emp* apresentam correlações médias de 0,58 (fins comerciais) e 0,60 (fins habitacionais). Entre as variáveis de área construída para fins comerciais, a maior correlação é observada entre *W2Emp* e *W2ACTC* (0,73) e entre as variáveis de área construída para fins habitacionais, *W2Emp* e *W2ACTH* (0,71).

4.3.2. Testando a Hipótese de Atenuação Espacial

O objetivo principal é responder em que medida os salários dos indivíduos seriam afetados pela aglomeração de trabalhadores. Em função da escala espacial fina utilizada para a estimação dos modelos econométricos, o efeito da aglomeração sobre os salários dos trabalhadores poderia ser ativado por relações face a face entre os trabalhadores. Testa-se a hipótese de que os efeitos positivos sobre os salários seriam atenuados à medida que aglomerações mais longínquas são consideradas na especificação econométrica. No contexto intra-urbano, aglomerações localizadas em áreas mais distantes do posto de trabalho do indivíduo tendem a impor atrito às relações face a face, dado que os trabalhadores estão situados em vizinhanças.

O primeiro desafio que se coloca é a proposição de uma regra de escolha sobre a defasagem espacial a ser considerada na estimação econométrica. Em busca de uma abordagem empírica objetiva em que os dados possam “falar”, considera-se a otimização da função de log-

verossimilhança utilizando diferentes funções *kernel* numa abordagem de Regressão Ponderada Geograficamente, como discutida no terceiro capítulo desta tese.

A equação 3.1 é estimada para cada um dos 13 anos dos microdados da RAIS, a fim de obter a banda ótima. A função *kernel* tricúbica apresenta critérios de informação AIC relativamente menores e seus resultados para o tamanho da banda apontam para uma extensão média de 2 km, como demonstra a Tabela 4.7. Por esse critério, consideram-se na estimação do modelo econométrico defasagens que capturem uma extensão de 2 km partindo do centroide da célula em que o indivíduo está localizado. Esquemáticamente, as defasagens são expresas na Figura 4.1.

Tabela 4.7 – Função kernel, critério de informação AIC e tamanho das bandas

Ano	Gaussiana		Biquadrada		Tricúbica	
	AIC	Banda (Km)	AIC	Banda (Km)	AIC	Banda (Km)
2002	150453,9	0,910	150289,6	2,271	150197,6	2,309
2003	151015,5	0,901	150902,8	2,214	150763,0	1,991
2004	152529,7	0,886	152384,9	2,160	152233,8	1,979
2005	154075,7	0,905	153929,3	2,218	153810,6	2,095
2006	154806,5	0,903	154670,1	2,205	154538,4	2,063
2007	158084,3	0,911	157947,1	2,236	157831,5	2,105
2008	157403,6	0,879	157169,2	2,157	157033,3	2,095
2009	157891,6	0,869	157671,1	2,131	157522,1	2,023
2010	160279,4	0,910	160142,8	2,244	160025,8	2,099
2011	159271,1	0,859	158985,5	2,089	158823,2	2,038
2012	159199,8	0,857	158913,1	2,079	158750,4	2,038
2013	159784	0,879	159549,5	2,154	159408,2	2,094
2014	159668,7	0,871	159429,8	2,129	159279,6	2,063
Média		0,888		2,176		2,076

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014)

Os efeitos da densidade interna (efeito direto) e da aglomeração de primeira e segunda ordem (efeitos indiretos) sobre os salários estão apresentados na Tabela 4.8. Além da apresentação dos resultados principais, discute-se também a relevância dos controles de *sorting* e *matches* sobre os resultados. Essa discussão também é relevante devido ao debate existente na literatura quanto à penalidade que os prêmios salariais sofrem em face de estimações de modelos mais restritos (mais controles para efeitos fixos). Todos os resultados derivam da estimação da equação 4.13 e devem ser lidos como semi-elasticidades. Ademais, os coeficientes foram multiplicados por 100.000 e, portanto, os efeitos discutidos adiante devem sempre ser lidos como uma função de um choque de 100.000 trabalhadores adicionais.

Os resultados estimados pelo estimador POLS indicam a existência de prêmio salarial intra-urbano generalizado (direto e indireto) na RMSP, ver Tabela 4.8. O prêmio salarial para a

aglomeração medida no local de trabalho do indivíduo é da ordem de 0,068%. Por sua vez, o efeito indireto capturado pela aglomeração dos vizinhos de primeira ordem apresenta efeito positivo e quatro vezes mais elevado quando comparado ao efeito direto. Já o efeito indireto de segunda ordem é positivo, mas menor do que os efeitos anteriores, fornecendo, *a priori*, evidências contra a hipótese de atenuação do efeito da aglomeração.

Tabela 4.8 – Efeitos Estimados da Aglomeração sobre a Produtividade do Trabalhador

Variáveis	POLS	I	II	III	IV
Emp	0,068***	0,066***	0,092***	0,034***	0,064***
Wemp	0,276***	-0,027***	-0,084***	0,103***	0,066***
W2Emp	0,021***	-0,147***	-0,212***	-0,461***	-0,457***
Trabalhador	Não	Sim	Não	Não	Não
Firma	Não	Não	Sim	Não	Não
Célula	Não	Não	Não	Sim	Não
Indústria-Célula	Não	Não	Não	Não	Sim
Match	Não	Não	Não	Não	Não
R2	0,77	0,95	0,86	0,77	0,81
Obs.	4.573.205	4.573.205	4.573.205	4.573.205	4.573.205
Estat. F	168.660	721.881	241.301	230.803	247.336
Prob>F	0,00	0,00	0,00	0,00	0,00
Variáveis	V	VI	VII	VIII	IX
Emp	0,041***	0,071***	0,079***	0,068***	0,039***
Wemp	-0,140***	-0,093***	-0,062***	-0,061***	-0,109***
W2Emp	-0,143***	-0,178***	-0,241***	-0,217***	-0,232***
Trabalhador	Não	Sim	Sim	Sim	Sim
Firma	Não	Sim	Sim	Sim	Sim
Célula	Não	Não	Sim	Sim	Sim
Indústria-Célula	Não	Não	Não	Sim	Sim
Match	Sim	Não	Não	Não	Sim
R2	0,97	0,96	0,96	0,96	0,97
Obs.	4.573.205	4.573.205	4.573.205	4.573.205	4.573.205
Estat. F	1.477.694	745.962	744.803	733.410	816.155
Prob>F	0,00	0,00	0,00	0,00	0,00

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: os dados concernentes a Emp, WEmp e W2Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Cada modelo inclui controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios, exceto o modelo POLS que adiciona controles para gênero, idade, idade², setor da firma. Todos os coeficientes foram multiplicados por 100.000. Portanto, todos os resultados da estimação econométrica, que utiliza a forma funcional log-linear, devem ser entendidos como efeitos decorrentes da ampliação da aglomeração na ordem de 100.000 trabalhadores. Os desvios-padrão estão reportados entre parênteses. Significante a 1% (***), 5% (**) e 10% (*).

Ao controlar apenas pelo efeito fixo do trabalhador, o resultado é reduzido em apenas 3%, de 0,068 (estimador POLS) para 0,066. O hiato observado entre os coeficientes estimados por POLS e Efeito Fixo para áreas espaciais mais agregadas é muito mais elevado quando comparado ao estimado por Yankow (2006), Combes, Duranton e Gobillon (2008) e Silva (2017). Os efeitos indiretos de primeira e de segunda ordem passam a ter efeito decrescente

sobre o salário do trabalhador, fornecendo suporte para a hipótese de atenuação espacial da aglomeração. Ao controlar apenas pelo efeito fixo da firma, o prêmio salarial é de 0,092%, superior ao coeficiente estimado por POLS e Efeito Fixo de indivíduos. Quanto aos efeitos indiretos, a relação decrescente mantém-se e a magnitude negativa do efeito é aprofundada.

Buscando capturar as dotações fixas presentes no espaço (célula), a coluna III da Tabela 4.8 apresenta prêmio salarial (efeito direto) de 0,034%, de menor magnitude entre todas as regressões estimadas. O efeito indireto de primeira ordem é positivo e maior que o efeito anterior. O prêmio salarial indireto de segunda ordem é negativo. Nesse caso, a atenuação do efeito da aglomeração não é observada. Controlando pelas possíveis especializações setoriais fixas no espaço, o prêmio espacial é de 0,064% (coluna IV, Tabela 4.8) e os prêmios salariais indiretos se comportam como na regressão que controla o efeito fixo da célula apenas.

Ao controlar pelo efeito fixo do *match* trabalhador-firma-espço (coluna V), o prêmio direto é reduzido em relação ao POLS, apresentando um efeito positivo sobre o salário-hora do trabalhador em 0,041%. Os coeficientes que capturam o prêmio salarial indireto fornecem evidências de que o efeito da aglomeração é atenuado no espaço e afeta negativamente os salários do trabalhador. Portanto, uma realocação de trabalhadores para áreas vizinhas implicaria em redução do prêmio salarial intra-urbano.

Controlando o *sorting* dos trabalhadores e firmas simultaneamente na regressão (coluna VI, Tabela 4.8), os resultados apontam para a existência de prêmio salarial positivo, sendo negativos os prêmios indiretos. Adicionando o efeito fixo de célula (coluna VII), o prêmio salarial direto é ampliado, o efeito negativo da aglomeração de primeira ordem sobre o salário é atenuado e o efeito de segunda ordem é acentuado, relativamente ao modelo que controla apenas pelos efeitos fixos do trabalhador e firma.

Ao adicionar controle para o efeito fixo de especialização setorial no espaço (coluna VIII), o efeito da aglomeração no local onde o indivíduo trabalha apresenta um retorno ao salário-hora em 0,068%, valor 14% menor quando comparado ao modelo sem controle para especificação setorial fixa no espaço (coluna VII) e os efeitos negativos são levemente atenuados. Conjuntamente, esses resultados apontam para a importância da especialização setorial fixa no tempo, sobretudo em uma economia com *clusters* espaciais de atividades

como na RMSP, como atividades financeiras na região da Av. Paulista e da Av. Brigadeiro Faria Lima, por exemplo, entre outras.

A especificação econométrica proposta por Combes, Duranton e Gobillon (2008), associada à discussão sobre efeitos fixos feita por Abowd, Kramarz e Margolis (1999) e Woodcock (2008; 2015) apresenta resultados mais conservadores (coluna IX). Os resultados mais moderados derivam do fato de que uma parcela do efeito da aglomeração é também componente dos efeitos fixos. Nesse caso, o prêmio salarial intra-urbano estimado é o prêmio observado e líquido dos efeitos fixos associados, inclusive de uma parcela da aglomeração que é fixa no tempo e não observável.

Ao comparar o prêmio salarial direto estimado pelo modelo *baseline* (coluna IX) ao respectivo coeficiente estimado pelo estimador POLS, o efeito é reduzido em 74%. Essa redução é mais acentuada quando comparada à redução que ocorre na comparação do POLS com o controle de efeito fixo de indivíduos. Quando se compara o efeito direto estimado pela especificação de referência com a especificação apresentada na coluna II, o efeito é reduzido em 69% e em relação ao respectivo coeficiente estimado com controle para efeito fixo de firma, em 135%. Controlando ambos efeitos simultaneamente, a redução é de 102%. Essas instabilidades dos coeficientes estimados podem ser notadas também nos coeficientes que capturam o efeito indireto da aglomeração. Os resultados para esses coeficientes estimados a partir do modelo *baseline* são relativamente maiores, em módulo, quando comparados aos respectivos coeficientes da coluna I, II e VI e VIII.

Os resultados apontam para um padrão persistente do efeito da aglomeração na própria célula sobre o salário do trabalhador. Todavia, as aglomerações vizinhas, que capturam os efeitos indiretos da aglomeração, apresentam efeitos negativos e decrescentes com a ampliação da distância, exceto para os resultados das especificações das colunas III e IV (Tabela 4.8).

Em geral, os resultados apontam o efeito positivo da aglomeração sobre o salário, quando o posto de emprego do indivíduo se localiza na área de aglomeração e também mostram que o efeito é declinante em função do distanciamento. Esses resultados vão ao encontro da hipótese de atenuação espacial da aglomeração, abrindo espaço à avaliação adicional quanto ao efeito sobre os salários quando variações no tamanho das células são observadas. Dito de

outra forma, a redução do tamanho da área das células levaria a contatos mais próximos entre trabalhadores, podendo afetar os salários numa magnitude maior.

Nesse contexto, os resultados sugerem que fortalecimento de uma paisagem intra-urbana com aglomeração local dos empregos (no local em que o indivíduo trabalha) é mais eficiente, em termos de prêmio salarial, quando comparada às aglomerações vizinhas (fora do local em que o indivíduo trabalha). Ao considerar a possibilidade de alocação ou realocação espacial do trabalhador, os resultados sugerem que aglomerar é mais eficiente relativamente à estratégia de alocação em áreas menos densas. Assim, assumindo que efeitos positivos da aglomeração sejam ativados pelo mecanismo do contato face a face, os resultados sugerem que esse mecanismo é fragilizado à medida em que o contato do trabalhador com as aglomerações de primeira e segunda ordem apresentam maiores atritos derivados da distância.

Embora os resultados não sejam capazes de afirmar o efeito causal das relações discutidas, eles sugerem que os efeitos da aglomeração são atenuados espacialmente. Como discutido acima, os resultados estimados são enviesados devido à endogeneidade. Em face dessa questão, passa-se a apresentar os resultados estimados a partir da equação 4.17 por 2SLS-IV.

Utilizando as defasagens temporais da área construída por km^2 , assume-se que o choque exógeno defasado da ampliação da área construída afeta diretamente a concentração de trabalhadores. Nesse sentido, as construções de novas áreas são *proxies* para estimar o estoque da aglomeração de trabalhadores por célula e limita, pelo lado da oferta de áreas construídas, a ampliação do adensamento de trabalhadores no espaço.

Duas condições são necessárias para a consistência dos coeficientes utilizando o estimador de 2SLS: a) correlação do instrumento com a variável endógena e b) exogeneidade condicional do instrumento, isto é, requer-se que o resíduo da regressão do primeiro estágio seja não correlacionado com a área construída. A primeira condição é facilmente demonstrada tanto pelas correlações (Tabela 4.2) quanto pelas estimações do primeiro estágio. A segunda condição envolve que a aglomeração de trabalhadores seja explicada pelas áreas construídas, mas a parte não explicada dessa regressão (os resíduos) não esteja correlacionada com os instrumentos.

A Tabela G.1 fornece os resultados do primeiro estágio, utilizando as áreas construídas para fins comerciais e residenciais separadamente na regressão e suas respectivas defasagens. A utilização simultânea das áreas construídas para ambos fins deriva da complementariedade existente entre imóveis para fins comerciais e residenciais, isto é, imóveis construídos para moradia que passam a sediar empresas. Os coeficientes associados às variáveis instrumentais são todos estatisticamente significantes, fornecendo evidências adicionais sobre a correlação entre cada variável endógena e os instrumentos utilizados. Os resultados e a direção dos instrumentos não apresentam padrões específicos que permitam uma discussão mais aprofundada. Stock e Yogo (2003) demonstram que instrumentos fracos enviesam os coeficientes estimados. Segundo os autores, valores da estatística F maiores do que 10 são suficientes e sugerem que o viés de instrumento fraco é pequeno. Na Tabela 4.9, todos os testes F de primeiro estágio, para as diferentes especificações, sugerem que esse viés é pequeno em relação ao estimador POLS e que os instrumentos não são fracos.

Sobre a segunda hipótese de exogeneidade condicional do instrumento, a Tabela G.2 apresenta evidências que dão suporte, em termos estatísticos, aos instrumentos utilizados. Em outras palavras, estimaram-se os resíduos do primeiro estágio contra as variáveis instrumentais e controles e os testes F para cada modelo estimado não rejeitam a hipótese de que os coeficientes sejam nulos, ou seja, F -test iguais a 1. E a estatística t (t -test) associada a cada instrumento também não rejeita a hipótese nula. Esses testes estatísticos sugerem a exogeneidade condicional dos instrumentos em cada regressão. Como se nota, os resultados são robustos para qualquer uma das especificações estimadas, isto é, para todas as variações de efeito fixo. Portanto, além da exogeneidade inerente do fato de que áreas construídas defasadas no tempo não afetam o salário contemporâneo dos trabalhadores, os testes estatísticos fornecem suporte adicionais à sua validade.

A Tabela 4.9 apresenta os resultados da regressão de segundo estágio, estimada como descrito na equação 4.19. Cada coluna apresenta diferentes regressões, controlando por diferentes conjuntos de efeitos fixos. Os coeficientes do efeito direto (Emp) e os efeitos indiretos ($WEmp$ e $W2Emp$) reportam os efeitos marginais de incrementos da ordem de 100.000 trabalhadores na aglomeração direta e indireta (primera e segunda ordem) sobre o logaritmo do salário do trabalhador, como discutido anteriormente.

Tabela 4.9 – Efeitos Estimados da Aglomeração sobre a Produtividade do Trabalhador (2SLS)

	POLS	I	II	III	IV
Emp	1,891***	0,137***	0,191***	-0,276***	0,492***
WEmp	0,919***	0,533***	-0,157*	1,045***	0,181***
W2Emp	-1,634***	-0,654***	-0,039	-1,007***	-1,181***
Trabalhador	Não	Sim	Não	Não	Não
Firma	Não	Não	Sim	Não	Não
Célula	Não	Não	Não	Sim	Não
Indústria-Célula	Não	Não	Não	Não	Sim
Match	Não	Não	Não	Não	Não
F-test (1º Estágio)	127767	104302	95944	119031	118815
Prob>F	0,00	0,00	0,00	0,00	0,00
F-test (1º Estágio)	426364	285396	226085	307135	296181
Prob>F	0,00	0,00	0,00	0,00	0,00
F-test (1º Estágio)	568457	285396	317626	402471	388493
Prob>F	0,00	0,00	0,00	0,00	0,00
R ²	0,75	0,95	0,86	0,77	0,80
Obs	4573205	4573205	4573205	4573205	4573205.00
F	155760	719950	241269	230709	247183.75
Prob>F	0,00	0,00	0,00	0,00	0,00
	V	VI	VII	VIII	IX
Emp	1,661***	0,430***	1,294***	1,460***	1,785***
WEmp	-2,109***	-0,539***	-1,549***	-1,800***	-2,121***
W2Emp	-0,240	-0,034	-0,210	-0,126	-0,382
Trabalhador	Não	Não	Sim	Sim	Sim
Firma	Não	Não	Sim	Sim	Sim
Célula	Não	Não	Não	Sim	Sim
Indústria-Célula	Sim	Não	Não	Sim	Sim
Match	Não	Sim	Não	Não	Sim
F-test (1º Estágio)	76499	85410	92240	88966	73242
Prob>F	0,00	0,00	0,00	0,00	0,00
F-test (1º Estágio)	218043	205396	249352	242710	209548
Prob>F	0,00	0,00	0,00	0,00	0,00
F-test (1º Estágio)	305179	291105	340120	337254	297056
Prob>F	0,00	0,00	0,00	0,00	0,00
R ²	0,97	0,96	0,96	0,96	0,97
Obs	4573205	4573205	4573205	4573205	4573205
F-Stat	1425439	743846	718856	704977	767570
Prob>F	0,00	0,00	0,00	0,00	0,00

Nota: os dados concernentes a Emp, WEmp e W2Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Cada modelo inclui controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios, exceto o modelo POLS que adiciona controles para gênero, idade, idade², setor da firma. Todos os coeficientes foram multiplicados por 100.000. Portanto, todos os resultados da estimação econométrica, que utiliza a forma funcional log-linear, devem ser entendidos como efeitos decorrentes da ampliação da aglomeração na ordem de 100.000 trabalhadores e lidos como semi-elasticidade. Significante a 1% (***), 5% (**) e 10% (*).

A partir das estimações apresentadas na Tabela 4.9 percebe-se o viés de ampliação dos coeficientes estimados (em módulo) por 2SLS-IV, sugerindo que a variável omitida é correlacionada positivamente com a densidade de emprego e os valores estimados sem correção à endogeneidade tendem a subestimar o prêmio salarial intra-urbano e a atenuação espacial da aglomeração. Quanto ao efeito direto, a sobre-estimação em relação ao modelo

de referência com endogeneidade pode derivar da omissão de habilidades do trabalhador e/ou tecnologia da firma que justificaram o *sorting* espacial que, como discutido acima, impacta o salário-hora médio. O mesmo motivo poderia estar aprofundando o efeito indireto negativo. As habilidades ou tecnologias não observadas variantes no tempo determinam o *sorting* para a área em que a firma está localizada e o indivíduo trabalha, e não para as áreas vizinhas e, portanto, teria impacto negativo sobre o salário. Essa possível explicação pode, em alguma medida, justificar a insignificância estatística do efeito indireto de segunda ordem. Ou seja, o que o instrumento está trazendo para o modelo é uma parcela da habilidade do trabalhador e/ou tecnologia da firma, positiva e escondida no termo de erro, que tem função apenas no espaço em que seu emprego está localizado e até a primeira defasagem.

Uma possível interpretação desses resultados é que a aglomeração de trabalhadores captura um efeito de interação, tal como o transbordamento de conhecimento derivado da interação dos trabalhadores e este é limitado espacialmente. Tal capacidade de interação poderia ser a habilidade do trabalhador não observada. Assim, o componente de atenuação aponta para o atrito gerado pela ampliação do distanciamento do trabalhador com as áreas de aglomeração vizinhas, dificultado o transbordamento e absorção desse conhecimento pelo trabalhador.

A utilização em separado de áreas construídas para fins específicos está embasado no fato de que construções para fins comerciais e residenciais ativam canais diferentes da aglomeração e que áreas residenciais podem ser utilizadas para fins comerciais, como já dito acima. Como sensibilidade, consideram-se como instrumentos as áreas construídas para fins habitacionais, para fins comerciais e a soma dessas áreas, cada variável instrumental é utilizada em regressões distintas. Com esse exercício objetiva-se justificar que ao considerar apenas áreas construídas para fins comerciais deixa-se de fora uma parcela do prêmio salarial intra-urbano, considerar apenas áreas construídas para fins habitacionais para instrumentalizar aglomeração não é factível e nem mesmo agregar numa única variável as áreas construídas para fins diferenciados. Nesse sentido, é possível depreender que áreas construídas para fins residências é complementar às áreas para fins comerciais e não se apresentam como bons instrumentos para a aglomeração de trabalhador quando utilizada única e exclusivamente como variável instrumental.

A Tabela G.3 apresenta o primeiro estágio e as estimações dos resíduos dos primeiros estágios contra os instrumentos. Em ambos exercícios, os testes estatísticos fornecem suporte para utilização como instrumento. A Tabela 4.10 apresenta os resultados estimados apenas para a especificação *baseline*. Na coluna I, onde as áreas construídas para fins comerciais são utilizadas como instrumento, os resultados são próximos aos estimados no modelo *baseline* apresentado na coluna IX da Tabela 4.9. O prêmio salarial intra-urbano é de 1,59% enquanto o resultado estimado no modelo *baseline* é de 1,78%. A atenuação espacial também é observada. Aglomerações fora da área que aninha o posto de trabalho do indivíduo impactam negativamente o prêmio salarial urbano em 2,04% (no modelo *baseline* da Tabela 4.9, esse resultado é de -2,12%). O efeito indireto de segunda ordem não é estatisticamente significativo e apresenta sinal negativo, em consonância com o modelo principal.

Tabela 4. 10 – Efeitos estimados da aglomeração sobre a produtividade do trabalhador com diferentes variáveis instrumentais (2SLS)

	I	II	III
Emp	1,59**	-4,91	-2,63
Wemp	-2,04***	-23,05*	4,37
W ² Emp	-0,573	34,91	-0,46
Trabalhador	Sim	Sim	Sim
Firma	Sim	Sim	Sim
Célula	Sim	Sim	Sim
Indústria-Célula	Sim	Sim	Sim
Match	Sim	Sim	Sim
F-test (1º Estágio)	139992	5143	24936
Prob>F	0,000	0,000	0,000
F-test (1º Estágio)	397371	7311	61265
Prob>F	0,000	0,000	0,000
F-test (1º Estágio)	586913	5026	
Prob>F	0,000	0,000	0,000
R ²	0,922	0,971	0,969
Obs.	4573205	4573205	4573205
Estat. F	2.8e+05	7.7e+05	7.2e+05
Prob>F	0,000	0,000	0,000

Nota: os dados concernentes a Emp, WEmp e W²Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Na coluna I, utilizam-se as áreas construídas para fins comerciais como instrumento, nas colunas II e III são utilizadas como instrumentos as áreas construídas para fins habitacionais e a soma das áreas para fins comerciais e residenciais, respectivamente. Cada modelo inclui controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios, exceto o modelo POLS que adiciona controles para gênero, idade, idade², setor da firma. Todos os coeficientes foram multiplicados por 100.000. Portanto, todos os resultados da estimação econométrica, que utiliza a forma funcional log-linear, devem ser entendidos como efeitos decorrentes da ampliação da aglomeração na ordem de 100.000 trabalhadores e lidos como semi-elasticidade. Significante a 1% (***), 5% (**) e 10% (*).

As colunas II e III da Tabela 4.10 consideram as áreas residenciais e a soma das áreas residenciais e comerciais como instrumentos para aglomeração, respectivamente. Observa-

se que ambos instrumentos não apresetam significância estatística e a hipótese de atenuação espacial da economia de aglomeração não pode ser observada. Apenas o efeito indireto de primeira ordem, coluna II, é estatisticamente significativo a 10%.

Esse exercício de sensibilidade dos instrumentos sugere que considerar apenas a área construída para fins comerciais é o mais relevante para identificar o prêmio salarial intra-urbano, enquanto considerar áreas construídas para fins residenciais apenas se justifica considerando a sua complementariedade. Tal afirmação se sustenta ao comparar os prêmios salariais estimados a partir do modelo de referência e do apresentado na coluna I da Tabela 4.10. A diferença entre ambos coeficientes é de 0,2 pontos percentuais. Portanto, os instrumentos propostos e a utilização simultânea, mas em separado no mesmo modelo econométrico, aparenta ser mais efetivo e mais factível quando comparado às especificações expostas nas colunas II e III da Tabela 4.10. Ademais, considerar apenas áreas construídas para fins comerciais, assumindo que todas alocações dos trabalhadores se dariam exclusivamente nessas áreas, pouco afeta o resultado estimado por 2SLS-IV, mas omitiria uma parcela do prêmio salarial intra-urbano.

Em síntese, os resultados estimados por 2SLS-IV fornecem prêmios salariais intra-urbanos mais elevados quando comparados aos modelos estimados sem considerar a endogeneidade. Nesse sentido, os resultados derivados das estimações por 2SLS-IV sugerem que possíveis habilidades do trabalhador e/ou tecnologias da firma, temporalmente variantes no tempo, quando omitidas da estimação tenderiam a subestimar o prêmio salarial intra-urbano. Embora os sinais dos coeficientes associados aos efeitos diretos e indiretos sejam consistentes entre os dois estimadores utilizados, as significâncias estatísticas não são. Os modelos que corrigem a endogeneidade restringem a extensão do transbordamento do prêmio até as células que compõem a vizinhança de primeira ordem, enquanto nas estimações com variáveis endógenas o modelo considera o efeito até a vizinhança de segunda ordem.

Em geral, os modelos endógenos são mais conservadores, estimando prêmios salariais intra-urbanos substancialmente menores *vis-à-vis* aos modelos que tratam a endogeneidade, como observado em Rosenthal e Sgrange (2008) e Silva (2017). O efeito direto estimado no segundo grupo de modelos (coluna IX, Tabela 4.9) é quarenta e seis vezes maior do que o mesmo efeito estimado no primeiro grupo de modelos (coluna IX, Tabela 4.8). Já o efeito

indireto de primeira ordem (derivado do segundo grupo de modelos) é vinte vezes maior do que o mesmo efeito estimado pelo primeiro grupo de modelos. Portanto, os modelos com variáveis instrumentais ampliam os coeficientes estimados, o que, em alguma medida, deriva de omissões de características relevantes, seja dos trabalhadores e/ou das firmas. Todavia, o modelo endógeno fornece a estimação *lower bound* dos efeitos direto e indiretos.

4.3.3. Sensibilidade dos Resultados à Extensão das Áreas

Os resultados da subseção anterior apontam à atenuação espacial do efeito da aglomeração, sugerindo que a relação mais importante é aquela que ocorre onde o indivíduo tem seu posto de trabalho localizado. Uma dentre as varias possibilidades para esse resultado é que as economias de aglomeração estariam sendo ativadas por relações face a face entre os trabalhadores. Nessa subseção, buscam-se evidências adicionais de que tais relações são os canais pelos quais são ativadas externalidades positivas da concentração de trabalhadores que conectam economias de aglomeração e salário.

Aglomeração permanece sendo a soma do total de trabalhadores por célula. A estratégia empírica é reduzida à reestimação do modelo *baseline* considerando grids com células maiores e menores a 1 km² de área⁴⁵, observando o impacto das alterações da área de aglomeração sobre salários. Sabe-se que ao reduzir (ampliar) a área em que os indivíduos se aglomeram pode tanto ampliar (reduzir) o possível contato entre os trabalhadores como reduzir (ampliar) o total de trabalhadores aglomerados, resultando em possível *confounding* dos resultados estimados.

Inicialmente, visando capturar o efeito das alterações propostas sobre a média de trabalhadores aglomerados por célula, apresentam-se as estatísticas descritivas na Tabela 4.11 com as variações nas áreas das células dos grids. Normalizando o total de emprego por 0,5 km² para garantir a comparabilidade entre os grids com diferentes tamanhos de células, notam-se médias de emprego, na coluna *Média por 0,5 km²*, considerando grid com células com 0,5 km², mais elevadas relativamente às médias em áreas de 2 km². Esses resultados sugerem que o total de trabalhadores por área não segue uma função linear, isto é, reduzir

⁴⁵ É importante esclarecer que tanto para as células de 0,5 km² quanto para as células de 2 km² todo o grid foi refeito. Não foi particionada cada célula de 1 km² por duas de 0,5 km² e nem agregadas duas células de 1 km² para compor células com áreas de 2 km² cada.

(ampliar) a área pela metade não reduz (eleva) o total de trabalhadores na célula na mesma proporção. Esse resultado fornece base para o pressuposto de que a probabilidade de contatos por área tende a se elevar a medida que o espaço físico diminui. Em outras palavras, no caso da amostra da RMSP, a densidade de trabalhador por 0,5 km² é maior em áreas menores.

A redução da área de agregação dos trabalhadores possivelmente aumenta a probabilidade de contato entre os trabalhadores, todavia o total de trabalhadores aninhado por células se eleva quanto maior é o tamanho da célula. Na Tabela 4.11, observam-se valores médios de trabalhadores por células mais elevados quanto maior é a célula, ver coluna *Média* (valores não normalizados pela área). Essa questão também pode dificultar a identificação do real mecanismo que ativa os efeitos positivos da aglomeração. Resumidamente, assumem-se a existência de dois possíveis fatores: fator de interação e fator de nível de aglomeração. O primeiro diz respeito ao possível contato entre os trabalhadores e o segundo, ao total de trabalhadores dentro da célula.

Tabela 4. 11 – Estatística descritiva para diferentes tamanhos de células

	Média ¹	Desv. Pad. ¹	Min ¹	Max ¹	Média por 0,5 km ²
Emp (0,5km ²)	6.407	8.014	1	58.536	6.407
WEmp (0,5km ²)	4.015	4.574	0	24.154	4.015
W2Emp (0,5km ²)	3.298	3.400	0	20.588	3.298
W3Emp (0,5km ²)	2.977	2.867	0	13.546	2.977
Emp (1 km ²)	10.671	14.104	1	57.773	5.336
Wemp (1km ²)	7.235	7.815	0	22.321	3.618
W2emp (1km ²)	6.112	5.954	0	13.411	3.056
Emp (2km ²)	18.329	21.830	0	31.417	4.582
WEmp (2km ²)	12.859	12.920	0	15.478	3.215

Fonte: elaboração própria a partir dos dados da RAIS. Nota: ¹ consideram-se as suas respectivas áreas das células.

Caso o fator de interação (possível contato entre trabalhadores) sobreponha o fator de nível de aglomeração, no primeiro cenário, onde é estimada a especificação *baseline* e consideradas células com 0,5 km², espera-se uma ampliação do prêmio salarial enquanto no segundo cenário, modelo com especificação *baseline* e células com 2 km², uma redução do prêmio salarial é esperada, relativamente ao modelo *baseline* com células de 1 km². Se o fator de nível de aglomeração é mais relevante, os impactos sobre o salário dos trabalhadores devem ser maiores em áreas com maior número de trabalhadores aglomerados. Para ambas

variações do tamanho de células, considera-se apenas a especificação completa de efeitos fixos estimados por Efeitos Fixos e 2SLS-IV.

Este experimento empírico também abre espaço para testar a sensibilidade da atenuação espacial das economias de aglomeração em face de alterações das áreas de aglomeração. Todavia, a fim de manter constante o tamanho da janela com o modelo *baseline*, a vizinhança no primeiro cenário é estendida até a terceira ordem e no segundo cenário, até a primeira ordem apenas. Tal estratégia mantém constante a área total considerada em ambos os cenários e, inclusive, é comprável com o modelo *baseline* com células de 1 km². Estimações do primeiro estágio e dos respectivos resíduos contra os instrumentos são apresentados nas Tabela G.4 e Tabela G.5, os resultados fornecem suporte aos instrumentos utilizados. Na Tabela 4.12 são apresentados os resultados do segundo estágio.

Tabela 4.12 – Efeitos estimados da aglomeração sobre a produtividade do trabalhador com diferentes tamanhos de áreas das células

	I	II	III	IV
	0,5 Km ²		2 Km ²	
Emp	0,0795***	0, 2134**	0,0437***	-0,5428***
WEmp	-0.0728***	-2,391***	-0,1892***	2,430***
W2Emp	-0. 2547***	2,53***	-	-
W3Emp	-0, 2527***	0,931	-	-
F-test (1º Estágio)	-	13979	-	43257
Prob>F	-	0,00	-	0,00
F-test (1º Estágio)	-	22658	-	118922
Prob>F	-	0,00	-	0,00
F-test (1º Estágio)	-	22317	-	150313
Prob>F	-	0,00	-	0,00
R ²	0,972	0,972	0,972	0,972
Obs.	4.573.205	4.573.205	4.573.205	4.573.205
F	786.694	779.560	811.741	805.846
Prob>F	0,00	0,00	0,00	0,00

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: os dados concernentes a *Emp*, *WEmp* e *W2Emp* são agregações de trabalhadores por células quadradas com 1 km² de área. Os modelos apresentados nas colunas I e II são estimados segundo o modelo *baseline* sem instrumentos. As estimacões apresentadas nas colunas II e VI utilizam o estimador 2SLS, tendo como instrumentos as áreas construídas para fins habitacionais e comerciais, como no modelo *baseline* com IV. Cada modelo inclui controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios, exceto o modelo POLS que adiciona controles para gênero, idade, idade², setor da firma. Todos os coeficientes foram multiplicados por 100.000. Portanto, todos os resultados da estimação econométrica, que utiliza a forma funcional log-linear, devem ser entendidos como efeitos decorrentes da ampliação da aglomeração na ordem de 100.000 trabalhadores e lidos como semi-elasticidade. Significante a 1% (***), 5% (**) e 10% (*).

Nos modelos com instrumentos, os efeitos seguem na direção das hipóteses acima. Avaliando o efeito direto (células em que o trabalhador está localizado, com dimensão de 0,5 km²), o impacto da aglomeração sobre o salário é de 0,21%. Esse coeficiente passa a negativo,

-0,54%, ao considerar células de 2 km². Normalizando os coeficientes estimados por áreas de 1km², o impacto estimado sobre o salário-hora médio é de 0,42% e -0,27%, respectivamente. Considerando células com 1 km², o impacto da economia de aglomeração sobre salários é 1,78%.

Os resultados sugerem que o prêmio salarial intra-urbano é sensível ao tamanho da célula. À medida em que o tamanho das áreas que aninham o total de trabalho é reduzido, o prêmio salarial intra-urbano é também reduzido, relativamente ao resultado derivado do grid com células de 1 km². Embora a densidade de trabalhador (total de trabalhador por tamanho de área) seja mais elevada, o nível da aglomeração (total de trabalhador na célula) é menor quando comparado ao grid *baseline*. Nesse sentido, a hipótese levantada nesse trabalho é que a ampliação da densidade potencializaria contatados F2F, os quais ativariam efeitos positivos das economias de aglomeração. Os resultados encontrados vão na direção oposta, fornecendo argumentos de que o mais relevante seria o nível de aglomeração.

O resultado da estimação que utiliza a agregação de trabalhadores por 2 km² apresenta sinal negativo para o prêmio salarial, resultado que merece considerações. Ao estimar o impacto sobre os salários médios com áreas de células mais elevadas, controlando pelo conjunto de efeitos fixos, o possível contato entre os trabalhadores dessa célula é fragilizado. Em outras palavras, contatos F2F tendem a ser pulverizados, embora o nível médio da aglomeração (total de trabalhador dentro da célula) seja elevado, resultando na reversão do efeito positivo da aglomeração. Esse resultado sugere que a ampliação da área reduz a probabilidade de contatos entre os trabalhadores e o efeito do nível da aglomeração não é suficiente para manter o efeito positivo da economia de aglomeração.

Quanto à hipótese de atenuação espacial das economias de aglomeração, o modelo que considera áreas de 0,5 km² apresenta apenas o coeficiente associado à primeira defasagem espacial com sinal negativo. Esse sinal é revertido na defasagem espacial seguinte e na terceira defasagem é estatisticamente não significativa e menor, em termos de magnitude, em relação à anterior. O efeito de atenuação não é observado. Esse resultado pode estar associado ao total de área que corresponde aos vizinhos de primeira ordem, 16 km², enquanto no primeiro cenário os vizinhos de primeira ordem ocupam áreas de 4 km².

Focando principalmente nos efeitos direto, os testes levados a cabo não fornecem evidências suficientes para dizer que as economias de aglomeração são ativadas única e exclusivamente por contato entre trabalhadores no espaço intra-urbano. Os experimentos permitem dizer que contatos F2F agem conjuntamente com o efeito de nível de aglomeração e a possível fragilização das relações de contato, derivadas da ampliação das células, pode resultar em efeitos perniciosos sobre os salários.

Caso seja considerado o modelo endógeno, que é mais conservador *vis-à-vis* os modelos estimados por 2SLS-IV, ao comparar os resultados expostos nas colunas XI (da Tabela 4.8), I e III (da Tabela 4.12) e normalizado por 1 km², nota-se que o efeito de primeira ordem sobre o salário-hora é reduzido à medida que área das células dos grids aumentam. No modelo que considera o grid com célula de 0,5 km², o efeito é de 0,158%. Dobrando a área do grid anterior, o efeito direto é de 0,039% e, considerando células com 2 km², o efeito é de 0,022%. As estimações apontam que o declínio dos efeitos diretos não é linear, fortalecendo a hipótese da coexistência dos fatores derivados da potencial interação F2F dos trabalhadores e do nível de aglomeração.

4.3.4. Economias de Aglomeração e Externalidades do Capital Humano

Evidências apontam que as cidades são centros de inovação, produção e comercialização de ideias (JEFFE; TRAJTENBERG; HENDERSON, 1993). Todavia, a apropriação dessas ideias depende da capacidade de absorção dos indivíduos. No contexto do mercado de trabalho, a absorção é função do nível educacional do trabalhador (COHEN; LEVINTHAL, 1990; FALCÃO; SILVEIRA NETO, 2007).

Para Cohen e Levinthal (1990), firmas que conduzem atividades de pesquisa e desenvolvimento se apropriam de maior parcela das pesquisas conduzidas por outras empresas. O canal pelo qual esse mecanismo é ativado deriva do fato de que trabalhadores altamente qualificados apresentam maior habilidade de comunicação e, por consequência, maior habilidade de aprenderem e de se apropriarem do conhecimento que está pelo ar na forma tácita e envolto em “barulho”, como expresso por Storper e Venables (2004), resultando em aumento da produtividade do trabalho. Por outro lado, trabalhadores com nível educacional limitado apresentam maior potencial de se beneficiar do transbordamento de

conhecimento associado à aglomeração, todavia a capacidade de absorção é limitada pelo próprio nível educacional restrito, quando comparados aos trabalhadores mais educados.

Gordon e McCann (2000) apontam que encadeamentos locais fortes entre firmas é uma parte que explica a aglomeração. Mesmo que esses encadeamentos, tanto para trás quanto para frente, sejam localmente fortes, estes raramente derivam de elevados custos de transporte, mas sim do fato da informação associada à transação física ser custosa ou difícil de transmitir par alongas distâncias (STORPER; VENABLES, 2004). Nesse sentido, o benefício gerado pela aglomeração, que estimula o fluxo de conhecimento e informação, é relevante para trabalhadores e firmas cujo fluxo de conhecimento e informação importem (GLAESER, 1999; BACOLOD; BLUM; STRANGE, 2009), justificando a aglomeração.

Moretti (2004b) aponta para a existência de ao menos três potenciais fontes de retorno social à educação: a) pessoas educadas cometem menos crimes; b) pessoas educadas tomam decisões considerando mais informações na hora de votar; e, c) proximidade de pessoas mais educadas podem melhorar a produtividade. Moretti (2004a) encontra que externalidades do capital humano⁴⁶ sobre o salário são positivas para os trabalhadores de todos os níveis educacionais, mas o efeito é maior para os trabalhadores mais educados. Bathelt, Malmberg e Maskell (2004) e Stoper e Venables (2004) argumentam que efeitos de transbordamento tendem a ser mais acentuados entre trabalhadores mais educados. Rosenthal e Sgrange (2008) e Andersson, Klaesson e Larsson (2012) encontram efeitos positivos da aglomeração mais elevados para trabalhadores mais educados relativamente àqueles com menor nível educacional nos Estados Unidos e na Suécia, respectivamente.

Considerando a relevância do capital humano sobre a apropriação de informações relevantes, que geralmente está disposta em aglomerações de pessoas, nesta sub-seção testa-se se a interação dos trabalhadores impacta o prêmio salarial intra-urbano de forma heterogênea em função do nível educacional. Para tanto, a base de dados foi particionada considerando os quatro ciclos educacionais dos trabalhadores: trabalhadores com 1) até ensino médio

⁴⁶ Existem certamente elementos além do nível educacional que compõe o capital humano (Bacolod, Blum e Strange, 2009), mas educação é um componente relevante do capital humano (Winter, 2013). Assim, a discussão sobre capital humano nessa subseção se limita ao nível educacional, e, apenas nesse contexto, podem, ambos termos, ser entendidos como substitutos perfeitos.

incompleto (Ciclo I); 2) ensino médio completo (Ciclo II); 3) ensino superior incompleto (Ciclo III); e, 4) ensino superior completo e mestrado e/ou doutorado (Ciclo IV). Como discutido acima, o modelo completo é menos parcimonioso e controla por um conjunto de auto seleção fixa, tanto do trabalhador quanto da firma. Por esse motivo, a discussão dos resultados a seguir está embasada e restrita ao modelo estimado *baseline*.

O estimador utilizado na regressão foi o 2SLS-IV e foram considerados os mesmos instrumentos do modelo *baseline*. A Tabela G.6 reporta os resultados do primeiro estágio e a Tabela G.7 traz os resultados estimados da média condicional dos resíduos estimados para cada uma das quatro regressões. Os resultados apontam para a correlação parcial dos instrumentos com as variáveis endógenas e exogeneidade dos instrumentos. Na Tabela 4.13 são reportadas as estatísticas F e *F-tests*, sendo todos maiores que 10 e estatisticamente significantes. Esses resultados apontam para a elevada correlação entre os instrumentos e variáveis endógenas.

Tabela 4.13 – Efeitos da aglomeração sobre a produtividade do trabalhador por ciclo educacional

Variáveis	Ciclo I	Ciclo II	Ciclo III	Ciclo IV
Emp	0,728***	2,611***	1,163***	3,634***
WEmp	-0,602***	-6,045***	0,110	0,318
W2Emp	-1,069***	2,657***	-1,330*	-6,281***
F-test (1º Estágio)	29638	20621	2568	13344
Prob>F	0,000	0,000	0,000	0,000
F-test (1º Estágio)	79174	60653	8188	45984
Prob>F	0,000	0,000	0,000	0,000
F-test (1º Estágio)	110877	90334	12798	58608
Prob>F	0,000	0,000	0,000	0,000
R ²	0,969	0,961	0,970	0,960
Obs.	1.813.489	1.432.227	257.504	1.069.985
F	415.661	204.224	37.248	124.222
Prob>F	0,000	0,000	0,000	0,000

Fonte: elaboração própria, a partir dos dados da RAIS. Nota: a variável dependente é $\ln(\text{salário-hora})$. Os dados concernentes a Emp, WEmp e W2Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios, exceto o modelo POLS que adiciona controles para gênero, idade, idade², setor da firma. Todos os coeficientes foram multiplicados por 100.000. Portanto, todos os resultados da estimação econométrica, que utiliza a forma funcional log-linear, devem ser entendidos como efeitos decorrentes da ampliação da aglomeração na ordem de 100.000 trabalhadores. Significante a 1% (***), 5% (**) e 10% (*).

A Tabela 4.13 reporta o impacto da aglomeração sobre o salário médio do trabalhador para um dado incremento de 100,000 trabalhadores. Avaliando o padrão geral, a aglomeração

medida pela densidade interna (efeito direto) impacta positivamente o salário-hora médio e, com exceção do Ciclo II, observa-se decaimento do efeito de aglomeração.

Focando a análise no efeito direto, os trabalhadores que possuem o nível educacional mais baixo (Ciclo I) absorvem relativamente menos do efeito positivo da aglomeração. Entre os trabalhadores que possuem ao menos graduação concluída (Ciclo IV), o efeito é mais expressivo, cerca de cinco vezes maior em comparação aos trabalhadores que compõem o Ciclo I. Esses resultados vão ao encontro da hipótese levantada por Cohen e Levinthal (1990), de que trabalhadores com menor nível educacional apresentam menor capacidade de absorção, e aos resultados encontrados em Falcão e Silveira Neto (2007), Rosenthal e Strange (2008), Andersson, Klaesson e Larsson (2012). Nesse caso, os resultados sugerem, considerando os ciclos extremos (Ciclo I e Ciclo IV), que os efeitos positivos gerados pelas economias de aglomeração beneficiam muito mais aos trabalhadores mais educados devido à sua capacidade de absorção.

O efeito de atenuação nos dois níveis mais baixos de educação é muito forte no tocante ao efeito de primeira ordem, maior do que nos dois ciclos mais altos. Já o efeito de segunda ordem entre os menos educados está presente apenas no caso do primeiro nível, mudando de sinal para o segundo nível. Para os dois ciclos superiores, a atenuação está presente, mas o efeito de primeira ordem, embora não significativo, ainda é positivo e o de segunda ordem é negativo e significativo, sendo menos intenso no terceiro ciclo e muito intenso nos níveis mais elevados de educação.

Esses resultados sugerem que a aglomeração é uma amenidade muito mais relevante para os trabalhadores com alto *skill* e, em alguma medida, os fatores que ativam as economias de aglomeração dentro de uma área de 1 km² é relativamente mais importante quando comparada àquelas que se dariam em vizinhanças circunvizinhas. Todavia, os trabalhadores de baixo *skill*, que observam uma perda do prêmio salarial intra-urbano logo na vizinhança de primeira ordem, e na de segunda ordem também, são relativamente mais sensíveis a qualquer realocação de empregados às células vizinhas (de primeira ordem) do que os trabalhadores com alto *skill*. Em síntese, os trabalhadores com menor nível educacional apresentam menor capacidade de absorver o conhecimento “solto pelo ar” e são mais

vulneráveis ao efeito de atenuação de primeira ordem. O mesmo não ocorre quando se considera o efeito de atenuação de segunda ordem.

4.3.5. Heterogeneidade das Economias de Aglomeração nos Subcentros

No capítulo anterior o esforço de pesquisa esteve concentrado em propor uma metodologia para identificar SDB e aplica-la à Região Metropolitana de São Paulo. Dos resultados anteriores, são considerados SBD as grandes aglomerações localizadas no centro expandido do município de São Paulo, nos municípios de Barueri e São Caetano do Sul. Nessa subseção, avaliam-se possíveis heterogeneidades no prêmio salarial intra-urbano direto e indireto em função das diferentes localizações espaciais do trabalhador, isto é, para cada uma das seguintes subdivisões SBD^{Q99} , SBD^{Q95} , SBD^{Q90} e $NSBD$. Em outras palavras, o objetivo é compreender se os trabalhadores dessas áreas se apropriam de forma diferente do prêmio salarial intra-urbano. Trabalhadores localizados em áreas menos densas, como no conjunto $NSBD$, teriam relações face a face mais restritas relativamente aos trabalhadores localizados em áreas mais densas, no SBD^{Q99} por exemplo. Nesse sentido, quão sensíveis são o prêmio salarial intra-urbano e a atenuação espacial nessas localidades?

A Tabela 4.14 apresenta os resultados da estimação da Equação 4.18 (2º estágio) para cada uma das quatro localidades. Novamente, todos os coeficientes estimados são normalizados por 100.000 e as regressões de 1º estágio e estimações das médias condicionais dos resíduos estão dispostas na Tabela G.8 e na Tabela G.9, respectivamente. O primeiro padrão geral observado é que o prêmio salarial direto é mais significativo para os trabalhadores que se localizam nas áreas periféricas da RM. Esse resultado sugere que o prêmio salarial associado à aglomeração é relativamente mais importante em regiões em que a aglomeração é menor. Nota-se que o prêmio salarial direto dos trabalhadores localizados no SBD^{Q99} é cerca de quarenta vezes menor que o prêmio direto auferido no $NSBD$. Isso pode ser explicado pela excassez relativa de aglomeração de trabalhadores nas áreas periféricas da RMSP, sugerindo que nessas áreas o retorno à aglomeração é mais significativo para o salário-hora dos trabalhadores nela localizados enquanto os trabalhadores localizados no SBD auferem prêmios salariais menores. Em resumo, quanto mais longe o trabalhador está do *core* de emprego da RMSP, mais importante é estar aglomerado.

Considerando-se os coeficientes de *WEmp* e *W2Emp*, nota-se que o benefício salarial de trabalhar em área relativamente mais densa na área NSBD, em comparação a áreas menos densas desse próprio espaço, é proporcionalmente maior do que a mesma comparação feita para os SBD identificados neste trabalho. Nestes últimos, trabalhar na vizinhança – em comparação a trabalhar na própria célula – produz redução salarial menor, proporcionalmente, do que nas áreas periféricas. Na região NSBD, estar no núcleo da aglomeração é muito mais importante, em termos de prêmio salarial, do que nas regiões mais densas.

Tabela 4.14 – Prêmio Salarial Intra-Urbano em Diferentes SBD (2º Estágio)

Variáveis	<i>SBD^{Q99}</i>	<i>SBD^{Q95}</i>	<i>SBD^{Q90}</i>	<i>NSBD</i>
Emp	0,312***	1,089***	1,771***	12,676***
WEmp	-1,527***	-1,977***	-2,131***	7,607*
W2Emp	0,777***	1,011***	-0,306*	-4,663**
F-test (1º Estágio)	3118	28368	49317	460
Prob>F	0,000	0,000	0,000	0,000
F-test (1º Estágio)	11078	83810	141819	3798
Prob>F	0,000	0,000	0,000	0,000
F-test (1º Estágio)	10840	120309	199533	16244
Prob>F	0,000	0,000	0,000	0,000
R ²	0,9747	0,9722	0,9713	0,9715
Obs.	1.530.924	3.239.272	4.016.538	556.667
F	508.839	759.976	847.176	111.103
Prob>F	0,000	0,000	0,000	0,000

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: a variável dependente é $\ln(\text{salário-hora})$. Os dados concernentes a *Emp*, *WEmp* e *W2Emp* são agregações de trabalhadores por células quadradas com 1 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios, exceto o modelo POLS que adiciona controles para gênero, idade, idade², setor da firma. Todos os coeficientes foram multiplicados por 100.000. Portanto, todos os resultados da estimação econométrica, que utiliza a forma funcional log-linear, devem ser entendidos como efeitos decorrentes da ampliação da aglomeração na ordem de 100.000 trabalhadores. Significante a 1% (***), 5% (**) e 10% (*).

Para todos quatro grupos avaliados, a atenuação está presente na vizinhança de primeira ordem, sendo menos acentuada na região *SBD^{Q99}*, com impacto sobre o salário dos trabalhadores de -1,53%. Para os *SBD^{Q95}* e *SBD^{Q90}*, as reduções são de 1,98% e 2,13%, respectivamente. Em termos percentuais, indivíduos localizados em postos de trabalho fora dos SBD identificados é que observam maior redução do salário devido à escassez de externalidades geradas pelas economias de aglomeração; isto é, queda 60% quando comparada ao efeito direto.

Ao analisar o efeito de segunda ordem, não se observa qualquer padrão entre os subcentros e áreas não centrais. No SBD^{Q99} e no SBD^{Q95} , os prêmios salariais intra-urbanos de segunda ordem são positivos. No SBD^{Q90} , o prêmio é negativo, mas mais elevado do que o efeito indireto de primeira ordem. Esse resultado se justifica e é esperado devido à configuração espacial do emprego na RMSP. A topologia da aglomeração apresenta picos acentuados de emprego seguidos por picos menos elevados. Nesse sentido, a média espacial de emprego de segunda ordem tende a considerar áreas que compõem os subcentros de ordem superior. Esse efeito passa a ser atenuado no SBD^{Q90} , pois considera áreas que correspondem ao *NSBD* e, por consequência, reduz a média espacial do emprego. As estimações econométricas relacionadas à categoria *NSBD*, por considerar áreas mais isoladas e com menor proximidade da grande mancha de emprego dos SBD, dão suporte à hipótese de atenuação. Esse comportamento característico das áreas de emprego na RMSP, inclusive, justifica a utilização de vizinhos até segunda ordem, pois suaviza o possível erro de medida que é inerente à identificação de aglomeração em ordens superiores de vizinhança.

4.4. CONCLUSÃO

Este capítulo apresentou uma avaliação empírica do efeito da economia de aglomeração sobre o logaritmo do salário médio dos trabalhadores. A abordagem econométrica unifica diferentes vertentes da literatura de Economia do Trabalho e Economia Regional e Urbana, discutindo a importância da auto-seleção do trabalhador, firma e disparidades intra-urbanas, tais como *clusters* espaciais setoriais. A novidade empírica está relacionada à utilização de uma base de dados longitudinal georreferenciada, que permite a estimação de uma especificação econométrica com controles para múltiplos efeitos fixos (trabalhador, firma, *macth* trabalhador-firma-espço, área, setor e setor-área). Por um lado, essa abordagem é por demais restritiva, pois parte do efeito da aglomeração está associada aos efeitos fixos e empiricamente é indissociável. Por outro, permite maior precisão quanto à estimação do prêmio salarial intra-urbano. Nesse sentido, é possível que as estimações estejam subestimando o efeito real da economia de aglomeração sobre o salário médio.

Outro importante avanço desta tese é a estimação do efeito de atenuação espacial da aglomeração. Dois pontos são relevantes. Primeiramente, a abordagem em painel e a

utilização de grid com células pequenas permitem que os SBD e CBD da cidade não sejam considerados previamente para o cálculo de gradiente de salários, utilizando uma nomenclatura de Economia Urbana. Ademais, a estratégia empírica permite que a possível não-linearidade do gradiente de salários possa ser observada via defasagens espaciais da aglomeração. Em segundo lugar, o modelo econométrico estimado dá suporte à hipótese de atenuação espacial da economia de aglomeração e sugere que a concentração de trabalhadores na Região Metropolitana de São Paulo tem impacto positivo sobre o salário, mas que transições da concentração de emprego para localidades mais distantes do trabalhador resulta em retornos negativos.

Adicionalmente, buscou-se demonstrar que a economia de aglomeração é ativada por relações de proximidades entre os trabalhadores. Os resultados das estimações apontam que aglomerações com trabalhadores mais dispersos no espaço resultam em efeitos menores da aglomeração sobre o salário dos trabalhadores. Nesse sentido, a hipótese que se levanta nessa tese, assim como em Jacobs (1969), é que as relações face a face ativam a economia de aglomeração e impactam positivamente sobre os salários. Todavia, como discutido na seção sobre o papel do capital humano na absorção dos efeitos positivos gerados pela aglomeração, os trabalhadores com maior nível de educação (*high-skill*) apropriam-se cinco vezes mais desse efeito positivo quando comparados aos trabalhadores com menor nível de educação (*low-skill*). Para os trabalhadores *high-skill*, a dispersão do efeito da aglomeração é menos intensa quando comparado aos trabalhadores *low-skill*. Esse padrão, tanto do efeito direto quanto dos efeitos indiretos dos trabalhadores *high-skill*, está relacionado à capacidade do trabalhador médio em fazer conexões que permitam, de alguma forma, elevar a produtividade do trabalho, e essa se estende espacialmente mais além em comparação aos trabalhadores *low-skill*.

Ao estimar o prêmio salarial intra-urbano por SBD (considerando suas respectivas faixas de identificação), observa-se que o maior impacto da aglomeração reside fora dos SBD, devido ao elevado efeito marginal da aglomeração em áreas pouco densas. Por outro lado, em áreas com densidade muito elevadas, como SBD^{Q99} , o impacto é relativamente menor. Devido à presença de subcentralidades, ou concentrações menores, o efeito da atenuação não é observado para nenhum dos SBD, mas é possível notar em regiões não consideradas SBD. A

estimação do modelo *baseline* sugere não-linearidade do gradiente de salários e a avaliação de heterogeneidade por SBD fornece evidências adicionais, colocando em discussão a estimação de gradientes de preços com único gradiente em regiões policêntricas.

5. CONSIDERAÇÕES FINAIS

Nessa tese duas questões importantes são trazidas à discussão. A primeira diz respeito à identificação de centralidades de emprego. O problema mostra-se relevante desde os trabalhos seminais desenvolvidos pelos economistas da escola alemã. A importância dessas áreas está no fato de aninharem os efeitos derivados das economias de aglomeração. Nesse contexto, o primeiro capítulo discute a literatura empírica para identificação de subcentralidades de emprego e propõe uma abordagem empírica que supere a limitação dos métodos anteriores embasados em conhecimentos subjetivos ou regras não generalizáveis.

A nova abordagem de identificação de subcentralidades proposta nessa tese descarta a necessidade do pesquisador conhecer a distribuição de emprego da cidade avaliada. Adicionalmente, propõe uma regra de corte que é intertemporalmente ajustada à base de dados utilizada, o que indiretamente significa dizer que o valor de corte é uma função da dinâmica do mercado de trabalho, seja de crise ou de ascensão. Em linhas gerais, essa nova abordagem estima um modelo não paramétrico, com regressões ponderadas geograficamente (GWR), considerando bandas que otimizam a função de verossimilhança. Os resultados das médias espaciais estimadas são utilizados para definir o valor de corte, considerando os quantis de ordem 99%, 95% e 90% do total de emprego espacial médio estimado.

Para a aplicação dessa nova abordagem foi utilizada a base de microdados da RAIS. Nesse campo deu-se um passo adiante em termo de empiria. O avanço está na produção de uma base de dados sobre o mercado de trabalho georreferenciada, criando mais uma dimensão aos dados: além da dimensão temporal, agora é possível lidar com a dimensão espacial. Nessa perspectiva, esta tese trabalha, tanto no terceiro capítulo quanto no quinto, com uma base de dados inédita, permitindo explorar estratégias nunca antes testadas com os microdados da RAIS.

Como resultado das estimações da identificação de subcentros, três áreas centrais são encontradas na RMSP, localizadas em Barueri (SBD-BAR), no centro expandido de São Paulo (SBD-SAO) e em São Caetano do Sul (SBD-SCS). Essas áreas são consistentes apenas nos anos de 2002 e 2008, pois em 2014 o SBD-SCS perde em termos de participação relativa no mercado de trabalho da RMSP, passando a existir apenas duas áreas centrais. No ano de 2014, o SBD-SAO tem sua área de concentração de emprego ampliada enquanto o SBD-

BAR perde área de concentração. Esses resultados apontam para o maior crescimento relativo do emprego no município de São Paulo frente aos demais. Essa identificação deriva de valores de corte iguais ao quantil 99 da distribuição de médias espaciais de emprego. O mesmo exercício foi aplicado considerando valores de corte iguais aos quantis 95 e 90. O padrão encontrado é o surgimento de SBD menores na proximidade dos SBD identificados com cortes mais restritivos.

A segunda questão abordada nessa tese diz respeito ao prêmio salarial derivado das economias de aglomeração. Primeiramente, o georreferenciamento dos dados permite tratar a questão da unidade espacial de aninhamento dos dados, lidando com a questão da normalização do total de emprego pela área. Ao utilizar grids com células que possuem áreas idênticas (salvo na periferia), a questão da normalização é tratada. Outra vantagem em obter o georreferenciamento das firmas é poder avaliar, em alguma medida, se o contato face a face (F2F) é o canal que ativa os efeitos positivos das economias de aglomeração.

O objetivo central do quarto capítulo dessa tese foi avaliar o prêmio salarial intra-urbano e qual a extensão do efeito da economia de aglomeração sobre o salário, a hipótese de atenuação espacial. A utilização de defasagens do total de trabalhadores por célula demonstrou ser um avanço relativamente à estratégia empírica geralmente utilizada para calcular *wage gradient* na literatura empírica de Economia Urbana, a saber, de usar a distância de cada região ao CBD da cidade. Nesse caso, a estimação demanda o conhecimento *a priori* do CBD e/ou SBD. No caso desta tese, o conhecimento da âncora de preços é dispensado, pois todas as áreas apresentam sua própria aglomeração de trabalhadores. Assim, com essa abordagem, aqui denominada de *wage attenuation*, o decaimento do prêmio salarial não é monotônico, mas segue uma não linearidade em função do distanciamento espacial.

Uma contribuição adicional à literatura de Economia de Aglomeração está em utilizar uma base de dados longitudinal e controlar pelo *sorting* do trabalhador e firma, bem como pelo *match* trabalhador-firma-célula e pelas dotações fixas do espaço e da relação espaço-setor. Outro ponto relevante é considerar a endogeneidade inerente às economias de aglomeração, não tratadas mesmo com controle para o *sorting* decorrente de características não observáveis do trabalhador e da firma. Nesta tese foi utilizada a oferta de áreas construídas defasadas no

tempo. Os resultados estimados apontam para a existência de um prêmio salarial intra-urbano, mesmo em face de uma especificação tão restritiva quanto a de múltiplos efeitos fixos. Os resultados também dão suporte à hipótese de atenuação espacial das economias de aglomeração, sugerindo inferir que a relação de aglomeração mais relevante é aquela que circunscreve o local de trabalho do indivíduo, e não aquela que se dá em áreas vizinhas à sua célula.

Exercícios de heterogeneidade foram feitos. Uma vasta literatura em Economia discute o papel do capital humano e o impacto sobre produtividade. A literatura especializada argumenta que trabalhadores mais educados (em termos de anos de estudo) apresentam maior capacidade de absorção dos efeitos positivos da aglomeração quando comparados com trabalhadores com menos anos de estudo. Os testes estimados nesta tese mostram que trabalhadores com graduação completa, mestrado ou doutorado capturam maior parcela do efeito das economias de aglomeração do que trabalhadores com até o ensino médio completo. O efeito da aglomeração também foi avaliado para os diferentes SBD e NSBD encontrados no capítulo três. Os resultados apontam que SBD mais densos apresentam prêmios salariais menores do que aqueles menos densos, em termos de emprego. O NSBD é o que apresenta maior prêmio salarial intra-urbano relativamente aos SBD. Os resultados estimados sugerem que as economias de aglomeração são mais relevantes no NSBD, pois fornece um prêmio salarial maior do que um aumento marginal em qualquer SBD identificado.

Este trabalho dialoga de perto com a literatura de Economia Regional e Urbana e Economia do Trabalho e contribuiu em fornecer evidências adicionais sobre economias de aglomeração e seu impacto sobre salários. A estratégia empírica estimada com controles para múltiplos efeitos fixos aplicada aos microdados de uma grande base de dados longitudinal do mercado de trabalho formal da maior região metropolitana do Brasil é inovadora e nunca antes utilizada em nenhum trabalho empírico citado nessa tese.

Em síntese, esta tese desenvolveu uma abordagem metodológica objetiva de identificação de SBD, que permite avaliar a dinâmica do mercado de trabalho local. Ademais, identificou as regiões centrais, em termos de emprego, na RMSP. A geocodificação da base de dados da RAIS também é outro resultado relevante, que poderá ser utilizada em outros trabalhos que tenham como unidade de análise a RMSP e o foco esteja nas características espaciais do

mercado de trabalho. Um passo adiante foi dado ao estimar o prêmio salarial intra-urbano e testar a hipótese de atenuação espacial da aglomeração.

Novas pesquisas podem derivar tanto da nova base de dados criada quanto da abordagem metodológica de identificação de subcentralidades proposta. Com o uso da base de dados, é possível estudar o padrão de deslocamento do emprego para qualquer um dos municípios do Brasil. É possível, inclusive, observar se oferta de bens públicos, como metrô, tem efeito sobre os salários dos trabalhadores, uma vez que é possível conhecer a microlocalização do trabalhador no espaço e, também, saber se este indivíduo estava antes da estréia da estação e se permaneceu depois. Ademais, a partir dos SBD identificados, é possível reestimar modelos de preços hedônicos que estimam preços da terra e avaliar o efeito quando novos subcentros são considerados.

REFERÊNCIAS

ABOWD, J. M.; KRAMARZ, F.; MARGOLIS, D. N. High wage workers and high wage firms. **Econometrica**, v. 67, n. 2, p. 251-333, 1999.

AGUILERA, A.; MIGNOT, D. Urban sprawl, polycentrism and commuting. A comparison of seven French urban areas. **Urban Public Economics Review**, v. 1, p. 93-111, 2004.

AHLFELDT, G. M.; PIETROSTEFANI, E. The economic effects of density: A synthesis. (SERCDP0210). **CESifo Working Paper**, No 6744, 2017. Disponível em: https://www.econstor.eu/bitstream/10419/173020/1/cesifo1_wp6744.pdf. Acesso em 9 Jul. 2018.

AHLFELDT, G. M.; REDDING, S. J.; STURM, D. M.; WOLF, N. The economics of density: evidence from the Berlin wall. **Econometrica**, v. 83, n. 6, p. 2127-2189, 2016.

ALONSO, W. (1964). **Location and Land Use**. Cambridge: Harvard University Press.

ANAS, A.; ARNOTT, R.; SMALL, K. A. Urban spatial structure. **Journal of Economic Literature**, v. 36, n. 3, p. 1426-1464, 1999.

ANAS, A.; KIM, I. General equilibrium models of polycentric land use with endogenous congestion and job agglomeration. **Journal of Urban Economics**, v. 28, p. 318-325, 1996.

ANDERSON, N. B.; BOGART, W. T. Identifying and characterizing employment centers in polycentric metropolitan areas. **American Journal of Economics and Sociology**, v. 60, n. 1, 2001.

ANDERSON, N.; BOGART, W. The structure of sprawl: Identifying and characterizing employment centers in polycentric metropolitan areas. **American Journal of Economics and Sociology**, v. 60, p. 147-169, 2001.

ANDERSSON, M.; KLAESSON, J.; LARSSON, J. P. How Local are Spatial Density Externalities? Neighbourhood Effects in Agglomeration Economies. **Regional Studies**, v. 50, n. 6, p. 1082-1095, 2016.

ANSELIN, L., VARGA, A., ACS, Z. Local geographic spillovers between university research and high technology innovations. **Journal of Urban Economics**, v. 42, p. 422-448, 1997.

AUDRETSCH, D. B., FELDMAN, M. A. Knowledge spillovers and the geography of innovation and production. **American Economic Review**, v. 86, p. 630-640, 1996.

AUTANT-BERNARD, C. The geography of knowledge spillovers and technological proximity. **Economics of Innovation and New Technology**, v. 10, p. 237-254, 2001.

AZZONI, C. R. Teoria da Localização: uma análise crítica. São Paulo. IPE/USP, 200p. **Série Estudos Econômicos**, v. 19, 1982.

AZZONI, C.; SERVO, L. Education, cost of living and regional wage inequality in Brazil. **Papers in Regional Science**, v. 81, n. 2, p. 157-175, 2001.

BACOLOD, M.; BLUM, B. S.; STRANGE, W. C. Skill in the city. **Journal of Urban Economics**, v. 62, n. 2, p. 136-153, 2009.

BARROS, A. P. B. G.; KNEIB, E. C.; PAIVA, M.; TEDESCO, G. M. I. Analysis of trip generating developments by space syntax: a case study of Brasilia - Brazil. **Journal of Transport Literature**, v. 8, n. 3, p. 7-36, 2014.

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.

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

BATHELT, H.; MALMBERG, A.; MASKELL, P. Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. **Progress in Human Geography**, v. 28, p. 31-56, 2004.

BAUMONT, C. e BOURDON, F. Centres secondaires et recomposition économique des espace urbain, le cas de la communauté de l'agglomération dijonnaise (1990; 1999). **LATEC Working Paper 4**, Universidade de Bourgogne, Dijon, 2002.

BAUMONT, C., ERTUR, C. e GALLO, J. L. Spatial analysis of employment and population density: the case of the agglomeration of Dijon, 1999. **Geographical Analysis**, v. 36, n. 2, 2004.

BECKMANN, M. J. (1974). Spatial Equilibrium in the Housing Market. **Journal of Urban Economics**, v. 1, p. 99–107, 1974.

BENDER, B.; HWANG, H. Hedonic housing price indices and secondary employment centers. **Journal of Urban Economics**, v. 1, n. 17, p. 90-107, 1985.

BERLIANT, M.; KONISHI, H. The endogenous foundations of a city: population agglomeration and marketplaces in a location specific production economy. **Regional Science and Urban Economics**, v. 30, n. 3, p. 289-324, 2000.

BEYERS, W., Cyberspace or human space: wither cities in the age of telecommunications? In Aoyama, Y.; Wheeler, J.; Barney, W. (Eds.), **Cities in the Telecommunications Age: The Fracturing of Geographies**. New York: Routledge, 161-180, 2000.

BIDERMAN, C. Forças de atração e expulsão na Grande São Paulo. São Paulo: Escola de Administração de Empresas de São Paulo, Fundação Getúlio Vargas, 2001. (**Tese de Doutorado**).

BIVAND, R. S.; PEBESMA, E.; GÓMEZ-RUBIO, V. **Applied Spatial Data Analysis with R**, Springer-Verlag, New York, 2008.

Boamet, M. The monocentric model and employment location. **Journal of Urban Economics**, v. 36, p. 79-97, 1994.

BOGART, W. T.; FERRY, W. C. Employment centres in Greater Cleveland: evidence of evolution in a formerly monocentric city. **Urban Studies**, v. 36, n. 12, p. 2099-2110, 1999.

BOGART, W. T.; HWANG, H. Employment centers in Greater Cleveland: evidence of evolution in a formerly monocentric city. **Urban Studies**, v. 36, pp. 2099-2110, 1999.

BOURNE, L.S. Are new urban forms emerging? empirical tests for Canadian urban areas. **Canadian Geographer**, n. 33, p. 312-328, 1989.

BOWMAN, A. W.; AZZALINI, A. **Applied smoothing techniques for data analysis: the kernel approach with S-plus illustrations**. Oxford, Oxford Science Publications, 1997.

BRAKMAN, S.; GARRETSEN, H.; VAN MARREWIJK, C. **The New Introduction to Geographical Economics**. Cambridge, UK: Cambridge University Press, 2009.

BRIANT, A.; COMBES, P.; LAFOURCADE, M. Dots to boxes: do the size and shape of spatial units jeopardize economic geography estimations? **Journal of Urban Economic**, v. 67, p. 287-302, 2008.

BURGER, M. J., van OORT, F. G., van der KNAAP, G. A. A treatise on the geographical scale of agglomeration externalities and the modifiable areal unit problem. **Scienze Regionali**, v. 9, p. 19-40, 2010.

BURNS, M.C.; BOIX, M.; ROCA, J. Contrasting Indications of Polycentrism within Spain's Metropolitan Urban Regions. **Paper for the Eighth European Estate Society Conference**, Alicante, June 26-29, 2001.

CAIXA. **Programa de Integração Social – PIS**. Disponível em: <<http://www.caixa.gov.br/beneficios-trabalhador/pis/Paginas/default.aspx>>. Acesso em: 10 Mai. 2018.

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-c6f925639a711d35a1187c1b9c5483f3.pdf>>. Acesso em: 09, Jul. 2018.

CAMPOS, R. B. A. O mercado imobiliário residencial no município de São Paulo: uma abordagem de preços hedônicos espacial. **Nova Economia**, v.27, n.1, p. 303-337, 2017.

CAMPOS, R. B. A.; ALMEIDA, E. S. Decomposição espacial nos preços de imóveis residenciais no município de São Paulo. **Estudos Econômicos**, v. 48, n. 1, p. 5-38, 2018.

CAMPOS, R. B. A.; CHAGAS, A. S. Identificação e caracterização das subcentralidades de emprego no Município de São Paulo. In: 45º Encontro Nacional de Economia, Natal, 2017. **Anais...** Natal: ANPEC, 2017.

CASTELLS, M.; HALL, P.; HALL, P. G. **Technopoles of the world: the making of twenty-first-century industrial complexes**: Routledge, 1994.

CEM. **Shapefile Logradouro da Região Metropolitana**, edição 2016. São Paulo: CEM/CEBRAP, 2016.

CERVERO R., **America's Suburban Centers: The Land Use-Transportation Link**: Unwin Hyman, Winchester, MA. 1989.

CERVERO, R.; WU, K-L. Polycentrism, commuting, and residential location in the San Francisco Bay area. **Environment and Planning A**, v. 29, p. 865-886, 1997.

CHRISTALLER, W. **Central places in southern Germany**. Prentice Hall, Englewood Cliffs, 1966.

CICCONE A, H.. Productivity and the density of economic activity. **American Economic Review**, v. 86, n. 1, p. 54-70. 1996.

CLARK, W.A.V.; KUIJPERS-LINDE, M. Commuting in Restructuring Urban Regions **Urban Studies**, v. 31, pp. 465-483, 1994.

COFFEY, W. J. e SHEARMUR, R. G. The identification of employment centers in Canadian metropolitan areas: the example of Montreal, 1996. **The Canadian Geographer**, v. 45, nº 3, p. 371-386, 2001.

COHEN, W. M.; LEVINTHAL, D. A. Absortive capacity: a new perspective on learning and innovation. **Administrative Science Quarterly**, v. 35, p. 128-152, 1990.

COMBES, P.-P., DURANTON, G., GOBILLON, L. Spatial wage disparities: Sorting matters! **Journal of Urban Economics**, v. 63, n. 2, p. 723-742, 2008.

COMBES, P.P., MAYER, T., THISSE, J.F. **Economic Geography**. Princeton: Princeton University Press, 1st Edition, 2008.

COMBES, P.P.; GOBILLON, L. The empirics of agglomeration economies. In: **Handbook of Regional and Urban Economics**, v. 5A, North-Holland, Amsterdam, 1st Edition, pp. 247-348, 2015.

COOKE, T. Causality reconsidered: a note. **Journal of Urban Economics**, v. 5, p. 538-542, 1978.

CORREIOS. Disponível em <<http://www.correios.com.br/precisa-de-ajuda/o-que-e-cep-e-por-que-usa-lo/estrutura-do-cep>>. Acesso em 08 Julho 2018.

CRAIG, S. G., NG, P. T. Using quantile smoothing splines to identify employment subcentres in a multicentric urban area. **Journal of Urban Economics**, v. 49, 2001.

CRUZ, B. O.; NATICCHIONI, P. Falling urban wage premium and inequality trends: evidence for Brazil. **Investigaciones Regionales**, v. 24, p. 91-113, 2012.

DINIZ, C. C. Desenvolvimento polygonal no Brasil: nem desconcentração nem contínua polarização. **Revista Nova Economia**, v. 3, n. 1, p. 35-64, 1993.

DINIZ, C. C.; CAMPOLINA, B. A região metropolitana de São Paulo: reestruturação, re-espacialização e novas funções. **Revista Eure**, v. 33, n. 98, p. 27-43, 2007.

DIXIT, A. K.; STIGLITZ, J. E. Monopolistic competition and optimum product diversity. **American Economic Review**, v. 67, n. 3, p. 297-308, 1976.

DURANTON, G.; PUGA, D. Micro-foundations of urban agglomeration economies. In: **Handbook of regional and urban economics**. Elsevier. p. 2063-2117, 2004.

DURLAUF, S. N. Neighborhood effects. In Henderson, V.; Thisse, J.F (Eds.) **Handbook of Regional and Urban Economics**. Amsterdam: North-Holland. 2003.

EBERTS, R. W. An empirical investigation of intraurban wage gradients. **Journal of Urban Economics**, vol. 10, p. 50-60, 1981.

FABER, S. and PÁEZ, A. A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations. **Journal of Geographical Systems**, v. 9, p. 371-396, 2007.

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.

FAUSTO, B. **História do Brasil**. São Paulo: Editora da USP. 1994.

FELDMAN, M., AUDRETSCH, D., 1999. Innovation in cities: science-based diversity, specialization and localized competition. **European Economic Review**, v. 43, p. 409-429, 1999.

FERREIRA, C. M. C. Espaço, regiões e economia regional. In: Haddad, P. R. (Org.). **Economia regional: teorias e métodos de análise**. Fortaleza: Banco do Nordeste do Brasil, 1989.

FOTHERINGHAM, A. A.; BRUNSDON, C.; CHARLTON, M (2000). **Quantitative geography: perspective on spatial data analysis**. Sage, London, England.

FOTHERINGHAM, A. A.; BRUNSDON, C.; CHARLTON, M (2002). **Geographically Weighted Regression: the analysis of spatially varying relationship**. Wiley, New York, NY, USA.

FREGUGLICA, R. S.; PROCÓPIO, T. S. Efeito 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**, v. 43, n. 2, 2013.

FUJITA, M. A monopolistic competition model of spatial agglomeration: Differentiated products approach. **Regional Science and Urban Economics**, v. 18, p. 87-124, 1988.

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

FUJITA, M.; OGAWA, H. Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations. **Regional Science and Urban Economics**, v. 12, n. 2, p. 161-191, 1982.

FUNDAÇÃO INSTITUTO DE PESQUISAS ECONÔMICAS – FIPE. **Índice de Preços ao Consumidor – IPC-Fipe**. Disponível em: <<http://www.fipe.org.br/pt-br/indices/ipc/>>. Acesso em: 08 Jul. 2018.

GARREAU, J. **Edge City: Life on the New Frontier**. New York: Doubleday. 1991.

GETIS, A. Second-order analysis of point patterns: the case of Chicago as a multi-center urban region. **Professional Geographer**, v. 35, n. 1, p. 73-80, 1983.

GIULIANO, G.; SMALL, K. A. Subcenters in the Los Angeles region. **Regional Science and Urban Economics**, v. 21, p. 163-182, 1991.

GIULIANO; G. et al. Not all sprawl: evolution of employment concentrations in Los Angeles, 1980-2000. **Working Paper 2005-1002**, Lusk Center for Real Estate, University of Southern California. 2005.

GLAESER, E. Cities, information and economic growth. **Cityscape**, v. 1, p. 9-47, 1999.

GLAESER, E. L.; SCHEINKMAN, J. **Non-market interactions**: National Bureau of Economic Research, 2000.

GLAESER, E. O Triunfo da Cidade. 2ª ed. São Paulo: BEI Comunicação. 2016.

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, pp. 221-244, 2010.

GLAESER, E.; RESSEGER, M.; TOBIO, K. Inequality in cities. **Journal of Regional Science**, v. 49, n. 4, p. 617-646, 2009.

GLASER, E. **Triumph of the city: how our greatest invention make us richer, smarter, greener, healthier, and happier**. The Penguin Press. 2011.

GORDON, I. R.; McCANN, P. Industrial clusters: complexes, agglomeration, and/or social networks? **Urban Studies**, v. 27, p. 513-532, 2000.

GORDON, P., RICHARDSON, H. and WONG, H. The distribution of population and employment in a polycentric city: the case of Los Angeles, **Environment and Planning A**, v. 18, p. 161-173, 1986.

GORDON, P.; RICHARDSON, H. W. Beyond polycentricity. The dispersed metropolis, Los Angeles, 1970-1990. **Journal of the American Planning Association**, v. 62, n° 3, pp. 289-295, 1996.

GOTTLIEB, P. Residential amenities, firm location and economic development. **Urban Studies**, v. 32, n. 9, p. 1413-1436, 1995.

GRAHAM, S.; MARVIN, S. **Telecommunications and the City: Electronic Spaces, Urban Places**. London: Routledge. 1996.

GRANOVETTER, M. **Getting a job: a study in contacts and careers**. Chicago: University of Chicago Press, 1995.

GREENE, D. Recent trends in urban spatial structure. **Growth and Change**, v. 10, p. 29-40, 1980.

GRIFFITH, D. A. Evaluation the transformation from a monocentric to a polycentric city. **Professional Geographer**, v. 33, n. 2, p. 198-196, 1981.

GROOT, S. P. T; DE GROOT, H. L, F.; SMIT, M. J. Regional wage differences in the Netherlands: micro evidence on agglomeration externalities. **Journal of Regional Science**, v. 54, n. 3, p. 503-523, 2014.

GUILLAIN, R.; GALLO, J. BOITEUX-ORAIN, C. Changes in spatial and sectoral patterns of employment in Ile-de-France, 1978-97. **Urban Studies**, v. 43, n. 11, p. 2017-2098, 2006.

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. 628-648, 2010.

HADDAD, E. A.; BARUFI, A. M. B. From rivers to roads: spatial mismatch and inequality of opportunity in urban labor markets of a megacity. **Habitat International**, v. 68, p. 3-14, 2017.

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

HALSTEAD, J.M.; DELLER, S.C. Public Infrastructure in Economic Development and Growth: Evidence from Rural Manufacturers. **Journal of the Community Development Society**, v. 28, p. 97-116, 1997.

HARTWICK, P., HARTWICK, J. M. Efficient Resource Allocation in a Multinucleated City with Intermediate Goods. **Quarterly Journal of Economics**, v. 88, p. 340-352, 1974.

HEIKKILA, E.; GORDON, P.; KIM, J. I. What happened to the CBD-distance gradient?: land rents in a polycentric city. **Environment and Planning A**, v. 21, pp. 221-232, 1989.

HELSLEY, R. W.; SULLIVAN, A. M. Urban Subcenter Formation. **Regional Science and Urban Economics**, v. 21, p. 255-275, 1991.

HENDERSON, J. V. Industrial development in cities. **Journal of Political Economy**, v. 102, p. 1067-1091, 1991.

HENDERSON, J. V. Marshall's scale economies. **Journal of Urban Economics**, v. 53, p. 1-28, 2003.

HENDERSON, J. V.; MITRA, A. The new urban landscape: developers and edge cities. **Regional Science and Urban Economics**, v. 26, n. 6, p. 613-643, 1996.

HENDERSON, J. V.; SLADE, E. Development games in non-monocentric cities. **Journal of Urban Economics**, v. 34, p. 207-230, 1993.

HOTCHKISS, D.; WHITE, M. A simulation model of a decentralized metropolitan area with two-worker, 'traditional' and female-headed household. **Journal of Urban Economics**, n. 34, pp. 159-185, 1993.

ILHANFELDT, K. R. Intraurban wage gradients: evidence by race, gender, occupational class, and sector. **Journal of Urban Economics**, v. 32, p. 70-91, 1992.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Censo Demográfico**, 2000; 2010. Disponível em: <<https://ww2.ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm>>. Acesso em: 08 Jul. 2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. **Pesquisa Nacional por Amostra de Domicílios (PNAD)**. Centro de Estudos da Metrópole (CEM). 2002, 2004, 2014. Disponível em: <<http://web.fflch.usp.br/centrodametropole/1148>>. Acesso em: 08 Jul. 2018.

ISARD, W. **Location in a Space Economy**. Cambridge, Massachusetts: Massachusetts Institute of Technology Press. 1956.

JACOBS, J. **The economy of cities**. Nova York: Random House, 1969.

JAFFE, A; TRAJTENBERG, M.; HENDERSON, R. Geographic localization of knowledge spillovers as evidenced by patent citations, **Quarterly Journal of Economics**, 108, 577-598 1993 .

JEFFE, V.; TRAJTENBERG, M.; HENDERSON, R. Geographic localization of knowledge spillovers. **Quarterly Journal of Economics**, v. 108, p. 577-599, 1993.

KANE, K.; HIPPI, J. R.; KIM, J. H. Los Angeles employment concentration in the 21st century. **Urban Studies**, p. 1-26, 2016.

KIM, S. Labor specialization and the extent of the market. **Journal of Political Economy**, v. 97, p. 692-705, 1991.

KLOOSTERMAN, R.; MUSTERD, S. Polycentric urban region: as a research concept. **Urban Studies**, v. 38, n. 4, p. 619-629, 2001.

KNEIB, E. C. "Centralidades urbanas e sistemas de transporte público em Goiânia, Goiás." **Revista Brasileira de Gestão Urbana**, v. 8, n. 3, 2016.

KNEIB, E. C. Subcentros urbanos: contribuição conceitual e metodológica à sua definição e identificação para planejamento de transportes (**Tese de doutorado**). Universidade de Brasília, Brasília, 2008.

KRUGMAN, P. Increasing return and economic geography. **Journal of Political Economy**, n. 99, p. 448-99, 1991.

LAVETTI, K.; SCHMUTTE, I. M. Estimating compensating wage differentials with endogenous job mobility. **Working Paper**, Ohio State University, 2016.

LEMOES, M. B.; DINIZ, C. C., GUERRA, L. P. A nova configuração regional brasileira e sua geografia econômica. **Estudos Econômicos**, v. 33, n. 4, p. 665-700, 2003.

LÓPEZ, M. A. G.; MUÑOZ, I. Employment decentralisation: polycentric compaction or sprawl? the case of the barcelona metropolitan region 1986-1996', **Document de Treball**, Facultat de Ciències Econòmiques i Empresariales, Universitat Autònoma de Barcelona, 2005.

LOPEZ, M. G. Estructura espacial del empleo y economías de aglomeración: el caso de la industria en la región metropolitana de Barcelona. **Arquitectura, Ciudad y Entorno**, v. 2, n. 4, p. 519-553, 2007.

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

LUCAS, R. E. ROSSI-HANSBERG, E. On the internal structure of cities. *Econometrica*, v. 70, n. 4, p. 1445-1476, 2002.

MACIEL, V.; BIDERMAN, C. Assessing the effects of the São Paulo's metropolitan beltway on residential land prices. **Journal of Transport Literature**, v. 7, n. 2, p. 373-402, 2013.

MARSHALL, A. **Principles of economics**. London: Macmillan, 1890.

MCDONALD, J. F. The identification of urban employment subcenters. **Journal of Urban Economics**, v. 21, pp. 242-258, 1987.

MCDONALD, J. F.; MCMILLEN, D. P. Employment subcenters and land values in a polycentric urban area: the case of Chicago. **Environment and Planning A**, v. 22, p. 1561-1574, 1990.

MCDONALD, J. F.; PRATHER, P. J. Suburban employment centres: the case of Chicago. **Urban Studies**, v. 31, n. 2, p. 201-218, 1994.

MCMILLEN, D. P. Nonparametric employment subcenter identification. **Journal of Urban Economics**, v. 50, p. 448-473, 2001a.

MCMILLEN, D. P. Polycentric urban structure: the case of Milwaukee. **Economic Perspectives**. Federal Reserve Bank of Chicago, 2001b.

McMILLEN, D. P. Identifying sub-centres using contiguity matrices. **Urban Studies**, v. 40, p. 57-69, 2003.

McMILLEN, D. P. e SMITH, S. C. The number of subcenters in large urban areas. **Journal of Urban Economics**, v. 53, nº 3, p. 321-338, 2003.

MCMILLEN, D. P.; MCDONALD, J.F. Suburban subcenters and employment density in metropolitan Chicago. **Journal of Urban Economics**, v. 43, p. 157-180, 1998.

MCMILLEN, D.; D. P.; MCDONALD, J. F. A nonparametric analysis of employment density in a polycentric city. **Journal of Regional Science**, v. 37, n. 4, p. 591-612, 1997.

MCMILLEN, D.; SINGELL, L. D. Work location, residence location and the intraurban wage gradient. **Journal of Urban Economics**, v. 32, p. 195-213, 1992.

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**, v. 36, n. 3, p. 449-470, 2006.

MEYER, R; GROSTEIN, M. D.; BIDERMAN, C. **São Paulo Metrópole**. Editora da Universidade de São Paulo: Imprensa Oficial do Estado de São Paulo, 2004.

MIESZKOWSKI, P. e SMITH, B. Analyzing urban decentralization: the case of Houston. **Regional Science and Urban Economics**, v. 21, n. 2, p. 183-199, 1991.

MIESZKOWSKI, P. MILLS, E. S. The causes of metropolitan suburbanization. **Journal of Economic Perspectives**, v. 7, n. 3, p. 135-147, 1993.

MILLS, E. S. An Aggregative Model of Resource Allocation in a Metropolitan Area. **American Economic Review**, v. 57, n. 2, p. 197-210, 1967.

MILLS, E. S. **Studies in the Structure of the Urban Economy**. Baltimore: John Hopkins University Press. 1972.

MILLS, E.S.; PRICE, R. Metropolitan suburbanization and central city problems. **Journal of Urban Economics**, v. 15, p. 1-17, 1984.

MINCER, J.. Schooling, Experience, and Earnings. Human Behavior & Social Institutions No. 2. 1974. Disponível em: < <https://eric.ed.gov/?id=ED103621>>. Acesso em 08 de Julho de 2018.

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

MION, G., NATICCHIONI, P. The spatial sorting and matching of skills and firms. **Canadian Journal of Economics**, v. 42, n. 1, p. 28-55, 2009.

MODARRES, A. Policentricity and transit service. **Transportation Research Part A**, v. 37, p. 841-864, 2003.

MORETTI, E. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. **Journal of Econometrics**, v. 121, p. 175-212, 2004a.

MORETTI, E. Human capital externalities in cities. In Henderson, J. V.; Thisse, J.F. (Eds.), **Handbook of Regional and Urban Economics**. Amsterdam: Elsevier, 2004b.

MUÑIZ, I.; GALINDO, A.; GARCÍA, M. A. És Barcelona uma ciudad polcéntrica? **Document de Treball** No. 0390, Departament d'Economia Aplicada, Universitat Autònoma de Barcelona, 2003.

MUTH, R. F. **Cities and Housing**. Chicago: University of Chicago Press. 1969.

MUTH, R. F. Numerical Solution of Urban Land-Use Models. **Journal of Urban Economics**, v. 2, p. 307–332, 1975.

NADALIN, V. **Evolução urbana e espraiamento na Região Metropolitana de São Paulo** In: Nadalin, V. Três ensaios sobre economia urbana e mercado de habitação em São Paulo. São Paulo: Programa de Pós-Graduação em Economia; Faculdade de Economia, Administração e Contabilidade; Universidade de São Paulo, 2010.

OGAWA, H., e FUJITA, M. Equilibrium land use patterns in a non-monocentric city. **Regional Science and Urban Economics**, v. 20, nº 4, p. 455–475, 1980.

OPENSHAW, S.; TAYLOR, P. J. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In Wrigley, N. (Ed.). **Statistical Applications in the Spatial Sciences**, p. 127-144. Pion, London.

OVERMAN, H. Can we learn anything from economic geography proper? **Journal of Economic Geography**, v. 4, p. 501–516, 2004.

PAN; Q; MA, L. Employment subcenter identification: a GIS-based method. Texas Southern University. **Science and Urban Economics**, v. 21, n. 2, p. 63-82, 2005.

PAPAGEORGIOU, G. J. Rent distribution models within a multicentral framework. **Environment and Planning A**, v. 3, p. 267–282, 1971.

RAMOS, F. R. Análise espacial de estruturas intra-urbanas: o caso de São Paulo. **Dissertação de Mestrado** em sensoriamento Remoto. INPE, São Jose dos Campos, SP. 2004.

READFEARN, C. L. The topography of metropolitan employment: identifying centers of employment in a polycentric urban area. **Journal of Urbana Economics**, v. 61, p. 519-541, 2007.

REDFEARN, C. The topography of metropolitan employment: identifying centers of employment in a polycentric urban area. **Journal of Urban Economics**, v. 61, p. 519-541, 2007.

REES, A.; SCHULTS, G. P. **Workers and Wages in an Urban Labor Market**. Chicago, USA: University of Chicago Press, 1970.

RICHARDSON, H. W. et al. Residential property values, the CBD, and multiple nodes: Further analysis. **Environment and Planning A**, v. 22, p. 829–833, 1990.

ROBACK, J. Wages, rents, and the quality of life. **The Journal of Political Economy**, v. 90, n. 6, p. 1257-1278, 1982.

ROCHA, R. M.; SILVEIRA NETO, R. M.; GOMES, S. M. F. P. O. Maiores cidades, maiores habilidades produtivas: ganhos de aglomeração ou atração de habilitados? Uma análise para as cidades brasileiras. **Revista Econômica do Nordeste**, vol. 42, No. 4, p. 675-695, 2011.

ROMANOS, M. C. Household Location in a Linear Multi-Center Metropolitan Area. **Regional Science and Urban Economics**, vol. 7, p. 233-250, 1979.

ROSENTHAL, S. S., STRANGE, W. C. Geography, industrial organization and agglomeration. **Review of Economics and Statistics**, v. 84, p. 377-393, 2003.

ROSENTHAL, S. S.; STRANGE W. C. The attenuation of human capital spillovers. **Journal of Urban Economics**, v. 64, p. 373-389, 2008.

ROSENTHAL S.; STRANGE W., Evidence on the nature and sources of agglomeration economies. In: Henderson, J.V., Thisse, J.F. (Eds.), **Handbook of Regional and Urban Economics 4**. Elsevier, Amsterdam, pp. 2119-2171, 2004.

ROSS, S; YINGER, J. Comparative static analysis of open urban models with a full labor market and suburban employment. **Regional Science and Urban Economics**, n. 25, pp. 575-605, 1995.

SASAKI, K., e MUN, S. A Dynamic Analysis of a Multiple-Center Formation in a City. **Journal of Urban Economics**, v. 40, n° 3, p. 257-278, 1996.

SHEAMUR, R.; COFFEY, W. J. A tale of four cities: intrametropolitan employment distribution in Toronto, Montreal, Vancouver, and Ottawa-Hull, 1981-1996. **Enviroment and Planning**, v. 34, p. 575-598, 2002.

SHUKLA, V.; WADDELL, P. Firm location and land use in discrete urban space: a study of the spatial structure of Dallas-Fort Worth. **Regional Science and Urban Economics**, v. 21, p. 225-253, 1991.

SILVA, D. L. G. Economias de aglomeração e heterogeneidade de trabalhador e firma na determinação de salários no Brasil. São Paulo: Programa de Pós-Graduação em Economia; Faculdade de Economia, Administração e Contabilidade; Universidade de São Paulo, 2017. (Tese de Doutorado).

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.

SINGER, P. **Economia política da urbanização**. São Paulo: Editora Contexto, 3ª ed. 2017.

SIQUEIRA, L. P. R. Proposta metodológica para a identificação de subcentros urbanos: estudo de caso na Região Metropolitana de São Paulo. **Economia Aplicada**, v. 18, n. 1, p. 139-165, 2014.

SIVITANIDOU, R.; WHEATON, W.C. Wage and rent capitalization in the commercial real estate market. **Journal of Urban Economics**, n. 31, pp. 206-229, 1992.

SMALL, K. A.; SONG, S. Population and employment densities: structure and change. **Journal of Urban Economics**, v. 36, p. 292-313, 1994.

SMOLKA, M. O. O espaço do espaço na teoria econômica. **Literatura Econômica**, Rio de Janeiro, IPEA, v. 5, n. 6, p. 705-28, 1983.

SOLOW, R. Congestion costs and the use of land for streets. **Bell Journal**, v. 4, n. 2, p. 602-618, 1973.

STEINNES, D. Causality and intraurban location. **Journal of Urban Economics**, v. 4, p. 69-79, 1977.

STOCK, J.; YOGO, M. Testing for weak instruments in linear IV regression. **Working paper**, Harvard University, 2003.

STORPER, M.; VENABLES, A. J. Buzz: face-to-face contact and the urban economy. **Journal of Economic Geography**, v. 4, p. 351-370, 2004.

SULLIVAN, A. A general equilibrium model with agglomerative economies and decentralized employment. **Journal of Urban Economics**, v. 20, p. 55-74, 1986.

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

TIMOTHY, D.; WHEATON, W. C. Intra-urban wage variation, employment location and commuting times. **Journal of Urban Economics**, v. 50, p. 338-366, 2001.

VAN SOEST, D., GERKING S.; VAN OORT F. G. Spatial impacts of agglomeration externalities. **Journal of Regional Science**, v. 46, p. 881-899, 2006.

WHEATON, W. C. A Comparative Static Analysis of Urban Spatial Structure. **Journal of Economic Theory**, v. 9, p. 223-237, 1974.

WHEATON, W. Commuting, congestion, and employment dispersal in cities with mixed land use. **Journal of Urban Economics**, v. 55, n. 3, p.417-438, 2004.

WHITE, M. J. Firm Suburbanization and Urban Subcenters. **Journal of Urban Economics**, v. 3, n. 4, p. 323-343, 1976.

WHITE, M. J. Location Choice Behavior and Commuting Behavior in Cities with Decentralized Employment. **Journal of Urban Economics**, v. 24, p. 129-152, 1988.

WHITE, M. J. Commuting and congestion: A simulation model of a decentralized metropolitan area. **Journal of the American Real Estate and Urban Economics Association**, n. 18, pp. 335-368, 1990.

WHITE, M. J. Urban areas with decentralized employment: Theory and empirical work. In Chesire, P.; Mills E. S. (Eds.) **Handbook of Regional and Urban Economics**, v. 3 Applied Urban Economics, North-Holland, pp. 1375-1412, 1999.

WIEAND, K. An extension of the monocentric urban spatial equilibrium model to a multicenter setting: The case of the two center city. **Journal of Urban Economics**, v. 21, p. 259-271, 1987.

WINTERS, J.V. Human capital externalities and employment differences across metropolitan areas of the USA. **Journal of Economic Geography**, v. 13, n. 5, p. 799–822, 2013.

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

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

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data**. Cambridge: Massachusetts Institute of Technology – MIT Press, 2010.

WREDE, M. A continuous spatial choice logit model of polycentric city. **Regional Science and Urban Economics**, vol. 53, p. 68-73, 2015.

WRIGLEY, N. Analysing, modelling, and resolving the ecological fallacy. In Longley, P.; Batty, M. (Eds.) **Spatial Analysis: Modelling in a GIS Environment**, New York, John Wiley Sons, pp. 25–41, 1996.

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

YINGER, J. Urban Models with More than One Employment Center. **Journal of Urban Economics**, v. 31, n° 2, p. 181–205, 1992.

ZHANG, Y.; KOMEI, S. Effects of subcenter formation on urban spatial structure. **Regional Science and Urban Economics**, v. 27, n. 3, p. 297- 324, 1997.

ZHANG, Y.; KOMEI, S. Spatial structure in an open city with a subcenter. **Annals of Regional Science**, v. 34, n. 1, p. 37–53, 2000.

APÊNDICE

APÊNDICE A – Quadro-Resumo das Abordagens Empíricas

Tabela A.1: Revisão dos Trabalhos Empíricos por Método Empírico

Autor	Variável Dependente	Base de Dados	Divisão Espacial	Região / Ano	Resultado (Nº SBD)
Valor de Corte					
Greene (1980)	Densidade de emprego	US Census Bureau	Setor Censitário	Atlanta (EUA) / 1960 e 1970	5 e 9
				Buffalo (EUA) / 1960 e 1970	25 e 33
				Denver (EUA) / 1960 e 1970	22 e 20
				Fort Worth (EUA) / 1960 e 1970	5 e 9
				Baltimore (EUA) / 1960 e 1970	10 e 12
Giuliano e Small (1991)	Densidade de emprego Número de empregos	Census journey-to-work	Zonas de Análise de Transporte (AZs) - 1146 zonas	Los Angeles (EUA) / 1980	32
Small e Song (1994)	Densidade de emprego Número de empregos	California Department of Transportation	Zonas de Análise de Transporte (AZs) - 1135 zonas	RM Los Angeles (EUA) / 1980	6
Cervelo e Wu (1997)	Densidade de emprego Número de empregos	Census Transportation Planning	Áreas de Ponderação - 1382 observações	São Francisco (EUA) / 1990	22
Bogart e Ferry (1999)	Densidade de emprego Número de empregos	US Census Bureau	Zonas de Análise de Transporte (AZs) - 1146 zonas	Greater Cleveland (EUA) / 1990	9
Anderson e Borgart (2001)	Densidade de emprego	US Census Bureau	Zonas de Análise de Transporte (AZs)	Cleveland (EUA)/1990	9
				Indiánapolis	11

	Número de empregos				(EUA)/1990 Portland	11
					(EUA) / 1990 Saint Luis (EUA) / 1990	11
Shearmur e Coffey (2002)	Número de Emprego	Statistics Canada	Áreas de Ponderação - 613 observações	Áreas de Ponderação - 530 observações	Montreal (CAN) / 1996 Toronto (CAN) / 1996	16
			Áreas de Ponderação - 161 observações	Áreas de Ponderação - 214 observações	Ottawa-Hull (CAN)/1996 Vancouver (CAN) / 1996	17 7 13
Boumont e Bourdon (2002)	Densidade de emprego Número de empregos	Système Informatique pour le Répertoire des Entreprises et des Etablissements (SIRENE)	Áreas de Ponderação - 107 IRIS	Áreas de Ponderação - 107 IRIS	Dijon (FRA) / 1990 Dijon (FRA) / 1999	6 6
Pan e Ma (2006)	Densidade de emprego Número de Emprego Utilização de SIG - CONFIRMAR	Censo 1990 e Censo 2000	Zonas de Análise de Transporte (AZs)		Houston (EUA) / 1990 Houston (EUA) / 2000	11 12
Econometria						
McDonald (1987)	Densidade de emprego Taxa de emprego/população	Census Population Chicago Area Transportation Housing Journey to Work Survey	Chicago Area Transportation (CATS) - 44 zonas		RM Chicago (EUA) / 1970	9
Heikkila et al. (1989)	Valor da terra	Damar Corporation of Los Angeles	Sector Censitário - 10 928		Los Angeles (EUA) / 1980	8
Gordon, Richardson and Wong (1986)	Densidade de emprego Densidade de população	US Census Bureau	Áreas de Ponderação - 2403 observations		Los Angeles (EUA) / 1970 Los Angeles (EUA) / 1980	-
Shukla and Waddell (1991)	Probab. localização	County Business Patterns	Zip codes - 141 zips		Dallas-Fort Worth (EUA) / 1987	
McDonald e Prather (1994)	Densidade de emprego	Urban Transportation Planning Package	Quarter-square-mile areas (.5 mile x .5 mile) - 1616 observations		Chicago (EUA) / 1980	3
McMillen e McDonald (1998)	Densidade de emprego	Northeastern Illinois Plannig Commission	Square miles		Suburbio de Chicago (EUA) / 1990	20

Craig e Ng (2001)	Densidade de emprego	de	US Census Bureau	Census tracts - 578 observations	Harris County (EUA) /	7
McMillen (2001a)	Densidade de Emprego	de	Census Transportation Planning Package	Transportation Analysis Zone (TAZ)	Milwaukee (EUA) / 1990	1
McMillen (2001b)	Densidade de Emprego	de	Census Transportation Planning Package	Transportation Analysis Zone (TAZ) - 11,752 TAZ	Chicago (EUA) / 1990	
				Transportation Analysis Zone (TAZ) - 5,950 TAZ	Dallas (EUA) / 1990	33
				Transportation Analysis Zone (TAZ) - 2,512 TAZ	Houston (EUA) / 1990	28
				Transportation Analysis Zone (TAZ) - 3,288 TAZ	Los Angeles (EUA) / 1990	25
				Transportation Analysis Zone (TAZ) - 498 TAZ	Nova Orleans (EUA) / 1990	19
				Transportation Analysis Zone (TAZ) - 3,127 TAZ	São Francisco (EUA) / 1990	2
					São Francisco (EUA) / 1990	22
McMillen (2003)	Densidade de Emprego	de	Census Transportation Planning Package	Transportation Analysis Zone (TAZ)	Atlanta (EUA)	4
					Baltimore-Washington (EUA)	18
					Boston (EUA)	11
					New York (EUA)	38
McMillen e Smith (2003)	Densidade de Emprego	de	Census Transportation Planning Package	Transportation Analysis Zone (TAZ)	Philadelphia (EUA)	4
					62 cidades	Tabela 1, p. 328
Muniz et al. (2003)	Densidade de emprego	de	Spain Census Bureau	Census tracts - 3481 observations	RM Barcelona (ESP) / 1996	22
	Número de Emprego	de				
Lopez (2007)	Densidade de emprego	de	Seguridad Social	Não Informado	RM Barcelona (ESP) / 1991	83
					RM Barcelona (ESP) / 2003	119
Redfearn (2007)	Densidade de emprego	de	Southern California Association of Governments	Area de Ponderação - 2349 observations	Los Angeles (EUA) / 2000	41
Kane, Hipp and Kim (2016)	Nível de Emprego		Reference USA	Geocoded firms	Los Angeles (EUA) / 1997 e 2014	

Estatística Espacial

Modarres (2003)	Nível de Emprego	US Census	Census tracts -	Los Angeles (EUA) / 2001	124 clusters de 1ª ordem 18 clusters de 2ª ordem
Ramos (2004)	Nível de Emprego	Pesquisas Origem-Destino	Área de Ponderação	RM de São Paulo (BRA) / 1997	-
Boumont, Ertur e Gallo (2004)	Nível e Densidade de Emprego	Recensement Général de la Population Census Système Infonnatique pour le Répertoire des Entreprises et de Etablissements	Census tracts - 136 observations	Dijon (FRA) / 1999	2
Guillain et al. (2006)	Employment/population rate	Population Census Regional Employment Surveys	1300 áreas geográficas	Ile-de-France (FRA) / 1978 Ile-de-France (FRA) / 1997	3 7
Kneib (2008)	Gerador de Viagens Deslocamento Pendular Uso do Solo	Pesquisa Origem-Destino Censo	Área de Ponderação Setor Censitário	RM de Manaus (BRA)	10 12 7
Siqueira (2014)	Nível e Densidade de Emprego	Pesquisa Origem-Destino Censo	Área de Ponderação - 460 observações	RM São Paulo (BRA) / 2007	6
Método do Especialista					
Freitas et al. (2013)	Conjunto de respostas	de Questionários	-	Salvador (BRA)	3
Tedesco and Kneib (2013)	Conjunto de respostas	de Questionários	-	Manaus (BRA) Brasília (BRA)	22 25
Barros et al. (2014)	Conjunto de respostas	de Questionários	-	Brasília (BRA)	16
Kneib (2016)	Conjunto de respostas	de Questionários	-	Goiania (BRA)	10

Fonte: Elaboração própria a partir de artigos publicados em revistas acadêmicas

APÊNDICE B – Tratamento dos Microdados da RAIS para Georreferenciamento

A Tabela B.1 mostra o efeito das exclusões de observações sobre o total de firmas e trabalhadores nas bases anuais RAIS-firma (2002-2014) considerando cada estratégia de limpeza da base como Etapa. E, a Tabela B.2 traz informações sobre o impacto de cada etapa de ajuste da base RAIS-firma no número total de trabalhadores, considerando as bases anuais RAIS-trabalhador (2002-2014). Cabe ressaltar que a decisão por excluir observações se deu face as dificuldades de identificar informações-chaves para a pesquisa. O estudo tem como foco firmas que tenham funcionários no ano-base, isto é, RAIS Negativa diferente de 1. Assim, a Etapa I exclui firmas que não teve funcionários no ano-base. A Etapa II deleta observações de firmas cujo mesmo CEI e CNPJ identifiquem mais do uma firma no mesmo ano-base. Na Etapa III foram excluídas firmas cujos algarismos da coluna Região dos CEPs são diferentes de zero e o campo município identifique a firma como pertencente a RMSP (como descrito acima, o vetor do código postal na RMSP inicia por zero). A Etapa IV: dentre os 39 municípios que compõem a RMSP 14 apresentavam CEP Únicos no período analisado (apenas Arujá e Caeiras oficializaram o arruamento e passaram a ter CEP Múltiplos, e por isso foram considerados o ano-base do recebimento para geocodificar as firmas dos períodos anteriores). Portanto, os demais municípios que compõem o grupo CEP Únicos e não foram geocodificados usando a estratégia de endereço completo foram excluídos da base. Tal tratamento revela-se pouco relevante dada a composição dos empregos nesses municípios no total da amostra da RMSP, como pode ser observado na Tabela B.3. Após a geocodificação, na Etapa V, exclui-se firmas cuja microlocalização, o par latitude-longitude, identifique regiões fora do respectivo município (correção do processo de geocodificação). Na Etapa IV foram excluídas firmas cujos trabalhadores pertencessem aos setores: público, construção civil, eletricidade, gás e água, captação, tratamento e distribuição de água, e transporte.

Tabela B.1 – Tratamento dos microdados de estabelecimentos da RAIS: 2002-2014

Ano	RAIS	Etapa I	Etapa II	Etapa III	Etapa IV	Etapa V	Etapa VI	Etapa VI / Etapa I
2002	828.286	294.086	266.112	264.707	255.610	253.238	238.918	81%
2003	855.411	298.287	269.186	267.754	259.094	256.845	242.473	81%
2004	886.636	304.015	276.463	274.943	266.589	264.227	249.421	82%
2005	919.692	313.074	286.610	285.494	277.471	275.427	260.298	83%
2006	945.958	325.697	298.671	297.565	289.436	287.044	270.555	83%
2007	975.036	338.213	309.141	307.963	299.911	297.283	279.606	83%
2008	1.017.065	356.465	324.655	323.274	314.707	312.897	293.851	82%
2009	1.048.131	370.407	337.298	336.057	326.968	325.071	304.701	82%
2010	1.069.725	387.209	355.685	354.200	344.170	342.141	319.637	83%
2011	1.117.753	407.954	370.941	369.742	359.682	357.494	332.106	81%
2012	1.123.571	418.772	381.915	378.883	369.597	361.194	335.513	80%
2013	1.159.963	431.255	392.713	389.105	379.690	370.615	343.261	80%
2014	1.169.534	442.388	403.490	399.561	389.829	380.117	348.807	79%
Média	1.008.982	360.602	328.683	326.865	317.904	314.123	293.781	82%

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

Tabela B.2 – Tratamento dos microdados de trabalhadores da RAIS: 2002-2014

Ano	Etapa I	Etapa II	Etapa III	Etapa IV	Etapa V	Etapa VI	Etapa VI / Etapa I
2002	4.906.071	4.905.567	4.348.157	4.068.473	4.006.188	3.084.074	63%
2003	4.928.566	4.928.550	4.398.757	4.119.755	4.069.359	3.177.436	64%
2004	5.168.965	5.168.880	5.132.718	4.802.355	4.742.646	3.392.980	66%
2005	5.426.661	5.426.543	5.389.195	5.098.209	5.045.015	3.662.322	67%
2006	5.721.804	5.721.711	5.687.505	5.407.753	5.348.338	3.923.472	69%
2007	6.160.103	6.160.044	6.124.251	5.848.793	5.776.752	4.273.453	69%
2008	6.540.251	6.540.249	6.489.372	6.167.609	6.114.571	4.498.928	69%
2009	6.722.364	6.722.316	6.665.181	6.329.870	6.276.712	4.608.334	69%
2010	7.166.407	7.166.407	7.121.752	6.757.293	6.703.083	4.980.651	69%
2011	7.380.405	7.380.387	6.906.980	6.576.029	6.519.231	5.170.584	70%
2012	7.608.839	7.608.832	7.539.645	7.208.200	7.048.143	5.265.088	69%
2013	7.684.602	7.684.530	7.153.138	6.831.030	6.675.167	5.318.544	69%
2014	7.744.894	7.744.879	7.155.690	6.842.290	6.688.164	5.344.069	69%
Média	6.396.918	6.396.838	6.162.488	5.850.589	5.770.259	4.361.533	68%

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

Tabela B.3 – Total de Emprego por subgrupos de CEP da RAIS: 2002-2014

Ano	MCU	MMC	MSPCM	RMSP	MCU/ RMSP	MMC/ RMSP	MSPCM/ RMSP
2002	67.855	1.477.295	3.360.921	4.906.071	1,4%	30,1%	68,5%
2003	63.974	1.502.921	3.361.671	4.928.566	1,3%	30,5%	68,2%
2004	63.447	1.605.882	3.499.636	5.168.965	1,2%	31,1%	67,7%
2005	66.268	1.675.794	3.684.599	5.426.661	1,2%	30,9%	67,9%
2006	66.278	1.750.425	3.905.101	5.721.804	1,2%	30,6%	68,2%
2007	70.780	1.883.152	4.206.171	6.160.103	1,1%	30,6%	68,3%
2008	70.168	1.981.007	4.489.076	6.540.251	1,1%	30,3%	68,6%
2009	69.730	2.031.549	4.621.085	6.722.364	1,0%	30,2%	68,7%
2010	74.666	2.218.402	4.873.339	7.166.407	1,0%	31,0%	68,0%
2011	80.896	2.274.626	5.024.883	7.380.405	1,1%	30,8%	68,1%
2012	78.696	2.292.885	5.237.258	7.608.839	1,0%	30,1%	68,8%
2013	77.108	2.359.590	5.247.904	7.684.602	1,0%	30,7%	68,3%
2014	74.473	2.362.020	5.308.401	7.744.894	1,0%	30,5%	68,5%
Média	71.103	1.955.042	4.370.773	6.396.918	1,1%	30,6%	68,3%

Fonte: Elaboração própria a partir dos microdados da RAIS (2002-2014). Nota: MCU – Municípios com CEP Único (Arujá, Biritiba Mirim, Caeiras, Cajamar, Embu Guaçu, Francisco Morato, Guararema, Juquitiba, Mairiporã, Pirapora do Bom Jesus, Rio Grande da Serra, Salesópolis, Santa Isabel, São Lourenço); MMC – Municípios com CEP Múltiplos sem o Município de São Paulo; MSPCM – Município de São Paulo com CEP Múltiplos. Esses resultados derivam da RAIS-trabalhador sem qualquer tratamento. Representa uma desagregação da coluna Etapa I da Tabela B.2.

APÊNDICE C – Distribuição Espacial dos Emprego, Firma e Massa Salarial em 2D

Figura C.1: Distribuição Espacial do Nível de Emprego em 2D (ano 2002)

Fonte: Elaboração própria a partir dos dados da RAIS (2002)

Figura C.3: Distribuição Espacial do Nível de Emprego em 2D (ano 2014)

Fonte: Elaboração própria a partir dos dados da RAIS (2014)

Figura C.2: Distribuição Espacial do Nível de Emprego em 2D (ano 2008)

Fonte: Elaboração própria a partir dos dados da RAIS (2008)

Figura C.4: Distribuição Espacial das Firmas em 2D (ano 2002)

Fonte: Elaboração própria a partir dos dados da RAIS (2002)

Figura C.5: Distribuição Espacial das Firms em 2D (ano 2008)

Fonte: Elaboração própria a partir dos dados da RAIS (2008)

Figura C.6: Distribuição Espacial Massa Salarial em 2D (ano 2002)

Fonte: Elaboração própria a partir dos dados da RAIS (2002)

Figura C.7: Distribuição Espacial das Firms em 2D (ano 2014)

Fonte: Elaboração própria a partir dos dados da RAIS (2014)

Figura C.8: Distribuição Espacial Massa Salarial em 2D (ano 2008)

Fonte: Elaboração própria a partir dos dados da RAIS (2008)

**Figura C.9: Distribuição Espacial Massa Salarial em 2D
(ano 2014)**

Fonte: Elaboração própria a partir dos dados da RAIS (2014)

**Figura C.11: Distribuição Espacial dos SBD em 2D por
Quantil (ano 2008)**

Fonte: Elaboração própria a partir dos dados da RAIS (2014)

**Figura C.10: Distribuição Espacial dos SBD em 2D por
Quantil (ano 2002)**

Fonte: Elaboração própria a partir dos dados da RAIS (2014)

**Figura C.12: Distribuição Espacial dos SBD em 2D por
Quantil (ano 2014)**

Fonte: Elaboração própria a partir dos dados da RAIS (2014)

APÊNDICE D – Estimções de Regressões Localmente Ponderadas com diferentes Kernel

Tabela D.1: Regressões Poderadas Geograficamente

	2002	2008	2014	2002	2008	2014
	I	II	III	IV	V	VI
Função Kernel	Gaussiana	Gaussiana	Gaussiana	Biquadrada	Biquadrada	Biquadrada
Min	9,9e ⁻¹¹	1,4e ⁻¹¹	1,2e ⁻¹⁰	0,00	0,00	0,00
1° Q	1,2	1,4	1,9	0,00	0,00	0,68
Mediana	13,82	18,30	19,12	1,57	2,55	14,57
3° Q	178,8	235,0	287,6	97,28	3.146,66	261,16
Max	11.332	18.464	2.159	21.477	34.345	24.788
Global	340	496	589	340	496	589
AIC	151.879	158.707	161.103	147.967	154.437	160.561
Quase-global R ²	0,55	0,57	0,59	0,75	0,77	0,62
Num. Obs.	9.071	9.071	9.071	9.071	9.071	9.071

Fonte: elaboração própria, a partir dos dados da RAIS (2002; 2008; 2014). Nota: os quartis se referem a distribuição dos coeficientes estimados pelo estimado GWR

APÊNDICE E – Características dos centros de negócios por valores de corte

Tabela E.1: Subcentros da Região Metropolitana de São Paulo – Quantil de ordem 99%

continua

Zona	Sigl	2002				2008				2014			
		Empreg	Firm	Massa Salarial	Grid	Empreg	Firm	Massa Salarial	Grid	Empreg	Massa Salarial	Firma	Grid
Leste	GR	42.470	2.334	42.970	30	88.362	4.392	142.203	55	131.097	5.855	324.281	58
Leste	IQQ	516	28	256	1	1.979	172	1.667	2	979	119	1.453	1
Leste	MJC	13.137	1.407	10.044	8	21.931	1.362	24.111	10	23.022	1.384	42.436	11
Leste	POA	1.588	129	836	4	5.420	135	3.830	3	11.439	628	17.592	4
Leste	SNO	12.959	1.321	12.100	10	15.701	1.493	16.806	9	26.303	1.772	41.655	9
Norte	CJM	6.309	124	3.478	6	-	-	-	-	-	-	-	-
Oeste	BAR	12.101	638	11.497	9	17.227	911	26.178	9	26.740	1.039	54.810	13
Oeste	CBA	4.173	390	2.807	3	12.740	684	10.001	6	7.062	653	12.298	4
Oeste	JAN	-	-	-	-	358	10	269	1	1.834	33	5.086	2
Oeste	OSA	14.705	979	14.148	13	19.640	1.440	29.687	17	15.102	998	35.527	10
Oeste	STP	11.613	682	9.932	16	30.904	1.425	45.926	12	23.640	783	59.829	11
Sudeste	DIA	1.284	75	973	1	750	36	1.216	1	13.551	468	35.317	4
Sudeste	MA	7.111	761	6.148	5	19.680	1.284	29.360	10	19.576	1.420	49.982	8
Sudeste	RPI	-	-	-	-	2.335	355	2.405	1	-	-	-	-
Sudeste	SAD	21.002	1.501	26.983	19	21.912	1.739	37.078	16	26.241	2.067	69.788	16
Sudeste	SCS	1.598	78	1.061	1	5.771	521	6.784	2	6.263	519	11.425	2
Sudeste	SBC	27.775	2.094	34.767	24	60.256	3.474	122.465	33	93.827	4.955	319.453	43
Sudoeste	COT	-	-	-	-	12.520	352	18.410	7	10.208	317	27.062	4
Sudoeste	EM	16.363	496	11.549	13	16.763	499	17.500	7	9.303	379	13.106	3
Sudoeste	JBA	4.498	80	3.508	1	-	-	-	-	-	-	-	-
Sudoeste	SLS	6.701	196	3.784	4	-	-	-	-	-	-	-	-
Sudoeste	TBS	8.579	503	9.909	9	14.267	703	21.947	9	15.233	808	39.017	9
Zona	BRE	229	26	134	1	-	-	-	-	-	-	-	-
Zona Leste	AR	7.971	952	5.857	5	13.537	1.280	15.229	5	6.909	742	13.390	3
Zona Leste	VPR	4.766	492	3.711	4	9.861	699	11.429	5	10.229	1.009	20.213	6
Zona Leste	TAT	-	-	-	-	-	-	-	-	1.038	44	1.688	1
Zona Leste	VFO	11.909	1.699	8.069	11	20.976	2.084	19.338	11	9.497	1.170	15.062	6
Zona Leste	SLU	5.951	753	4.655	6	6.747	752	8.439	5	9.679	1.377	15.303	6
Zona Leste	PEN	18.388	2.431	13.477	16	26.951	2.862	28.322	15	33.141	3.411	55.803	15
Zona Leste	ARI	15.104	1.489	10.525	10	18.869	1.642	19.781	10	26.138	2.270	44.960	11
Zona Leste	SAP	425	59	224	1	564	81	505	1	8.344	772	17.269	5
Zona Leste	CN	904	82	716	4	-	-	-	-	-	-	-	-

continuação

Zona	Sigl	2002				2008				2014			
		Empreg	Firm	Massa Salarial	Grid	Empreg	Firm	Massa Salarial	Grid	Empreg	Massa Salarial	Firma	Grid
Zona Leste	AAL	546	92	273	1	988	120	785	1	1.388	149	1.986	1
Zona Leste	VM	7.436	925	5.279	6	11.322	1.122	10.933	6	8.251	921	13.981	4
Zona Leste	CAR	4.186	548	3.178	2	5.474	653	6.639	2	2.947	338	6.570	1
Zona Leste	CLD	5.439	658	3.722	8	7.384	802	7.475	7	14.254	995	23.977	7
Zona Leste	SMT	6.497	496	5.472	4	11.412	732	14.471	4	16.043	1.256	31.862	6
Zona Leste	PQC	8.186	267	9.198	4	10.279	387	17.193	4	6.601	243	17.065	3
Zona Leste	ITQ	497	65	274	1	7.235	516	10.599	3	5.608	672	9.419	4
Zona Leste	JBO	-	-	-	-	-	-	-	-	746	44	1.537	1
Zona Leste	VJA	3.868	444	3.771	4	3.252	290	5.508	2	3.945	323	10.144	2
Zona Leste	SMI	7.294	754	7.205	6	11.343	917	13.341	4	10.667	1.117	17.537	4
Zona Leste	ER	1.468	77	1.251	2	-	-	-	-	-	-	-	-
Zona Norte	VGL	5.711	745	4.138	5	5.304	721	6.584	4	6.550	719	12.170	4
Zona Norte	STN	8.515	1.293	6.313	6	12.684	1.355	15.344	6	7.030	835	12.732	3
Zona Norte	MA	5.232	557	7.158	4	7.001	629	11.888	4	8.797	849	20.355	4
Zona Norte	TUC	13.114	1.621	10.379	10	14.689	1.758	15.247	9	22.695	2.419	37.089	10
Zona Norte	VM	16.528	1.427	16.041	13	9.036	1.119	9.940	7	11.086	1.531	17.756	8
Zona Norte	TRE	1.303	27	1.845	1	-	-	-	-	2.336	45	7.663	1
Zona Norte	JAC	6.870	561	5.531	5	2.975	128	4.358	1	3.606	167	7.874	1
Zona Norte	VM	2.544	188	2.492	1	3.523	220	4.761	1	10.378	777	19.835	3
Zona Norte	SDO	12.618	876	18.410	12	18.963	990	38.951	11	20.672	1.138	58.360	11
Zona Norte	FRE	10.112	1.380	7.594	9	15.391	1.775	15.776	10	21.253	2.276	35.363	11
Zona Norte	LIM	3.404	407	2.974	3	3.504	467	3.613	3	4.884	575	8.657	3
Zona Norte	CVE	3.514	453	2.246	3	3.526	541	3.619	3	4.309	651	6.855	3
Zona Norte	PIR	724	136	494	1	883	155	1.030	1	957	186	1.534	1
Zona Norte	CAC	1.109	186	849	1	1.421	201	1.455	1	1.703	260	2.650	1
Zona Oeste	VSO	18.675	2.030	18.339	16	27.970	1.953	53.837	14	18.389	1.699	37.154	10
Zona Oeste	RTA	7.180	262	11.236	5	1.119	55	1.358	1	-	-	-	-
Zona Oeste	RPE	6.721	578	14.020	5	8.837	706	18.826	4	12.482	895	61.330	6
Zona Oeste	JRE	4.649	320	5.349	4	4.244	302	5.905	2	5.707	323	13.896	3
Zona Oeste	VLE	1.648	66	4.314	1	1.472	82	2.366	1	-	-	-	-
Zona Oeste	BUT	6.903	224	14.952	4	-	-	-	-	12.396	466	80.440	4
Zona Oeste	API	3.233	314	5.023	3	1.358	104	4.044	1	5.164	505	16.612	3
Zona Oeste	JAG	3.210	248	3.086	3	1.717	153	2.932	2	1.781	192	4.807	2
Zona Sul 1	MO	943	23	1.619	1	1.767	50	4.891	2	802	33	4.701	1
Zona Sul 1	SAC	20.584	1.645	23.973	13	24.568	2.196	38.825	14	25.100	2.402	60.342	12
Zona Sul 1	CUR	5.640	569	5.409	7	6.914	560	9.348	4	8.321	738	16.987	5

													conclusão
Zona	Sigl	Empreg	Firm	2002 Massa Salarial	Grid	Empreg	Firm	2008 Massa Salarial	Grid	Empreg	2014 Massa Salarial	Firma	Grid
Zona Sul 1	JAB	14.194	1.498	11.707	10	17.630	1.558	20.275	9	14.335	1.523	27.471	8
Zona Sul 1	CBE	1.305	149	1.214	1	1.553	163	2.024	2	-	-	-	-
Zona Sul 2	CL	5.646	429	4.241	6	15.541	1.114	17.103	10	11.303	1.248	20.203	7
Zona Sul 2	VA	3.961	434	4.427	3	7.057	580	9.918	3	286	55	608	1
Zona Sul 2	JDS	2.499	279	2.582	4	5.749	544	6.867	6	5.014	463	11.533	5
Zona Sul 2	CA	12.124	1.342	10.110	9	14.961	1.526	18.075	9	17.395	1.674	29.461	8
Zona Sul 2	SA	69	9	75	1	110	16	200	1	100	18	315	1
Zona Sul 2	PDR	2.647	238	2.609	4	4.659	321	7.812	4	7.663	404	20.842	4
Zona Sul 2	CD	8.054	467	12.441	5	14.859	881	35.094	7	21.688	1.479	55.909	10
Zona Sul 2	SOC	945	113	963	2	1.420	147	1.918	2	1.322	198	2.759	2
Total		547.674	47.21	537.894	454	797.115	58.99	1.142.014	454	932.349	67.793	2.192.14	454
Média		7.502	647	7.368	6	11.722	868	16.794	7	13.711	997	32.237	7
Desvio Padrão		7.238	614	7.837	6	13.670	836	23.552	8	19.404	1.045	54.389	9
Mínimo		69	9	75	1	110	10	200	1	100	18	315	1
Máximo		42.470	2.431	42.970	30	88.362	4.392	142.203	55	131.097	5.855	324.281	58

Fonte: elaboração própria a partir dos microdados da RAIS.

Tabela E.2: Subcentros da Região Metropolitana de São Paulo – Quantil de ordem 95%**continua**

Região/Zon	Sigla	2002				2008				2014			
		Empreg	Firma	Massa Salarial	Grid	Empreg	Firma	Massa	Grid	Empreg	Firma	Massa Salarial	Grid
Leste	GRU	36.838	2.501	36.183	12	55.232	3.205	84.369	13	47.883	3.515	94.595	10
Leste	IQQ	1.407	103	907	1	-	-	-	-	-	-	-	-
Leste	MJC	3.947	610	2.219	1	10.264	1.175	9.471	2	18.187	1.868	30.319	3
Leste	POA	23.622	762	15.271	4	12.852	580	16.988	2	-	-	-	-
Leste	SNO	4.043	90	3.501	2	9.790	82	10.498	2	-	-	-	-
Norte	CJM	18.604	126	8.410	4	-	-	-	-	-	-	-	-
Oeste	BAR	41.243	981	41.266	11	60.675	1.250	113.232	13	49.055	1.107	136.987	10
Oeste	CBA	4.748	189	6.173	4	3.172	36	7.950	2	3.958	60	10.090	2
Oeste	OSA	19.609	1.880	20.923	5	34.126	2.533	50.853	6	50.107	3.338	110.456	9
Oeste	STP	18.032	948	15.110	7	2.998	144	5.390	4	10.712	280	41.883	3
Sudeste	DIA	34.596	1.718	36.753	12	50.077	2.182	79.965	13	37.333	2.012	90.685	9
Sudeste	SAD	74.481	5.242	72.739	22	104.030	5.997	157.042	22	128.801	6.966	285.179	24
Sudeste	SCS	9.378	643	7.533	3	41.568	1.858	80.378	6	51.176	2.612	140.227	7
Sudeste	SBC	110.017	5.673	145.122	36	166.309	7.216	363.859	40	136.459	7.176	353.959	30
Sudoeste	EMB	5.958	202	3.573	1	-	-	-	-	-	-	-	-
Zona Centro	CMB	29.954	1.652	39.660	6	38.295	1.829	58.018	6	38.662	2.014	93.659	6
Zona Centro	LIB	2.632	391	2.377	1	3.671	423	3.814	1	-	-	-	-
Zona Centro	PRI	-	-	-	-	7.829	771	7.500	1	7.259	784	12.621	1
Zona Centro	BRE	11.226	754	10.296	3	15.317	766	21.790	3	32.477	881	59.288	3
Zona Leste 1	MO	35.329	3.341	34.772	11	55.092	3.974	70.893	11	72.126	4.906	147.740	12
Zona Leste 1	ARA	20.541	2.309	14.742	7	22.901	2.692	25.852	7	41.031	3.939	76.675	9
Zona Leste 1	VPR	4.482	459	4.967	4	8.349	697	9.846	4	9.367	665	16.769	3
Zona Leste 1	TAT	46.590	3.979	53.655	15	64.719	5.103	93.614	15	81.089	6.214	157.223	14
Zona Leste 1	VFO	-	-	-	-	-	-	-	-	18.379	1.356	23.436	5
Zona Leste 1	BEL	10.989	714	12.925	2	11.907	812	19.787	2	16.311	847	32.953	2
Zona Leste 1	PEN	-	-	-	-	-	-	-	-	1.253	147	2.010	1
Zona Leste 1	VMT	-	-	-	-	-	-	-	-	6.754	489	13.306	2
Zona Leste 1	CAR	1.849	198	1.516	1	2.881	276	3.335	1	7.820	765	14.654	2
Zona Leste 2	PQC	-	-	-	-	-	-	-	-	6.819	245	9.888	1
Zona Leste 2	ITQ	-	-	-	-	-	-	-	-	6.968	143	21.129	1
Zona Leste 2	ERM	1.989	25	2.399	1	-	-	-	-	-	-	-	-
Zona Norte 1	VGL	34.322	2.866	38.714	12	47.347	3.064	75.344	12	68.436	3.739	164.035	13
Zona Norte 1	STN	30.591	2.846	28.172	11	39.695	3.415	52.733	10	63.509	4.938	128.399	13
Zona Norte 1	VM	9.142	765	9.072	4	12.429	871	16.832	4	7.511	427	21.872	2
Zona Norte 2	SDO	2.114	154	2.584	1	2.212	186	3.846	1	6.766	292	28.214	2

													conclusão
Região/Zon	Sigla	2002				2008				2014			
		Empreg	Firma	Massa Salarial	Grid	Empreg	Firma	Massa	Grid	Empreg	Firma	Massa Salarial	Grid
Zona Norte 2	FRE	6.624	599	5.769	3	12.189	913	15.041	4	14.258	1.027	30.259	4
Zona Norte 2	LIM	4.889	437	4.146	2	6.504	515	7.290	2	7.522	616	14.082	2
Zona Norte 2	CVE	3.462	394	2.903	1	4.672	480	5.573	1	5.512	555	10.460	1
Zona Oeste	VSO	1.873	168	1.365	1	9.981	1.075	12.938	4	38.074	2.024	121.904	8
Zona Oeste	RPE	12.428	700	19.946	5	20.345	900	56.715	6	15.922	959	61.726	4
Zona Oeste	JRE	14.231	540	25.213	8	33.137	747	83.084	10	29.754	922	101.887	9
Zona Oeste	VLE	28.442	2.019	31.789	9	43.810	2.455	74.201	9	56.219	3.034	175.861	10
Zona Oeste	IBI	7.038	457	14.188	1	9.195	505	14.776	1	14.964	513	52.796	1
Zona Oeste	PIN	9.369	1.469	10.023	4	13.967	1.785	24.503	4	19.985	2.123	59.162	4
Zona Oeste	PRD	19.047	2.273	19.796	6	19.198	1.803	25.296	4	40.094	3.731	95.126	6
Zona Oeste	BUT	6.014	405	8.479	4	31.625	877	86.181	8	16.114	561	51.483	4
Zona Oeste	API	1.783	215	1.861	1	4.894	485	8.190	3	2.136	230	5.594	1
Zona Oeste	LAP	46.767	2.894	56.753	9	38.134	1.945	52.842	5	39.359	2.295	92.358	5
Zona Oeste	BFU	17.385	568	25.779	4	15.347	567	22.499	3	13.825	661	39.789	3
Zona Oeste	JAG	-	-	-	-	2.438	134	3.913	1	2.855	132	6.207	1
Zona Sul 1	MO	16.904	944	26.117	9	18.697	1.122	30.060	9	22.447	1.288	63.016	8
Zona Sul 1	SAC	21.330	1.485	39.576	11	10.616	786	14.874	4	21.257	1.332	39.253	6
Zona Sul 1	CUR	23.330	2.374	26.245	9	26.112	2.490	33.646	9	30.291	2.754	63.066	8
Zona Sul 1	JAB	37.781	2.610	68.084	10	62.189	3.181	172.370	10	58.786	3.973	158.983	11
Zona Sul 1	CBE	22.446	2.828	22.471	10	30.938	2.757	63.608	8	58.069	5.225	172.577	11
Zona Sul 1	SAU	18.808	2.253	21.390	5	30.751	3.035	45.791	6	24.225	1.963	60.550	4
Zona Sul 1	VM	12.927	1.509	12.270	5	16.799	1.703	22.910	5	29.931	1.953	59.885	5
Zona Sul 1	IPI	32.749	2.014	35.398	9	49.714	2.839	74.566	10	58.234	3.405	142.928	10
Zona Sul 1	MOE	4.792	34	5.779	1	5.796	55	11.659	1	12.970	47	38.508	1
Zona Sul 2	CLM	8.834	584	6.965	2	11.249	762	15.170	2	15.802	890	26.758	2
Zona Sul 2	VAN	33.223	1.055	99.945	7	25.471	876	76.587	6	48.964	1.928	233.237	8
Zona Sul 2	JDS	10.633	294	20.977	3	14.837	375	46.341	3	19.640	663	99.149	4
Zona Sul 2	CAD	1.674	176	1.439	1	4.070	344	5.543	2	4.896	406	10.988	2
Zona Sul 2	SAM	43.047	2.314	61.329	9	56.934	3.086	95.011	10	62.588	4.248	177.552	11
Zona Sul 2	CGR	28.899	1.365	44.463	8	46.732	1.660	106.829	8	57.621	1.917	170.389	8
Zona Sul 2	SOC	4.072	305	4.842	2	6.684	373	11.255	2	7.260	468	17.293	2
Total		1.149.07	77.40	1.396.829	363	1.566.76	91.76	2.762.45	363	1.875.78	113.56	4.771.076	363
Média		19.476	1.312	23.675	6	27.487	1.610	48.464	6	31.793	1.925	80.866	6
Desvio Padrão		19.322	1.270	26.354	6	28.331	1.506	57.988	6	28.590	1.804	73.435	5
Mínimo		1.407	25	907	1	2.212	36	3.335	1	1.253	47	2.010	1
Máximo		110.017	5.673	145.122	36	166.309	7.216	363.859	40	136.459	7.176	353.959	30

Fonte: elaboração própria a partir dos microdados da RAIS.

Tabela E.3: Subcentros da Região Metropolitana de São Paulo – Quantil de ordem 90%

Zona	Sigla	2002				2008				2014			
		Emprego	Firma	Massa	Cel.	Emprego	Firma	Massa	Cel.	Emprego	Firma	Massa	Cel.
Oeste	BAR	55.342	1.566	71.243	5	62.087	1.914	150.549	4	94.556	2.476	377.924	5
Sudeste	SCS	56.533	1.795	62.021	5	13.845	533	15.234	1	-	-	-	-
Zona Centro	CMB	4.103	594	3.840	1	4.551	632	5.406	1	6.127	648	12.418	1
Zona Centro	BVI	49.658	4.296	53.642	3	83.466	4.599	110.014	3	108.218	4.966	232.377	3
Zona Centro	LIB	24.918	1.714	31.958	3	36.596	1.941	66.136	3	49.968	2.684	148.897	4
Zona Centro	SEE	73.116	6.635	93.367	3	116.154	7.525	184.699	3	117.292	8.217	300.838	3
Zona Centro	PRI	39.990	4.442	27.272	5	46.688	4.840	49.189	4	52.003	5.528	124.025	4
Zona Centro	BRS	6.341	650	5.564	1	10.857	713	13.437	1	8.412	748	20.956	1
Zona Centro	CON	41.715	3.857	53.550	3	52.736	4.218	90.747	3	62.291	4.561	176.679	3
Zona Centro	SCE	15.707	2.023	11.010	1	19.332	2.310	20.559	1	21.236	2.412	39.616	1
Zona Centro	REP	17.235	2.165	16.399	1	20.910	2.531	24.911	1	24.976	2.450	55.021	1
Zona Centro	BRE	10.729	543	14.071	1	17.372	596	29.056	1	19.306	650	53.590	1
Zona Leste 1	MOO	3.966	303	4.540	1	7.661	350	7.244	1	-	-	-	-
Zona Leste 1	VPR	3.467	181	3.036	1	-	-	-	-	-	-	-	-
Zona Norte 1	STN	-	-	-	-	10.374	256	14.116	1	13.676	268	30.607	1
Zona Oeste	IBI	85.931	7.778	164.105	9	141.431	9.982	446.111	9	201.233	11.646	973.748	9
Zona Oeste	JDP	139.220	9.991	242.061	11	228.516	11.690	534.020	11	252.890	12.985	895.791	11
Zona Oeste	PIN	24.504	2.169	31.033	3	40.371	2.564	69.348	3	38.038	2.806	162.538	3
Zona Oeste	PRD	38.796	2.949	50.666	5	71.754	4.236	126.969	7	50.783	2.879	166.097	5
Zona Oeste	LAP	-	-	-	-	51.436	1.367	89.338	4	62.820	1.488	148.082	4
Zona Oeste	BFU	11.668	840	14.477	2	42.151	1.135	69.953	3	44.769	1.358	120.655	3
Zona Sul 1	MOR	66.590	3.464	176.106	7	109.799	4.478	436.207	7	167.186	5.475	1.017.398	9
Zona Sul 1	CBE	19.181	1.926	20.685	2	32.300	2.696	43.477	3	12.129	1.028	24.531	2
Zona Sul 1	SAU	24.710	2.026	30.047	3	36.276	1.801	65.939	2	68.016	3.458	217.647	4
Zona Sul 1	VMN	78.624	6.614	138.276	7	127.802	7.518	295.636	7	142.538	8.400	513.254	7
Zona Sul 1	IPI	9.390	448	8.185	1	-	-	-	-	-	-	-	-
Zona Sul 1	MOE	12.387	1.242	15.800	3	17.037	1.456	25.741	3	17.493	1.507	50.825	3
Zona Sul 2	VAN	-	-	-	-	18.917	436	105.861	1	26.225	543	204.852	1
Zona Sul 2	SAM	21.022	1.333	29.454	4	26.212	1.202	79.305	3	22.004	781	111.228	2
Total		934.843	71.544	1.372.410	91	1.446.631	83.519	3.169.201	91	1.684.185	89.962	6.179.594	91
Média		35.956	2.752	52.785	4	53.579	3.093	117.378	3	67.367	3.598	247.184	4
Desvio Padrão		32.526	2.532	61.736	3	51.969	3.015	143.389	3	64.781	3.462	294.543	3
Mínimo		3.467	181	3.036	1	4.551	256	5.406	1	6.127	268	12.418	1
Máximo		139.220	9.991	242.061	11	228.516	11.690	534.020	11	252.890	12.985	1.017.398	11

Fonte: elaboração própria a partir dos microdados da RAIS.

Tabela E.4: Participação do trabalhador por gênero nos SBD por quantil

Estatística	2002		2008		2014	
	Homem	Mulher	Homem	Mulher	Homem	Mulher
	Quantil 99					
Média	55,2%	44,8%	47,5%	52,5%	50,3%	49,7%
Desvio Padrão	6,5%	6,5%	6,4%	6,4%	5,7%	5,7%
Mínimo	39,5%	31,9%	31,1%	36,8%	37,8%	35,7%
Máximo	68,1%	60,5%	63,2%	68,9%	64,3%	62,2%
	Quantil 95					
Média	60,3%	39,7%	42,5%	57,5%	51,9%	48,1%
Desvio Padrão	7,7%	7,7%	7,0%	7,0%	8,6%	8,6%
Mínimo	34,2%	18,7%	28,9%	36,9%	22,4%	24,4%
Máximo	81,3%	65,8%	63,1%	71,1%	75,6%	77,6%
	Quantil 90					
Média	61,3%	38,7%	41,8%	58,2%	54,1%	45,9%
Desvio Padrão	9,2%	9,2%	9,5%	9,5%	8,6%	8,6%
Mínimo	35,5%	11,2%	8,9%	34,7%	29,2%	24,0%
Máximo	88,8%	64,5%	65,3%	91,1%	76,0%	70,8%

Fonte: elaboração própria a partir dos microdados da RAIS.

Tabela E.5: Parcela de Trabalhadores por Categorias de Salário Mínimo nos SBD

continua

Tabela E15: Parcela de Trabalhadores por Categorias de Salário Mínimo nos SDE									
Estatística	SM 1	SM 2	SM 3	SM 4	SM 5	SM 6	SM 7	SM 8	SM 9
	2002								
	Quartil 99%								
Média	2,7%	12,3%	29,1%	14,4%	8,6%	10,6%	8,2%	6,0%	8,2%
Desvio Padrão	0,9%	6,3%	6,9%	2,0%	1,6%	3,0%	2,6%	2,4%	5,6%
Mínimo	1,6%	5,2%	15,9%	9,0%	6,4%	5,2%	2,7%	1,6%	1,3%
Máximo	6,1%	29,1%	46,5%	18,4%	11,9%	18,2%	12,5%	11,2%	27,4%
	Quartil 95%								
Média	3,1%	14,4%	28,2%	14,6%	9,0%	10,7%	7,6%	5,5%	6,8%
Desvio Padrão	1,1%	9,1%	6,6%	2,7%	2,2%	3,0%	2,4%	2,7%	5,9%
Mínimo	1,8%	1,2%	10,4%	8,1%	3,4%	2,1%	1,0%	0,6%	0,4%
Máximo	8,2%	55,2%	43,4%	23,5%	16,2%	18,5%	12,9%	11,6%	32,9%
	Quartil 90%								
Média	3,2%	17,5%	32,2%	14,2%	8,2%	9,5%	6,1%	4,4%	4,7%
Desvio Padrão	1,7%	9,7%	8,6%	2,8%	2,6%	4,2%	3,3%	3,4%	5,0%
Mínimo	1,4%	3,2%	6,2%	6,9%	2,4%	3,8%	0,7%	0,0%	0,0%
Máximo	13,0%	51,7%	59,4%	24,2%	23,3%	33,5%	24,2%	20,9%	23,9%
	2008								

Estatística									conclusão
	SM 1	SM 2	SM 3	SM 4	SM 5	SM 6	SM 7	SM 8	SM 9
Quartil 99%									
Média	4,8%	36,2%	21,8%	9,9%	5,8%	6,6%	5,3%	4,1%	5,5%
Desvio Padrão	1,5%	11,7%	5,8%	2,1%	1,5%	2,0%	2,5%	3,1%	6,6%
Mínimo	2,9%	9,3%	8,2%	5,2%	2,8%	2,3%	1,8%	0,8%	0,3%
Máximo	8,2%	59,5%	35,7%	14,0%	8,5%	10,5%	12,3%	14,2%	29,5%
Quartil 95%									
Média	5,0%	35,6%	23,4%	10,7%	6,3%	6,8%	5,0%	3,4%	3,7%
Desvio Padrão	1,9%	9,8%	4,4%	2,1%	1,6%	2,0%	2,0%	2,0%	3,6%
Mínimo	1,5%	6,8%	14,2%	6,4%	2,6%	2,8%	1,5%	0,6%	0,2%
Máximo	12,4%	58,2%	35,2%	18,2%	14,4%	14,0%	10,8%	9,0%	16,1%
Quartil 90%									
Média	5,0%	42,4%	23,4%	10,2%	5,5%	5,6%	3,5%	2,3%	2,2%
Desvio Padrão	1,9%	11,1%	4,1%	3,0%	2,0%	2,4%	1,9%	1,9%	2,5%
Mínimo	2,0%	13,8%	13,0%	3,6%	0,6%	0,8%	0,2%	0,0%	0,0%
Máximo	12,2%	69,9%	34,8%	24,5%	12,2%	16,0%	10,6%	10,6%	13,0%
2014									
Quartil 99%									
Média	5,4%	35,5%	20,0%	10,1%	6,2%	7,3%	5,6%	4,7%	5,2%
Desvio Padrão	1,3%	9,3%	4,6%	1,7%	1,1%	1,9%	2,4%	3,2%	5,2%
Mínimo	3,1%	14,6%	8,7%	7,1%	4,3%	4,2%	2,0%	1,0%	0,6%
Máximo	7,9%	50,7%	31,4%	13,7%	8,5%	10,7%	12,1%	14,8%	22,2%
Quartil 95%									
Média	5,9%	40,6%	22,8%	9,6%	5,5%	5,9%	4,0%	2,9%	2,8%
Desvio Padrão	2,1%	8,5%	4,9%	1,8%	1,3%	1,6%	1,9%	2,1%	2,8%
Mínimo	3,0%	24,1%	13,7%	4,9%	2,4%	1,7%	0,8%	0,4%	0,1%
Máximo	15,4%	60,9%	36,4%	14,1%	8,6%	9,6%	8,8%	9,9%	13,1%
Quartil 90%									
Média	5,5%	44,5%	22,6%	9,9%	5,1%	5,0%	3,0%	2,3%	2,1%
Desvio Padrão	2,1%	11,5%	4,0%	3,3%	1,9%	2,1%	2,0%	3,0%	3,5%
Mínimo	1,0%	11,0%	8,9%	4,8%	2,2%	1,4%	0,5%	0,1%	0,0%
Máximo	13,9%	61,4%	35,0%	23,7%	15,0%	10,8%	11,0%	20,1%	18,8%

Fonte: Elaboração própria a partir dos dados da RAIS. Nota: SM1 (menor que 1 SM); SM2 (entre 1 + 2 SM); SM3 (entre 2 + 3 SM); SM4 (entre 3 + 4 SM); SM5 (entre 4 + 5 SM); SM6 (entre 5 + 7 SM); SM7 (entre 7 + 10 SM); SM8 (entre 10 + 15); SM9 (maior que 15 SM).

Tabela E.6: Parcela de Trabalhadores por Anos de Experiência nos SBD**continua**

Estatística	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7	Exp8	Exp9
2002									
Quartil 99%									
Média	32,7%	18,4%	11,9%	7,2%	5,4%	4,6%	3,7%	3,3%	12,9%
Desvio Padrão	5,7%	2,1%	1,2%	1,0%	0,6%	0,8%	0,9%	0,9%	3,7%
Mínimo	24,6%	14,9%	9,1%	5,0%	3,9%	2,9%	1,9%	1,6%	5,9%
Máximo	44,4%	23,0%	14,3%	10,0%	7,0%	5,8%	5,5%	5,5%	19,7%
Quartil 95%									
Média	33,3%	17,9%	11,9%	7,0%	5,1%	4,5%	3,4%	3,2%	13,6%
Desvio Padrão	8,0%	2,6%	1,6%	1,1%	0,9%	1,1%	0,8%	1,0%	8,2%
Mínimo	12,7%	9,8%	6,5%	3,5%	1,8%	1,8%	1,1%	0,6%	1,5%
Máximo	53,9%	24,4%	15,8%	10,1%	7,0%	8,1%	5,6%	5,6%	60,1%
Quartil 90%									
Média	33,5%	18,6%	12,6%	7,3%	5,5%	4,6%	3,4%	3,0%	11,4%
Desvio Padrão	8,6%	2,7%	3,8%	2,0%	1,7%	1,2%	1,6%	1,4%	7,2%
Mínimo	15,0%	9,3%	4,3%	1,7%	0,6%	0,7%	0,2%	0,2%	0,2%
Máximo	76,6%	24,1%	36,9%	16,3%	11,9%	8,9%	13,9%	12,0%	44,7%
2008									
Quartil 99%									
Média	40,3%	17,4%	10,2%	7,1%	4,4%	3,1%	2,8%	2,4%	12,1%
Desvio Padrão	5,2%	1,9%	1,4%	1,5%	1,0%	0,5%	0,7%	0,5%	4,1%
Mínimo	30,9%	12,5%	7,7%	5,5%	3,2%	2,5%	2,0%	1,6%	7,0%
Máximo	49,7%	20,3%	13,8%	11,0%	8,1%	4,0%	4,9%	3,3%	25,2%
Quartil 95%									
Média	38,0%	17,7%	9,8%	6,8%	4,9%	3,6%	3,0%	2,6%	13,6%
Desvio Padrão	6,8%	3,5%	1,2%	1,2%	1,1%	1,0%	0,9%	0,8%	6,0%
Mínimo	8,8%	3,8%	4,9%	3,0%	2,7%	1,8%	1,3%	1,1%	5,4%
Máximo	54,6%	35,6%	12,4%	10,1%	8,7%	9,1%	7,0%	5,5%	42,7%
Quartil 90%									
Média	37,7%	18,1%	10,6%	7,0%	5,1%	3,9%	3,1%	2,6%	11,9%
Desvio Padrão	7,4%	2,4%	1,9%	1,3%	0,9%	1,7%	1,1%	0,9%	4,8%
Mínimo	16,4%	10,6%	2,8%	2,5%	1,4%	0,3%	0,6%	1,1%	3,6%
Máximo	77,1%	25,3%	14,9%	9,9%	7,3%	15,3%	8,1%	5,5%	31,8%
2014									
Quartil 99%									
Média	35,4%	17,1%	11,1%	7,9%	5,6%	3,9%	3,2%	2,4%	13,3%
Desvio Padrão	4,6%	1,4%	1,1%	1,0%	0,5%	0,6%	0,4%	0,4%	3,6%
Mínimo	26,9%	14,0%	9,6%	6,6%	4,8%	3,1%	2,5%	1,8%	7,7%

									conclusão
	Exp1	Exp2	Exp3	Exp4	Exp5	Exp6	Exp7	Exp8	Exp9
Máximo	43,4%	20,2%	14,7%	11,1%	6,9%	5,5%	4,3%	3,4%	21,1%
	Quartil 95%								
Média	36,8%	17,2%	11,3%	7,6%	5,6%	3,7%	3,1%	2,3%	12,2%
Desvio Padrão	5,7%	2,2%	3,2%	0,9%	0,8%	0,7%	0,9%	0,6%	4,6%
Mínimo	20,6%	10,7%	8,1%	5,3%	2,9%	2,1%	1,7%	1,4%	1,6%
Máximo	51,3%	26,6%	33,8%	10,2%	8,0%	5,9%	8,1%	5,5%	32,0%
	Quartil 90%								
Média	33,5%	17,4%	11,2%	8,4%	6,0%	4,0%	3,4%	2,5%	13,6%
Desvio Padrão	6,0%	2,4%	1,6%	2,9%	1,1%	1,4%	0,7%	0,8%	6,9%
Mínimo	10,0%	8,0%	7,7%	6,1%	3,7%	1,9%	1,9%	1,3%	5,7%
Máximo	49,1%	21,4%	17,6%	27,2%	10,6%	11,8%	5,2%	5,1%	47,0%

Fonte: Elaboração própria a partir dos dados da RAIS. Nota: Exp1 (com menos de 1 ano no emprego); Exp2 (1 ─ 2 anos); Exp3 (2 ─ 3 anos); Exp4 (3 ─ 4 anos); Exp5 (4 ─ 5 anos); Exp6 (5 ─ 6 anos); Exp7 (6 ─ 7 anos).

Tabela E.7: Parcela de Trabalhadores por Categorias de Instrução Educacional nos SBD

continua

Estatística	Grau 1	Grau 2	Grau 3	Grau 4	Grau 5	Grau 6	Grau 7	Grau 8	Grau 9
	2002								
	Quartil 99%								
Média	0,7%	3,8%	8,0%	10,8%	16,8%	9,4%	28,6%	7,2%	14,6%
Desvio Padrão	0,2%	1,2%	3,3%	4,1%	5,9%	1,7%	5,1%	3,9%	7,3%
Mínimo	0,5%	2,0%	3,1%	4,5%	8,4%	6,8%	19,3%	1,8%	3,9%
Máximo	1,2%	6,2%	19,9%	20,2%	34,9%	13,1%	36,7%	21,3%	34,5%
	Quartil 95%								
Média	1,0%	4,3%	8,8%	12,0%	18,0%	10,5%	28,4%	5,4%	11,6%
Desvio Padrão	0,7%	2,0%	3,9%	3,2%	3,7%	2,2%	5,2%	2,2%	7,0%
Mínimo	0,1%	0,8%	3,3%	5,0%	8,1%	5,1%	11,5%	0,7%	0,7%
Máximo	3,5%	11,9%	31,1%	20,9%	29,3%	17,7%	42,2%	13,0%	38,8%
	Quartil 90%								
Média	0,9%	4,1%	9,8%	13,1%	20,8%	11,1%	26,4%	4,0%	9,8%
Desvio Padrão	0,9%	2,5%	4,6%	3,3%	5,9%	2,6%	6,6%	2,3%	7,5%
Mínimo	0,0%	0,9%	3,3%	5,3%	9,8%	4,3%	12,6%	0,0%	1,1%
Máximo	5,4%	17,3%	34,3%	22,3%	41,0%	19,3%	54,4%	16,8%	37,6%
	2008								
	Quartil 99%								
Média	0,2%	2,0%	3,8%	6,3%	11,7%	7,1%	42,0%	7,3%	19,6%

Estatística									conclusão
	Grau 1	Grau 2	Grau 3	Grau 4	Grau 5	Grau 6	Grau 7	Grau 8	Grau 9
Desvio Padrão	0,1%	1,9%	2,1%	2,7%	4,9%	2,5%	8,8%	2,9%	13,3%
Mínimo	0,0%	0,2%	0,7%	1,2%	3,7%	2,1%	15,6%	2,1%	4,5%
Máximo	0,6%	10,5%	11,0%	14,8%	22,5%	14,2%	57,4%	13,9%	64,4%
Quartil 95%									
Média	0,3%	2,1%	4,2%	7,3%	13,5%	8,6%	44,0%	5,6%	14,2%
Desvio Padrão	0,4%	0,8%	1,5%	2,1%	3,6%	2,2%	7,5%	1,9%	7,2%
Mínimo	0,0%	0,6%	0,5%	1,4%	1,5%	1,5%	30,9%	2,1%	3,7%
Máximo	3,0%	4,7%	9,3%	11,9%	21,5%	14,7%	72,6%	11,7%	37,1%
Quartil 90%									
Média	0,3%	2,3%	4,7%	8,1%	16,4%	9,8%	43,5%	4,3%	10,8%
Desvio Padrão	0,2%	1,8%	2,1%	3,4%	4,8%	2,1%	7,0%	1,8%	6,1%
Mínimo	0,0%	0,0%	1,6%	3,0%	8,4%	3,6%	28,7%	0,6%	0,6%
Máximo	0,9%	12,1%	15,8%	26,0%	30,4%	14,1%	61,0%	11,6%	36,7%
2014									
Quartil 99%									
Média	0,1%	1,3%	2,1%	4,0%	7,8%	6,3%	44,7%	6,0%	27,8%
Desvio Padrão	0,1%	0,7%	1,0%	1,7%	2,8%	2,1%	9,2%	1,6%	13,6%
Mínimo	0,0%	0,5%	0,6%	1,2%	2,7%	1,9%	17,5%	3,3%	9,8%
Máximo	0,3%	3,5%	4,9%	7,2%	12,1%	11,4%	57,3%	9,2%	66,1%
Quartil 95%									
Média	0,1%	1,4%	2,8%	4,7%	9,8%	7,1%	50,3%	4,7%	19,1%
Desvio Padrão	0,1%	0,6%	4,7%	2,6%	3,2%	1,9%	9,3%	1,2%	9,9%
Mínimo	0,0%	0,1%	0,3%	0,5%	0,8%	0,4%	25,9%	1,6%	4,2%
Máximo	0,5%	3,4%	37,3%	16,5%	22,7%	11,9%	80,9%	9,3%	64,8%
Quartil 90%									
Média	0,1%	1,5%	2,6%	4,6%	11,1%	8,9%	52,6%	3,5%	15,1%
Desvio Padrão	0,1%	1,3%	1,4%	2,0%	3,5%	2,3%	7,9%	1,1%	8,9%
Mínimo	0,0%	0,0%	0,3%	1,3%	5,0%	3,7%	21,2%	1,0%	3,4%
Máximo	0,6%	8,0%	8,6%	16,7%	21,0%	15,6%	69,7%	7,0%	53,6%

Fonte: Elaboração própria a partir dos dados da RAIS (2002). Nota: G1 (analfabetos); G2 (até o 5º ano do fundamental incompleto); G3 (5º ano do fundamental completo); G4 (6º a 9º ano do ensino fundamental incompleto); G5 (fundamental completo); G6 (médio incompleto); G7 (médio completo); G8 (superior incompleto); G9 (superior completo, mestrado, doutorado).

Tabela E.8: Parcela de Trabalhadores por Atividade Setorial nos SBD**continua**

Estatística	S1	S4	S6	S7	S9	S14	S16	S19
2002								
Quartil 99%								
Média	24,9%	17,3%	-	16,5%	-	-	6,6%	7,2%
Desvio Padrão	12,3%	7,0%	-	7,1%	-	-	8,4%	9,1%
Mínimo	2,6%	5,3%	-	4,4%	-	-	0,4%	0,1%
Máximo	52,0%	31,4%	-	33,7%	-	-	38,7%	34,2%
Quartil 95%								
Média	20,0%	12,4%	5,1%	18,2%	-	6,0%	-	-
Desvio Padrão	13,2%	10,8%	3,7%	10,2%	-	6,9%	-	-
Mínimo	0,0%	0,0%	0,1%	0,6%	-	0,0%	-	-
Máximo	60,7%	73,5%	22,0%	55,2%	-	30,4%	-	-
Quartil 90%								
Média	16,3%	11,3%	-	22,3%	-	6,0%	-	5,4%
Desvio Padrão	19,5%	13,6%	-	11,4%	-	6,0%	-	13,7%
Mínimo	0,0%	0,9%	-	1,7%	-	0,0%	-	0,0%
Máximo	92,6%	89,9%	-	48,9%	-	26,4%	-	93,3%
2008								
Quartil 99%								
Média	33,2%	16,0%	6,2%	13,1%	-	-	-	6,1%
Desvio Padrão	13,7%	10,5%	4,4%	6,2%	-	-	-	9,7%
Mínimo	5,0%	4,7%	1,5%	3,9%	-	-	-	0,0%
Máximo	58,7%	59,8%	19,7%	29,4%	-	-	-	40,8%
Quartil 95%								
Média	21,5%	12,5%	6,3%	18,5%	-	5,4%	-	-
Desvio Padrão	11,0%	6,6%	4,7%	8,0%	-	7,6%	-	-
Mínimo	1,1%	0,3%	0,6%	0,9%	-	0,0%	-	-
Máximo	55,7%	31,4%	22,6%	36,6%	-	47,8%	-	-
Quartil 90%								
Média	12,9%	14,1%	5,4%	23,5%	-	5,5%	-	-
Desvio Padrão	11,3%	12,9%	3,5%	11,0%	-	5,5%	-	-
Mínimo	0,3%	0,0%	0,0%	2,0%	-	0,0%	-	-
Máximo	62,6%	55,0%	14,3%	49,3%	-	28,2%	-	-
2014								
Quartil 99%								
Média	33,9%	15,4%	-	12,7%	-	-	6,3%	7,8%
Desvio Padrão	15,5%	6,8%	-	5,3%	-	-	6,6%	10,4%
Mínimo	4,0%	4,8%	-	4,0%	-	-	0,5%	0,0%

								conclusão
Estatística	S1	S4	S6	S7	S9	S14	S16	S19
Máximo	59,2%	29,9%	-	23,4%	-	-	30,4%	35,6%
Quartil 95%								
Média	25,3%	13,8%	7,0%	18,3%	-	-	-	7,2%
Desvio Padrão	12,8%	8,4%	7,7%	7,1%	-	-	-	12,8%
Mínimo	0,5%	1,0%	0,0%	1,5%	-	-	-	0,1%
Máximo	62,8%	46,5%	42,1%	37,8%	-	-	-	82,8%
Quartil 90%								
Média	15,6%	13,9%	-	24,9%	7,5%	-	-	6,2%
Desvio Padrão	10,1%	10,9%	-	11,7%	12,7%	-	-	14,0%
Mínimo	1,5%	1,3%	-	2,0%	0,0%	-	-	0,0%
Máximo	50,2%	49,0%	-	52,7%	88,8%	-	-	85,6%

Fonte: Elaboração própria a partir dos dados da RAIS. Nota: S1: Comercialização e administração de imóveis, valores mobiliários, e serviços técnicos; S4: Indústria de produtos alimentícios, bebidas e álcool etílico; S6: Comércio atacadista; S7: Comércio varejista; S9: Ensino; S14: Indústria Química; S16: Instituições de crédito, seguros e capitalização; S19: Serviços médicos, odontológicos e veterinários.

Tabela E.9: Teste de Sensibilidade à Área da Célula

	2002			2008			2014		
	Quantil 99%								
	SBD_{2Km^2}								
SBD_{1Km^2}	0	1	Total	0	1	Total	0	1	Total
0	8.955	25	8.980	8.955	25	8.980	8.951	29	8.980
1	21	70	91	18	73	91	16	75	91
Total	8.976	95	9.071	8.973	98	9.071	8.967	104	9.071
	SBD_{4Km^2}								
SBD_{1Km^2}	0	1	Total	0	1	Total	0	1	Total
0	8.944	36	8.980	8.952	28	8.980	8.945	35	8.980
1	31	60	91	23	68	91	30	61	91
Total	8.975	96	9.071	8.975	96	9.071	8.975	96	9.071
	Quantil 95%								
	SBD_{2Km^2}								
SBD_{1Km^2}	0	1	Total	0	1	Total	0	1	Total
0	8.558	59	8.617	8.952	64	8.617	8.546	71	8.617
1	61	393	454	23	397	454	68	386	454
Total	8.619	452	9.071	8.975	461	9.071	8.614	457	9.071
	SBD_{4Km^2}								
SBD_{1Km^2}	0	1	Total	0	1	Total	0	1	Total
0	8.516	101	8.617	8.522	95	8.617	8.521	96	8.617
1	83	371	454	77	377	454	78	376	454
Total	8.599	472	9.071	8.599	472	9.071	8.599	472	9.071
	Quantil 90%								
	SBD_{2Km^2}								
SBD_{1Km^2}	0	1	Total	0	1	Total	0	1	Total
0	8.064	99	8.163	8.072	91	8.163	8.059	104	8.163
1	87	821	908	85	823	908	94	814	908
Total	8.151	920	9.071	8.157	914	9.071	8.153	918	9.071
	SBD_{4Km^2}								
SBD_{1Km^2}	0	1	Total	0	1	Total	0	1	Total
0	7.993	170	8.163	7.999	164	8.163	8.003	160	8.163
1	138	770	908	132	776	908	128	780	908
Total	8.131	940	9.071	8.131	940	9.071	8.131	940	9.071

Fonte: elaboração própria a partir dos microdados da RAIS.

APÊNDICE F – Tratamento dos Microdados da RAIS para Base Longitudinal

A Tabela F.1 mostra o efeito das exclusões de observações sobre o total de trabalhadores nas bases anuais da RAIS (2002-2014) após o tratamento aplicado para georreferenciamento (descrito na Seção 3.2 dessa tese) considerando cada estratégia de limpeza da base como Etapa. Toda a exclusão visa a alcançar consistência das informações apresentadas na RAIS, buscando corrigir erros sistemáticos e, principalmente, à formatação de uma base de dados longitudinal. Na Etapa I excluiu-se trabalhadores cujo contratos de trabalhos apresentam *missing* ou identificador (Id) incompleto. Em seguida, na Etapa II, foram excluídos todos os vínculos com contratos de trabalho com rendimentos nulos. Trabalhadores com contrato de trabalho com horas trabalhadas menor ou igual a 20 horas semanais foram deletados da base de dados – Etapa III. Na Etapa IV foram excluídos os trabalhadores com divergência na informação do gênero reportado em cada contrato de trabalho, isto é, trabalhadores com gênero diferente da média. Na Etapa V excluiu-se os trabalhadores com mais de cinco vínculos ativos e dentre estes, excluiu-se os contratos de trabalho cujas horas trabalhadas são inferiores à média das horas contratadas em todos os vínculos do trabalhador. Na Etapa VI foram excluídos todos os trabalhadores cujo contratos de trabalho tem horas trabalhadas inferior à média. Na Etapa VII considera apenas contratos de trabalhos mais atuais, excluindo aqueles que estão abaixo da média. Para trabalhadores que permaneceram com mais de um contrato de trabalho após a aplicação das etapas V-VII, escolheu-se apenas um após a aleatorização desses contratos por trabalhador – Etapa VIII.

Tabela F.1 – Tratamento dos microdados de trabalhadores da RAIS: 2002 -2014

Ano	RAIS	Etapa I	Etapa II	Etapa III	Etapa IV	Etapa V	Etapa VI	Etapa VII	Etapa VIII	Etapa VIII /	Georref / RAIS
2002	4.906.071	3.083.921	3.010.280	2.965.464	2.965.464	2.960.484	2.952.375	2.940.988	2.940.195	60%	63%
2003	4.928.566	3.177.260	3.106.736	3.067.886	3.067.886	3.063.049	3.054.813	3.042.749	3.041.385	62%	64%
2004	5.168.965	3.392.818	3.305.025	3.265.573	3.265.573	3.259.939	3.251.495	3.238.682	3.237.407	63%	66%
2005	5.426.661	3.662.139	3.556.420	3.511.762	3.511.762	3.506.210	3.496.391	3.480.356	3.478.310	64%	67%
2006	5.721.804	3.923.469	3.799.121	3.752.035	3.752.035	3.746.614	3.736.396	3.720.179	3.718.940	65%	69%
2007	6.160.103	4.273.449	4.121.926	4.070.806	4.070.806	4.065.428	4.054.702	4.036.892	4.035.048	66%	69%
2008	6.540.251	4.498.925	4.350.148	4.297.300	4.297.300	4.291.964	4.280.623	4.258.402	4.255.791	65%	69%
2009	6.722.364	4.608.329	4.457.825	4.403.021	4.403.021	4.399.249	4.387.266	4.368.293	4.366.628	65%	69%
2010	7.166.407	4.980.647	4.814.730	4.757.931	4.757.931	4.754.398	4.741.059	4.718.508	4.716.859	66%	69%
2011	7.380.405	5.170.559	5.003.420	4.944.447	4.944.447	4.940.676	4.925.262	4.893.047	4.889.103	66%	70%
2012	7.608.839	5.265.088	5.093.457	5.032.979	5.032.979	5.028.594	5.012.035	4.980.164	4.978.451	65%	69%
2013	7.684.602	5.318.540	5.147.644	5.086.616	5.086.616	5.082.129	5.064.298	5.029.258	5.028.373	65%	69%
2014	7.744.894	5.344.065	5.163.723	5.098.908	5.098.908	5.094.568	5.076.285	5.041.914	5.040.534	65%	69%
Média	6.396.918	4.361.478	4.225.420	4.173.441	4.173.441	4.168.716	4.156.385	4.134.572	4.132.848	64%	68%

Fonte: elaboração própria, a partir dos microdados da RAIS

APÊNDICE G – Tabelas Complementares dos Resultados

Tabela G.1 – Estimação de primeiro estágio da aglomeração sobre salário

		POLS	I	II	III	IV	V	VI	VII	VIII	IX
Emp											
ACTH		0,003***	0,003***	0,004***	0,003***	0,003***	0,003***	0,004***	0,003***	0,004***	0,003***
WACTH		-0,019***	-0,006***	-0,005***	0,006***	0,006***	0,008***	-0,005***	0,007***	0,007***	0,007***
W2ACTH		0,048***	0,070***	0,083***	0,081***	0,080***	0,068***	0,081***	0,072***	0,071***	0,068***
ACTC		0,028***	0,016***	0,009***	-0,013***	-0,014***	-0,012***	0,009***	-0,013***	-0,013***	-0,013***
WACTC		0,140***	0,111***	0,082***	0,057***	0,045***	0,035***	0,078***	0,048***	0,036***	0,030***
W2ACTC		0,086***	0,037***	-0,003***	-0,112***	-0,118***	-0,079***	0,002***	-0,084***	-0,085***	-0,078***
F-test	(1°	127767	104302	95944	119031	118815	76499	85410	92240	88966	73242
R²		0,595	0,856	0,925	0,947	0,960	0,952	0,925	0,957	0,960	0,960
WEmp											
ACTH		-0,002	0,001	0,002	0,003	0,002	0,002	0,002	0,003	0,003	0,003
WACTH		0,009	0,015	0,016	0,018	0,019	0,018	0,016	0,018	0,018	0,018
W2ACTH		0,006	0,028	0,038	0,043	0,043	0,039	0,037	0,040	0,040	0,039
ACTC		-0,019	-0,007	0,000	0,004	0,004	0,002	0,000	0,003	0,003	0,002
WACTC		0,194	0,125	0,069	0,019	0,017	0,013	0,067	0,017	0,015	0,013
W2ACTC		0,162	0,095	0,025	-0,101	-0,104	-0,078	0,026	-0,085	-0,086	-0,079
F-test	(1°	426364	285396	226085	307135	296181	218043	205396	249352	242710	209548
R²		0,619	0,876	0,946	0,967	0,974	0,968	0,946	0,973	0,974	0,974
W2Emp											
ACTH		0,002	0,001	0,003	0,002	0,002	0,002	0,003	0,002	0,002	0,002
WACTH		-0,001	0,015	0,008	0,012	0,012	0,012	0,008	0,012	0,012	0,012
W2ACTH		0,019	0,028	0,034	0,036	0,036	0,035	0,034	0,035	0,035	0,035
ACTC		-0,015	-0,007	-0,006	-0,004	-0,004	-0,003	-0,005	-0,003	-0,003	-0,003
WACTC		0,081	0,125	0,017	-0,001	-0,001	0,003	0,017	0,002	0,002	0,003
W2ACTC		0,200	0,095	0,071	-0,031	-0,034	-0,023	0,071	-0,025	-0,027	-0,024
F-test	(1°	568457	285396	317626	402471	388493	305179	291105	340120	337254	297056
R²		0,687	0,876	0,960	0,974	0,978	0,975	0,960	0,977	0,979	0,979
Trabalhador		Não	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Sim
Firma		Não	Não	Sim	Não	Não	Não	Sim	Sim	Sim	Sim
Célula		Não	Não	Não	Sim	Não	Não	Não	Sim	Sim	Sim
Indústria-Célula		Não	Não	Não	Não	Sim	Não	Não	Não	Sim	Sim
Match		Não	Não	Não	Não	Não	Sim	Não	Não	Não	Sim

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: Os dados concernentes a Emp, WEmp e W2Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios, exceto o modelo POLS que adiciona controles para gênero, idade, idade², setor da firma. Significante a 1% (***), 5% (**) e 10% (*).

Tabela G.2 – Estimação da média condição dos resíduos do primeiro estágio

	POLS	I	II	III	IV	V	VI	VII	VIII	IX
Residual (Emp)										
ACTH	-1,8e ⁻¹²	7,3e ⁻¹⁴	4,9e ⁻¹⁴	-3,9e ⁻¹⁴	8,7E ⁻¹⁴	-7,9e ⁻¹⁵	-3,5e ⁻¹⁴	-1,3e ⁻¹³	-2,2e ⁻¹³	-2,3e ⁻¹⁴
WACTH	-4,6e ⁻¹²	-7,7e ⁻¹³	-2,4e ⁻¹³	-2,5e ⁻¹³	-7,6e ⁻¹³	-6,9e ⁻¹³	-2,0e ⁻¹³	-4,5e ⁻¹³	-4,9e ⁻¹³	-4,1e ⁻¹³
W2ACTH	5,4e ⁻¹²	-1,8e ⁻¹²	-1,1e ⁻¹³	2,7e ⁻¹³	5,6e ⁻¹³	-2,0e ⁻¹²	-7,6e ⁻¹³	-1,2e ⁻¹²	-1,8e ⁻¹²	-2,2e ⁻¹²
ACTC	8,3e ⁻¹²	-8,9e ⁻¹⁴	-1,3e ⁻¹⁵	-7,7e ⁻¹⁴	3,9e ⁻¹⁴	-2,9e ⁻¹³	3,0e ⁻¹⁴	-1,7e ⁻¹³	-2,2e ⁻¹³	-2,7e ⁻¹³
WACTC	-3,3e ⁻¹¹	1,6e ⁻¹²	1,7e ⁻¹³	6,0e ⁻¹³	3,8e ⁻¹³	-2,0e ⁻¹²	1,9e ⁻¹³	-1,0e ⁻¹²	-1,7e ⁻¹²	-1,7e ⁻¹²
W2ACTC	5,4e ⁻¹²	5,4e ⁻¹³	-5,2e ⁻¹³	-5,0e ⁻¹³	1,8e ⁻¹³	-3,9e ⁻¹²	-1,0e ⁻¹²	-2,8e ⁻¹²	-1,6e ⁻¹²	-3,2e ⁻¹²
F-test	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Residual (wEmp)										
ACTH	-1,0e ⁻¹²	-4,9e ⁻¹⁵	-2,5e ⁻¹⁴	-4,4e ⁻¹⁴	-1,3e ⁻¹⁴	-6,7e ⁻¹⁴	-4,3e ⁻¹⁴	-6,7e ⁻¹⁴	-1,7e ⁻¹³	-3,9e ⁻¹⁴
WACTH	2,2e ⁻¹²	-2,5e ⁻¹³	-1,7e ⁻¹³	2,7e ⁻¹³	-2,7e ⁻¹³	-3,5e ⁻¹³	-1,1e ⁻¹⁴	-3,1e ⁻¹³	-2,2e ⁻¹³	-3,7e ⁻¹³
W2ACTH	-4,9e ⁻¹²	-9,0e ⁻¹³	1,4e ⁻¹³	2,8e ⁻¹⁴	3,7e ⁻¹³	-1,6e ⁻¹²	-5,8e ⁻¹³	-8,7e ⁻¹³	-1,9e ⁻¹²	-1,2e ⁻¹²
ACTC	4,1e ⁻¹²	-7,5e ⁻¹⁴	8,8e ⁻¹⁵	-3,8e ⁻¹⁴	2,9e ⁻¹⁴	-1,9e ⁻¹³	4,5e ⁻¹⁴	-1,1e ⁻¹³	-4,0e ⁻¹³	-1,4e ⁻¹³
WACTC	2,0e ⁻¹²	5,7e ⁻¹³	1,7e ⁻¹³	-3,7e ⁻¹³	1,5e ⁻¹³	-1,5e ⁻¹²	1,4e ⁻¹³	-6,6e ⁻¹³	-8,3e ⁻¹³	-1,1e ⁻¹²
W2ACTC	5,6e ⁻¹²	7,8e ⁻¹³	-2,3e ⁻¹³	-4,5e ⁻¹³	3,6e ⁻¹⁴	-3,1e ⁻¹²	-5,7e ⁻¹³	-1,8e ⁻¹²	-4,0e ⁻¹²	-2,5e ⁻¹²
F-test	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Residual(w2emp)										
ACTH	-3,1e ⁻¹³	-4,9e ⁻¹⁵	-4,9e ⁻¹⁴	6,3e ⁻¹⁴	9,9e ⁻¹⁴	-5,9e ⁻¹⁴	-9,9e ⁻¹⁵	-4,6e ⁻¹⁴	7,7e ⁻¹⁴	-3,7e ⁻¹⁴
WACTH	2,7e ⁻¹²	-2,5e ⁻¹³	-1,2e ⁻¹³	-3,5e ⁻¹³	-6,5e ⁻¹³	-2,5e ⁻¹⁴	-9,8e ⁻¹⁴	-5,2e ⁻¹⁶	-1,8e ⁻¹³	9,1e ⁻¹⁴
W2ACTH	-2,1e ⁻¹²	-9,0e ⁻¹³	1,4e ⁻¹³	4,9e ⁻¹³	7,0e ⁻¹³	-1,8e ⁻¹²	-5,1e ⁻¹³	-1,1e ⁻¹²	-2,5e ⁻¹²	-1,6e ⁻¹²
ACTC	5,7e ⁻¹³	-7,5e ⁻¹⁴	3,9e ⁻¹⁵	-2,2e ⁻¹⁴	6,3e ⁻¹⁴	-2,2e ⁻¹³	7,0e ⁻¹⁴	-1,3e ⁻¹³	-1,0e ⁻¹²	-1,9e ⁻¹³
WACTC	3,7e ⁻¹²	5,7e ⁻¹³	9,3e ⁻¹⁴	5,5e ⁻¹⁴	2,9e ⁻¹³	-1,4e ⁻¹²	-7,8e ⁻¹⁴	-7,9e ⁻¹³	-4,7e ⁻¹³	-1,1e ⁻¹²
W2ACTC	-8,2e ⁻¹²	7,8e ⁻¹³	-3,3e ⁻¹³	-3,7e ⁻¹³	-2,0e ⁻¹³	-2,6e ⁻¹²	-3,6e ⁻¹³	-1,4e ⁻¹²	-6,3e ⁻¹²	-2,0e ⁻¹²
F-test	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Trabalhador	Não	Sim	Não	Não	Não	Não	Sim	Sim	Sim	Sim
Firma	Não	Não	Sim	Não	Não	Não	Sim	Sim	Sim	Sim
Célula	Não	Não	Não	Sim	Não	Não	Não	Sim	Sim	Sim
Indústria-Célula	Não	Não	Não	Não	Sim	Não	Não	Não	Sim	Sim
Match	Não	Não	Não	Não	Não	Sim	Não	Não	Não	Sim

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: Os dados concernentes a Emp, WEmp e W²Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios, exceto o modelo POLS que adiciona controles para gênero, idade, idade², setor da firma. Significante a 1% (***), 5% (**) e 10% (*).

Tabela G.3 – Estimações de primeiro estágio da aglomeração sobre o salário e da média condicional dos resíduos com diferentes variáveis instrumentais

	Emp	WEmp	W2Emp	Resíduos (Emp)	Resíduos (WEmp)	Resíduos (W2Emp)
Área Comercial IV						
ACTC	-0,017***	-0,0005***	-0,005***	-2.96e ⁻¹³	-2.24e ⁻¹³	-1.83e ⁻¹³
WACTC	0,0488***	0,0331***	0,019***	-5.90e ⁻¹²	-4.32e ⁻¹²	-3.60e ⁻¹²
W2ACTC	-0,023***	-0,042***	0,008***	-1.28e ⁻¹¹	-9.55e ⁻¹²	-8.03e ⁻¹²
F-test (1º estágio)	139992	397371	586913	0	0	0
Prob>F	0,00	0,00	0,00	1	1	1
R ² Ajust.	0,978	0,956	0,966	0,00	0,00	0,00
Obs.	4573205	4573205	4573205	4573205	4573205	4573205
Área Residencial IV						
ACTH	0,0032***	0,0027***	0,0021***	-2.92e ⁻¹⁴	-5.72e ⁻¹⁴	-5.21e ⁻¹⁴
WACTH	0,0077***	0,0182***	0,0123***	-4.66e ⁻¹³	-1.73e ⁻¹³	3.45e ⁻¹⁴
W2ACTH	0,0667***	0,0373***	0,0349***	-2.04e ⁻¹²	-1.45e ⁻¹²	-1.50e ⁻¹²
F-test (1º estágio)	5143	7311	5026	0	0	0
Prob>F	0,00	0,00	0,00	1	1	1
R ² Ajust.	0,960	0,974	0,978	0,00	0,00	0,00
Obs.	4573205	4573205	4573205	4573205	4573205	4573205
Áreas Comercial e Residencial IV						
ACTHC	0,0042***	0,0020***	-0,0001***	-1.12e ⁻¹³	-9.60e ⁻¹⁴	-8.43e ⁻¹⁴
WACTHC	0,0193***	0,0170***	0,0095***	-1.11e ⁻¹²	-8.01e ⁻¹³	-7.85e ⁻¹³
W2ACTHC	0,0221***	0,0155***	0,0172***	-3.71e ⁻¹²	-2.59e ⁻¹²	-2.19e ⁻¹²
F-test (1º estágio)	24936	61265	58728	0	0	0
Prob>F	0,00	0,00	0,00	1	1	1
R ² Ajust.	0,968	0,957	0,967	0,00	0,00	0,00
Obs.	4.573.205	4573205	4573205	4573205	4573205	4573205

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: Os dados concernentes a Emp, WEmp e W2Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios. Significante a 1% (***), 5% (**) e 10% (*).

Tabela G.4 – Estimações de primeiro estágio da aglomeração sobre o salário e da média condicional dos resíduos com células 0,5 km²

	Emp	WEmp	W2Emp	W3Emp	Resíduos (Emp)	Resíduos (WEmp)	Resíduos (W2Emp)	Resíduos (W3Emp)
ACTH	0,009***	0,002***	0,003***	0,002***	-8.50e ⁻¹⁴	-5,4e ⁻¹⁴	-5,7e ⁻¹⁴	-4,6e ⁻¹⁴
WACTH	0,012***	0,012***	0,005***	0,006***	-4.85e ⁻¹³	-3,2e ⁻¹³	-2,7e ⁻¹³	-2,8e ⁻¹³
W2ACTH	0,008***	0,008***	0,011***	0,006***	-6.34e ⁻¹³	-6,2e ⁻¹³	-4,9e ⁻¹³	-4,1e ⁻¹³
W3ACTH	0,045***	0,024***	0,018***	0,018***	-2.23e ⁻¹²	-1,4e ⁻¹²	-1,3e ⁻¹²	-1,2e ⁻¹²
ACTC	0,017***	0,009***	-0,005***	-0,002***	-1.54e ⁻¹³	-1,1e ⁻¹³	-8,7e ⁻¹⁴	-7,7e ⁻¹⁴
WACTC	0,052***	0,029***	-0,001***	-0,002***	-1.38e ⁻¹²	-8,2e ⁻¹³	-8,1e ⁻¹³	-6,9e ⁻¹³
W2ACTC	-0,080***	0,010***	0,020***	0,018***	-3.24e ⁻¹²	-2,2e ⁻¹²	-1,8e ⁻¹²	-1,6e ⁻¹²
W3ACTC	0,025***	-0,002***	0,047***	0,060***	-6.03e ⁻¹²	-4,4e ⁻¹²	-3,8e ⁻¹²	-3,2e ⁻¹²
F-test (1º estágio)	20462	32012	33805	48502	0,00	0,00	0,00	0,00
Prob>F	0,00	0,00	0,00	0,00	1	1	1	1
R ²	0,93	0,96	0,96	0,97	0,00	0,00	0,00	0,00
Obs.	4573205	4573205	4573205	4573205	4573205	4573205	4573205	4573205

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: Os dados concernentes a Emp, WEmp e W2Emp são agregações de trabalhadores por células quadradas com 0,5 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios. Significante a 1% (***), 5% (**) e 10% (*).

Tabela G.5 – Estimações de primeiro estágio da aglomeração sobre o salário e da média condicional dos resíduos com células 2 km²

	Emp	Wemp	Resíduos (Emp)	Resíduos (Wemp)
ACTH	0,0095***	0,0040***	-2,45e ⁻¹³	-2,1e ⁻¹³
WACTCH	0,0203***	0,0079***	-1,84e ⁻¹²	-1,6e ⁻¹²
ACTC	0,0559***	0,0002	-7,34e ⁻¹³	-6,9e ⁻¹³
WACTC	0,0317***	0,0266***	-6,97e ⁻¹²	-5,8e ⁻¹²
F-test (1º estágio)	44347	14634	0,00	0,00
Prob>F	0,00	0,00	1	1
R ²	0,93	0,96	0,00	0,00
Obs.	4573205	4573205	4573205	4573205

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: Os dados concernentes a Emp, Wemp e W²Emp são agregações de trabalhadores por células quadradas com 2 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios. Significante a 1% (***), 5% (**) e 10% (*).

Tabela G.6 – Estimação de primeiro estágio da aglomeração sobre o salário e média condicional por ciclo educacional

	Ciclo I			Ciclo II		
	Emp	Wemp	W2Emp	Emp	Wemp	W2Emp
ACTH	0,0013***	0,0001***	0,0018***	0,0059***	0,0050***	0,0017***
WACTH	0,0034***	0,0152***	0,0105***	0,0095***	0,0125***	0,0115***
W2ACTH	0,0673***	0,0446***	0,0375***	0,0617***	0,0380***	0,0347***
ACTC	-0,0229***	0,0035***	-0,0045***	-0,0076***	0,0001***	-0,0019***
WACTC	0,0411***	0,0130***	0,0065***	0,0096***	0,0074***	0,0011***
W2ACTC	-0,0898***	-0,0913***	-0,0326***	-0,0731***	-0,0570***	-0,0104***
F-test (1º estágio)	29638	79174	110877	20621	60653	90334
Prob>F	0,00	0,00	0,00	0,00	0,00	0,00
R ² Ajust.	0,953	0,969	0,974	0,964	0,979	0,983
Obs.	1813489	1813489	1813489	1432227	1432227	1432227
	Ciclo III			Ciclo IV		
	Emp	Wemp	W2Emp	Emp	Wemp	W2Emp
ACTH	0,0035***	0,0030***	0,0019***	0,0068***	0,0053***	0,0029***
WACTH	0,0074***	0,0201***	0,0139***	0,0175***	0,0299***	0,0166***
W2ACTH	0,0772***	0,0370***	0,0359***	0,0594***	0,0248***	0,0276***
ACTC	-0,0113***	0,0007***	-0,0029***	-0,0007***	0,0007***	-0,0019***
WACTC	0,0396***	0,0192***	-0,0042***	0,0187***	0,0207***	0,0039***
W2ACTC	-0,0467***	-0,0476***	-0,0094***	-0,0220***	-0,0522***	-0,0213***
F-test (1º estágio)	2568	8188	12798	13344	45984	58608
Prob>F	0	0	0	0	0	0
R ²	0,962	0,978	0,982	0,967	0,980	0,984
Obs.	257504	257504	257504	1069985	1069985	1069985

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: Os dados concernentes a Emp, Wemp e W²Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios. Significante a 1% (***), 5% (**) e 10% (*). Ciclo I diz respeito a trabalhadores com até ensino médio incompleto, Ciclo II considera apenas trabalhadores com Ensino Médio completo, Ciclo III considera trabalhos com Ensino Superior incompleto e Ciclo IV, trabalhadores com Ensino Superior completo, mestrado e/ou doutorado.

Tabela G.7 – Estimação da média condicional dos resíduos do primeiro estágio por ciclo educacional

	Ciclo I			Ciclo II		
	Emp	Wemp	W2Emp	Emp	Wemp	W2Emp
ACTH	1,69e ⁻¹⁴	9,59e ⁻¹⁵	9,55e ⁻¹⁵	7,73e ⁻¹⁵	1,29e ⁻¹⁴	1,19e ⁻¹⁴
WACTH	2,12e ⁻¹³	2,36e ⁻¹³	8,78e ⁻¹⁴	1,72e ⁻¹³	3,63e ⁻¹⁴	4,14e ⁻¹⁴
W2ACTH	1,06e ⁻¹²	8,28e ⁻¹³	9,34e ⁻¹³	3,04e ⁻¹³	2,43e ⁻¹³	2,34e ⁻¹³
ACTC	1,60e ⁻¹³	1,33e ⁻¹³	1,28e ⁻¹³	3,38e ⁻¹⁴	2,47e ⁻¹⁴	2,21e ⁻¹⁴
WACTC	8,89e ⁻¹³	7,65e ⁻¹³	7,40e ⁻¹³	4,51e ⁻¹³	2,85e ⁻¹³	2,57e ⁻¹³
W2ACTC	2,37e ⁻¹²	1,97e ⁻¹²	1,67e ⁻¹²	9,08e ⁻¹³	5,13e ⁻¹³	4,66e ⁻¹³
F-test	0,00	0,00	0,00	0,00	0,00	0,00
Prob>F	1	1	1	1	1	1
R ²	0,00	0,00	0,00	0,00	0,00	0,00
Obs.	1813489	1813489	1813489	1432227	1432227	1432227
	Ciclo III			Ciclo IV		
	Emp	Wemp	W2Emp	Emp	Wemp	W2Emp
ACTH	3,49e ⁻¹⁵	4,97e ⁻¹⁵	-8,02e ⁻¹⁶	-4,96e ⁻¹⁴	6,53e ⁻¹⁴	2,99e ⁻¹⁴
WACTH	8,78e ⁻¹⁵	-6,77e ⁻¹⁵	1,39e ⁻¹⁴	5,08e ⁻¹⁴	-1,29e ⁻¹³	-1,54e ⁻¹³
W2ACTH	3,08e ⁻¹⁴	7,11e ⁻¹⁴	2,00e ⁻¹⁴	2,32e ⁻¹⁴	-3,06e ⁻¹³	-4,47e ⁻¹⁴
ACTC	2,04e ⁻¹⁶	5,77e ⁻¹⁵	3,92e ⁻¹⁵	-8,15e ⁻¹⁴	1,30e ⁻¹³	4,56e ⁻¹⁴
WACTC	-8,51e ⁻¹⁵	6,46e ⁻¹⁴	2,61e ⁻¹⁴	3,25e ⁻¹³	-7,82e ⁻¹³	-2,99e ⁻¹³
W2ACTC	2,46e ⁻¹³	-1,53e ⁻¹⁴	7,82e ⁻¹⁴	-9,28e ⁻¹³	8,51e ⁻¹³	1,09e ⁻¹³
F-test	0,00	0,00	0,00	0,00	0,00	0,00
Prob>F	1	1	1	1	1	1
R ²	0,00	0,00	0,00	0,00	0,00	0,00
Obs.	257504	257504	257504	1069985	1069985	1069985

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: Os dados concernentes a Emp, WEmp e W2Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios. Significante a 1% (***), 5% (**) e 10% (*). Ciclo I diz respeito a trabalhadores com até ensino médio incompleto, Ciclo II considera apenas trabalhadores com Ensino Médio completo, Ciclo III considera trabalhos com Ensino Superior incompleto e Ciclo IV, trabalhadores com Ensino Superior completo, mestrado e/ou doutorado.

Tabela G.8 – Estimação de primeiro estágio da aglomeração sobre o salário por SBD

	<i>SBD</i>^{Q99}			<i>SBD</i>^{Q95}		
	Emp	Wemp	W2Emp	Emp	Wemp	W2Emp
ACTH	0,0010***	-0,0004***	0,0010***	0,0036***	0,0028***	0,0025***
WACTH	-0,0159***	0,0187***	0,0068***	0,0049***	0,0192***	0,0130***
W2ACTH	0,0424***	-0,0079***	-0,0009***	0,0642***	0,0327***	0,0305***
ACTC	-0,0186***	0,0014***	-0,0066***	-0,0145***	0,0006***	-0,0043***
WACTC	-0,0100***	-0,0244***	-0,0227***	0,0186***	0,0046***	-0,0041***
W2ACTC	-0,0802***	-0,0923***	-0,0436***	-0,0994***	-0,0941***	-0,0371***
F-test (1º estágio)	3118	11078	10840	28368	83810	120309
Prob>F	0,00	0,00	0,00	0,00	0,00	0,00
R ² Ajust.	0,948	0,968	0,976	0,955	0,969	0,974
Obs.	1530924	1530924	1530924	3239272	3239272	3239272

	<i>SBD</i>^{Q90}			<i>NSBD</i>		
	Emp	Wemp	W2Emp	Emp	Wemp	W2Emp
ACTH	0,0030***	0,0024***	0,0020***	0,0016***	0,00002***	0,0011***
WACTH	0,0064***	0,0173***	0,0119***	-0,0001***	0,0026***	0,0050***
W2ACTH	0,0665***	0,0378***	0,0340***	0,0008***	0,0027***	0,0059***
ACTC	-0,0132***	0,0013***	-0,0036***	0,0058***	-0,0073***	0,0058***
WACTC	0,0259***	0,0107***	0,0010***	-0,0671***	-0,0089***	0,0116***
W2ACTC	-0,0844***	-0,0842***	-0,0282***	-0,0359***	-0,0083***	0,0117***
F-test (1º estágio)	49317	141819	199533	460	3798	16244
Prob>F	0,00	0,00	0,00	0,00	0,00	0,00
R ² Ajust.	0,958	0,972	0,976	0,913	0,935	0,969
Obs.	4016538	4016538	4016538	556667	556667	556667

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: Os dados concernentes a Emp, Wemp e W2Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios. Significante a 1% (***), 5% (**) e 10% (*). ***SBD*^{Q99}**, ***SBD*^{Q95}** e ***SBD*^{Q90}** são centros identificados considerando os quartis 99%, 95% e 90% da distribuição de médias espaciais calculadas a partir do modelo GWR, respectivamente. ***NSBD*** são áreas que não constam em nenhuma das anteriores.

Tabela G.9 – Estimação da média condição dos resíduos do primeiro estágio por SBD

	<i>SBD</i>^{Q99}			<i>SBD</i>^{Q95}		
	Emp	Wemp	W2Emp	Emp	Wemp	W2Emp
ACTH	1,73e ⁻¹³	1,45e ⁻¹³	1,59e ⁻¹³	-8,55e ⁻¹⁴	-9,60e ⁻¹⁴	-8,65e ⁻¹⁴
WACTH	2,24e ⁻¹²	2,18e ⁻¹²	1,69e ⁻¹²	-9,84e ⁻¹³	-6,42e ⁻¹³	-7,22e ⁻¹⁵
W2ACTH	8,77e ⁻¹²	6,73e ⁻¹²	5,99e ⁻¹²	-3,22e ⁻¹²	-2,26e ⁻¹²	-2,50e ⁻¹²
ACTC	6,84e ⁻¹³	4,70e ⁻¹³	4,26e ⁻¹³	-3,61e ⁻¹³	-2,29e ⁻¹³	-2,60e ⁻¹³
WACTC	4,90e ⁻¹²	3,59e ⁻¹²	3,16e ⁻¹²	-1,81e ⁻¹²	-1,27e ⁻¹²	-1,19e ⁻¹²
W2ACTC	3,06e ⁻¹³	2,55e ⁻¹²	2,08e ⁻¹²	-2,80e ⁻¹²	-2,42e ⁻¹²	-1,62e ⁻¹²
F-test	0,00	0,00	0,00	0,00	0,00	0,00
Prob>F	1	1	1	1	1	1
R ² Ajust.	0,00	0,00	0,00	0,00	0,00	0,00
Obs.	1.530.924	1.530.924	1.530.924	3.239.272	3.239.272	3.239.272

	<i>SBD</i>^{Q90}			<i>NSBD</i>		
	Emp	Wemp	W2Emp	Emp	Wemp	W2Emp
ACTH	-9,23e ⁻¹⁴	-9,72e ⁻¹⁴	-1,16e ⁻¹³	-1,17e ⁻¹⁴	-7,70e ⁻¹⁵	-6,86e ⁻¹⁵
WACTH	-1,00e ⁻¹²	-6,72e ⁻¹³	-1,25e ⁻¹³	-4,13e ⁻¹⁴	-2,80e ⁻¹⁴	-2,46e ⁻¹⁴
W2ACTH	-4,07e ⁻¹²	-2,78e ⁻¹²	-2,91e ⁻¹²	-1,03e ⁻¹³	-6,46e ⁻¹⁴	-5,30e ⁻¹⁴
ACTC	-5,11e ⁻¹³	-3,74e ⁻¹³	-3,48e ⁻¹³	-1,28e ⁻¹⁴	-1,05e ⁻¹⁴	-9,95e ⁻¹⁵
WACTC	-3,25e ⁻¹²	-2,15e ⁻¹²	-2,00e ⁻¹²	-1,02e ⁻¹³	-6,42e ⁻¹⁴	-5,72e ⁻¹⁴
W2ACTC	-5,95e ⁻¹²	-4,43e ⁻¹²	-3,51e ⁻¹²	1,55e ⁻¹³	8,74e ⁻¹⁴	6,16e ⁻¹⁴
F-test	0,00	0,00	0,00	0,00	0,00	0,00
Prob>F	1	1	1	1	1	1
R ² Ajust.	0,00	0,00	0,00	0,00	0,00	0,00
Obs.	4016538	4016538	4016538	556667	556667	556667

Fonte: elaboração própria, a partir dos microdados da RAIS (2002-2014). Nota: Os dados concernentes a Emp, Wemp e W2Emp são agregações de trabalhadores por células quadradas com 1 km² de área. Todos os modelos incluem efeito fixo de trabalhador, firma, área, setor-área, *match* trabalhador-firma-área e controles para experiência, experiência² e educação do trabalhador, tamanho (faixa de total de trabalhador na firma) e especialização da firma e dummies de municípios. Significante a 1% (***), 5% (**) e 10% (*). ***SBD*^{Q99}**, ***SBD*^{Q95}** e ***SBD*^{Q90}** são centros identificados considerando os quartis 99%, 95% e 90% da distribuição de médias espaciais de emprego calculadas a partir do modelo GWR, respectivamente. ***NSBD*** são áreas que não constam em nenhuma das anteriores.