

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

Diana Lúcia Gonzaga da Silva (UFBA)

Carlos Roberto Azzoni (USP)

**Resumo:** O objetivo deste artigo é identificar a contribuição dos efeitos de aglomeração e do *sorting* das heterogeneidades dos trabalhadores e firmas para a determinação dos salários individuais e dos salários dos arranjos populacionais do Brasil (mercados de trabalho locais). A maior produtividade das áreas densas pode ser atribuída à concentração de trabalhadores e firmas mais produtivos: *sorting*. O primeiro estágio do modelo estima os efeitos de localização a partir de uma equação salarial, em um painel pareado de trabalhadores e firmas com microdados da RAIS (2002-2014). O segundo estágio decompõe os efeitos de localização em componentes locais e atributos não observados das firmas e trabalhadores. A estratégia de identificação considera os efeitos fixos dos trabalhadores e firmas, uma variável instrumental para a densidade e os dados de satélite da luminosidade noturna para mensurar a área dos arranjos. O modelo principal encontrou um efeito da densidade sobre os salários locais de 4,9%, o qual é superior ao *lower bound* da literatura (3%).

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

**Abstract:** The goal of this study is to identify the contribution of agglomeration effects and spatial sorting of worker and firm for the determination of individual and local wages in Brazilian urban agglomerations (labor market areas). The higher productivity of dense areas can be attributed to the concentration of more productive workers and firms: *sorting*. The first stage of model estimates the location effects as from a wage equation in a matched panel of workers and firms with microdata of RAIS (2002-2014). The second stage decomposes the location effects into local components and unobserved attributes of firms and workers. The identification strategy considers the fixed effects of workers and firms, an instrumental variable to density and satellite data on illumination to estimate the area in each local labour markets. The model of preference indicates a density effect on wages of 4.9%, much higher than the literature lower bound (3%).

**Keywords:** Wages. Agglomeration Economies. *Sorting*. Firm heterogeneity. Worker heterogeneity.

Código JEL: J24, J31, R23, C23

## 1 Introdução

Diversos estudos em Economia buscam compreender os determinantes e as disparidades salariais. As pesquisas em Economia do Trabalho avançaram no controle de características observáveis e não observáveis dos trabalhadores, a partir da equação minceriana (MINCER, 1974). Entretanto, permanece um diferencial salarial favorável aos grandes centros urbanos. Isso motivou o surgimento de estudos que integram Economia Urbana e Economia do Trabalho para explicar o diferencial salarial espacial, através das economias de aglomeração urbana. Nessa abordagem, a localização é incluída como um determinante salarial.

A distribuição espacial de salários no Brasil apresenta um comportamento considerado historicamente desigual<sup>1</sup>. Os dados da Relação Anual de Informações Sociais (RAIS) do Ministério do Trabalho e Emprego (MTE) mostram que, entre as regiões metropolitanas (RMs), o salário médio de João Pessoa correspondeu a 65% do salário médio da RM de São Paulo e 49% do salário da RM de Brasília em 2014. Por outro lado, os salários nas regiões metropolitanas são maiores do que as médias dos seus estados. Isso pode sugerir a existência de ganhos de aglomeração nos mercados de trabalho mais densos, se o diferencial de custo de vida não for contabilizado. No entanto, o custo de vida mais elevado pode ser o resultado da maior concentração populacional e das atividades produtivas, que torna a disponibilidade de terra e de outros fatores mais escassa. Menezes e Azzoni (2006) mostraram que o diferencial de custo de vida não foi individualmente capaz de explicar o diferencial de salários entre as regiões metropolitanas do Brasil, usando dados da Pesquisa Nacional por Amostras de Domicílios (PNAD) do Instituto Brasileiro de Geografia e Estatística (IBGE), para o período de 1981 a 2003. As características regionais e do capital humano foram mais relevantes.

As diferenças observadas de salários podem ocorrer por diversos fatores, dentre eles os diferenciais de custo de vida e de mobilidade, os diferenciais de produtividade, as dotações de amenidades locais, as segmentações regionais nos mercados de trabalho, tais como, formal, informal, setorial e por tamanho de empresas, dentre outros (SAVEDOFF, 1990; SERVO; AZZONI, 2002; SILVEIRA NETO; AZZONI, 2004; ROCHA; SILVEIRA NETO; GOMES, 2011). Azzoni (1999) mostrou que 40% do diferencial de renda pessoal e familiar, entre os estados do Brasil, era explicado pelo componente regional, usando dados dos Censos Demográficos de 1960-1991. Silveira Neto e Azzoni (2004) indicaram que as características pessoais, ocupacionais e os custos de vida eram mais importantes do que as amenidades locais para o diferencial de renda entre áreas urbanas das regiões Nordeste e Sudeste em 2002. A educação apresentou a maior contribuição para as desigualdades salariais entre grandes áreas metropolitanas do Brasil, 17,6% e 31%, entre 1992 e 1997, em Servo e Azzoni (2002).

De acordo com Barros, Franco e Mendonça (2007), a desigualdade no rendimento do trabalho tem declinado no Brasil desde 1995, apresentando queda mais intensa entre 2001 e 2005. Menezes-Filho e Komatsu (2015) apontam que o crescimento real do salário mínimo, mais intenso nos anos 2000, contribuiu para a redução da dispersão salarial entre 2004 e 2014. A educação é considerada o principal determinante da redução na desigualdade de renda do país, a partir dos anos 2000<sup>2</sup>. Barros, Henriques e Mendonça (2000) mostraram que a educação é responsável por 40% dos diferenciais salariais no Brasil e o impacto de um ano adicional de educação sobre os salários chega a 16%. A redução na desigualdade educacional dos trabalhadores brasileiros contribuiu em 17% para a redução da desigualdade salarial, entre 2001 e 2005. No mesmo período, houve uma intensificação na queda dos diferenciais de

---

<sup>1</sup> Ver Langoni (2005).

<sup>2</sup> Ver Barros, Henriques e Mendonça (2000) e Menezes-Filho, Fernandes e Picchetti (2007).

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

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

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

Em um estudo recente, o IBGE (2015) mostrou que, aproximadamente 56% da população do Brasil residia em arranjos populacionais, definidos a partir da intensidade dos fluxos para trabalho e estudo e das características urbanas, com base nos dados do Censo Populacional de 2010. Por sua vez, os arranjos concentravam em torno de 82% do emprego formal em 2010, sendo 57% somente nos grandes arranjos (população acima de 750 mil). De acordo com os dados da RAIS, os salários médios nos arranjos populacionais de dezembro de 2014, eram aproximadamente 20% maiores do que nos municípios que não pertenciam a algum arranjo e 40% maiores nos grandes arranjos. Esses dados conjuntamente reforçam os indícios da existência de fatores específicos locais em áreas urbanas densas, que dinamizam os mercados de trabalho locais e elevam a produtividade nessas áreas.

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

Por outro lado, os ganhos associados às características de trabalhadores e firmas localizados nas áreas densas podem ser confundidos com ganhos de aglomeração. Se a heterogeneidade das firmas e trabalhadores dos grandes centros urbanos não é considerada, deve existir um viés nas estimativas dos efeitos de localização, associado ao *sorting* espacial na determinação de salários. O *sorting* por habilidades surge se trabalhadores com melhores atributos tendem a se aglomerar em mercados de trabalho locais densos e qualificados. A seleção de firma pode surgir se apenas as mais produtivas sobrevivem nos grandes centros

urbanos. No entanto, até o momento não foi encontrado algum estudo no Brasil que tenha considerado os efeitos de firma na determinação dos efeitos de localização sobre os salários.

O controle da seleção de trabalhadores mais produtivos para as firmas que pagam maiores salários requer a disponibilidade de dados longitudinais pareados (*matched employer-employee data*) de trabalhador e firma. A limitação empírica para estimar modelos com heterogeneidade não observada de trabalhador e firma, usando o método de Mínimos Quadrados Ordinários (MQO), surge da complexidade computacional para incluir uma grande dimensão de identificadores nas estimações econométricas com dados longitudinais. Abowd, Kramarz e Margolis (1999) desenvolveram abordagens alternativas para estimar os parâmetros de equações salariais, dada a limitação computacional da solução de MQO. Desde o estudo de Abowd, Kramarz e Margolis (1999), a literatura internacional em Economia do Trabalho tem buscado soluções estruturais e estatísticas factíveis computacionalmente para estimar os parâmetros de interesse, em modelos que lidam com múltiplos efeitos fixos (ABOWD; CREECY; KRAMARZ, 2002; ANDREWS; SCHANK; UPWARD, 2006; GUIMARÃES; PORTUGAL, 2010; McCAFFREY, 2012; MITTAG, 2012; GAURE, 2013).

Este trabalho contribui para a literatura internacional e nacional sobre os determinantes salariais espaciais. Em relação à literatura internacional, contribui com a inclusão de um grupo mais amplo de controles, a exemplo dos efeitos de firma. Além disso, o estudo sugere uma medida mais adequada para a densidade, utilizando a área com luminosidade noturna para mensurar a área com efetiva ocupação humana e corrigir a subestimação dos efeitos de aglomeração. O estudo também contribui com o uso do instrumento *Bartik*, tradicionalmente utilizado nos estudos empíricos da Economia do Trabalho, para lidar com a endogeneidade da densidade e identificar o efeito exógeno da aglomeração. Para a literatura nacional, o avanço corresponde ao controle simultâneo do *sorting* espacial da firma e do trabalhador nos salários individuais e na identificação dos efeitos de aglomeração. O trabalho utilizará a definição do IBGE (2015) de arranjos populacionais, a qual se mostra mais adequada para a identificação dos mercados de trabalho locais aglomerados, tendo em vista a intensidade dos fluxos para trabalho e estudo e as características de urbanização adotadas na definição dessas categorias.

O objetivo desse trabalho é identificar a contribuição dos efeitos de aglomeração e do *sorting* espacial de trabalhador e firma para a determinação de salários individuais e para o diferencial salarial entre os arranjos populacionais do Brasil. O estudo propõe identificar a contribuição do *sorting* espacial a partir do controle dos efeitos de firma e trabalhador, recorrendo aos avanços metodológicos da literatura internacional. Um painel de micro dados combinados de trabalhador e firma, extraído da RAIS (MTE), será utilizado para estimar as equações de salário com as características observadas dos trabalhadores e do emprego, a heterogeneidade não observada dos trabalhadores e firmas e os efeitos de localização em arranjos populacionais, para o período entre 2002 e 2014. A estratégia de identificação propõe lidar com o *sorting* das heterogeneidades não observadas que resulta da atração e concentração dos trabalhadores e firmas mais produtivos nos mercado de trabalho locais densos. O estudo utiliza uma variável instrumental para lidar com a endogeneidade da densidade com os salários locais e utiliza uma medida mais adequada para a densidade.

Além desta introdução, este trabalho é composto por mais três seções. A segunda seção apresenta a metodologia da pesquisa e sua estratégia de identificação, o banco de dados, o modelo econométrico e os métodos de estimação. Os resultados são analisados na seção três. Por fim, a última seção apresenta as conclusões do trabalho.

## **2 Banco de Dados e Método**

Esta seção apresenta a metodologia do estudo, incluindo o seu banco de dados, modelo econométrico e estratégia de identificação.

## 2.1 Banco de Dados

Este estudo utiliza os micro dados da Relação Anual de Informações Sociais (RAIS) do Ministério do Trabalho e Emprego (MTE) como principal banco de dados para a estimação dos resultados de interesse. A RAIS é um banco de registros administrativos combinados de trabalhador e estabelecimento (*matched employer-employee data*), que são declarados compulsoriamente a cada ano pelas empresas do mercado de trabalho formal do Brasil. Para a construção do painel foram excluídos os registros sem o identificador do trabalhador (PIS), da firma (CNPJ) ou do município, sem remuneração e com carga horária semanal contratada inferior a 20 horas. Os trabalhadores da Administração Pública foram excluídos, tendo em vista que a lógica de funcionamento e de determinação salarial desse setor é diferente da lógica de mercado. O painel é construído considerando apenas o vínculo principal do trabalhador (mais antigo, maior carga horária). A geração da amostra adotou dois critérios necessários para a estratégia de identificação. O primeiro requer suficiente mobilidade de estabelecimentos entre as áreas de emprego aglomeradas (arranjos populacionais) adotadas neste estudo. A mobilidade permite identificar os efeitos fixos de firma e de arranjos separadamente. O segundo critério requer uma amostra aleatória representativa para os arranjos populacionais.

Para satisfazer o primeiro critério, cerca de 42% das observações da amostra é composta pelos trabalhadores dos estabelecimentos que apresentaram alguma mobilidade entre arranjos. Para garantir a representatividade dos arranjos e aleatoriedade dos dados foi extraída uma amostra de 1% aleatorizada e estratificada entre os arranjos, para compor o banco de dados. O banco de micro dados final é composto por 6.048.393 observações, que correspondem a um painel anual balanceado com 465.261 trabalhadores, 271.951 estabelecimentos e 1.081.990 combinações de trabalhador e estabelecimento, para o período entre 2002 e 2014.

Para a construção dos salários foi utilizado a remuneração nominal do trabalhador do mês de dezembro de cada ano. O Índice Nacional de Preços ao Consumidor Amplo – IPCA (IBGE) das regiões metropolitanas foi utilizado para deflacionar os salários nominais, assumindo dezembro de 2014 como o período base. O IPCA é calculado apenas para onze regiões metropolitanas do Brasil: Belém; Fortaleza; Recife; Salvador; São Paulo; Rio de Janeiro; Goiânia; Distrito Federal; Belo Horizonte; Curitiba e Porto Alegre. Os índices das regiões metropolitanas foram extrapolados para os estados correspondentes e o critério de maior aproximação foi utilizado para os estados que não possuíam uma RM contemplada<sup>3</sup>. O salário mensal real foi transformado em salário real hora, com as horas semanais contratadas. Tendo em vista as mudanças em função da criação e extinção de municípios ao longo do período analisado, o estudo fará uso de Áreas Mínimas Comparáveis (AMC).

## 2.2 Métodos

Para lidar com as diferentes dimensões da heterogeneidade salarial, este trabalho utiliza as estratégias de estimação desenvolvidas a partir do modelo de decomposição estatística de salários de Abowd, Kramarz e Margolis – AKM (1999). O modelo AKM é considerado a primeira análise estatística da heterogeneidade simultânea de trabalhador e firma na determinação de salários. O estudo propôs decompor o diferencial de salários setorial da França através a partir de modelo estimado em dois estágios. O primeiro estágio busca encontrar as estimativas dos efeitos setoriais a partir de uma equação de salários:

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

---

<sup>3</sup> Este procedimento foi adotado com base em Freguglia (2007).

onde,  $X$  é uma matriz ( $N^* \times P$ ) de características observáveis variantes no tempo,  $D$  é uma matriz ( $N^* \times N$ ) de indicadores para os indivíduos,  $F$  é uma matriz ( $N^* \times J$ ) de indicadores para a firma em que  $i$  trabalha em  $t$  ( $J$  é o número total de firmas),  $A$  ( $J \times K$ ) classifica cada uma das  $J$  firmas em um dos  $K$  setores,  $FA\kappa$  é, portanto, a variável indicadora de setor,  $(F\psi - FA\kappa)$  representa o efeito puro da firma,  $y$  é o vetor ( $N^* \times 1$ ) de salários,  $\varepsilon$  é o vetor de resíduos e  $N^* = NT$ . A versão da equação (1) sem o efeito puro de firma ( $M_{FA}F\psi$ ) e o efeito fixo de indivíduo ( $\theta$ ) implicará nos seguintes efeitos setoriais ( $\kappa^{**}$ ) inconsistentes:

$$\begin{aligned}\kappa^{**} &= \kappa + (A'F'M_XFA)^{-1}A'F'M_X(M_{FA}F\psi + D\theta) \\ &\equiv (A'F'M_XFA)^{-1}A'F'M_XF\psi + (A'F'M_XFA)^{-1}A'F'M_XD\theta.\end{aligned}\quad (2)$$

O modelo AKM adota a equação (2) para decompor os efeitos setoriais e verificar a contribuição dos efeitos de trabalhador, de firma e de ambos para tais diferenciais. O estudo propõe abordagens estatísticas alternativas à de mínimos quadrados para estimar o modelo completo, incluindo efeitos fixos de firmas e indivíduos. O método condicional inclui cada efeito fixo em diferentes estágios da estimação, até a completa identificação do modelo. Esse método consegue implementar a solução de MQO assumindo supostos adicionais de ortogonalidade. A estimação do modelo econométrico completo adotará a abordagem iterativa proposta por Guimarães e Portugal (2010) para estimar modelos com grande dimensão de efeitos fixos, tais como o modelo AKM. Correia (2014) desenvolveu um algoritmo de iteração de ponto fixo para implementar a abordagem no *software stata*. De maneira prática, o algoritmo faz uma regressão linear absorvendo (*demean*) os efeitos fixos indicados, como o Teorema de Frisch-Waugh-Lovell (FWL).

Os métodos tradicionais de painel, Efeitos Fixos (FE) e Mínimos Quadrados Ordinários Agrupado (POLS), serão adotados para a comparação com as estimativas do método condicional que estima o modelo completo. Além desses, o método de Mínimos Quadrados em Dois Estágios (2SLS) será utilizado para lidar com a endogeneidade da densidade do emprego no segundo estágio, na estimação com a variável instrumental.

### 2.3 Modelo Econométrico e Estratégia de Identificação

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

$$\begin{aligned}\ln(w_{it}) &= \beta_0 + \sum_e \beta_1^e educ_{it}^e + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + \beta_4 age_{it} + \beta_5 age_{it}^2 + \\ &\quad \beta_6 gender + \sum_s \beta_7^s sector_{it}^s + \beta_8 size\_firmsect_{j(i,t)} + \sum_t \beta_9^t T_t + \\ &\quad \sum_a \kappa_a lma_{it}^a + \theta_i + \psi_{j(i,t)} + \varphi_{i,j,t} + \varepsilon_{it}\end{aligned}\quad (3)$$

onde,  $\ln(w_{it})$  é o logaritmo natural do salário-hora real do trabalhador  $i$  no tempo  $t = \{2002, \dots, 2014\}$ ,  $educ_{it}^e$  são *dummies* para quatro ciclos escolares  $e = \{1, \dots, 4\}$  onde “até ensino médio incompleto” é a categoria base,  $exp_{it}$  é a experiência do trabalhador,  $age_{it}$  é a idade,  $gender$  é uma *dummy* de gênero (=1 se mulher),  $sector_{it}^s$  são *dummies* para os setores de atividade  $s = \{1, \dots, 16\}$  onde indústria de transformação é a categoria base,  $size\_firmsect_{j(i,t)}$  são *dummies* para o tamanho relativo da empresa  $j$  no setor de atividade correspondente e  $T_t$  são de *dummies* de tempo (2002 é a base). Para captar os efeitos de localização sobre os salários serão incluídos os efeitos fixos de cada arranjo populacional  $a = \{1, \dots, 365\}$ . Se o efeito fixo do arranjo (*Labor Market Area-LMA*) for positivo e significativo, o vetor  $\kappa_a$  será a magnitude do prêmio salarial urbano. As heterogeneidades não observadas do trabalhador ( $i$ ) e da firma ( $j$ ) são representadas pelos efeitos fixos  $\theta_i$  e  $\psi_{j(i,t)}$ . Por sua vez,

o efeito fixo combinado de trabalhador e firma  $\varphi_{i,j,t}$  representa o *matching* no mercado de trabalho local e  $\varepsilon_{it}$  é o termo de erro do modelo.

Os efeitos de localização serão estimados na equação salarial do primeiro estágio sem os efeitos fixos de indivíduo e firma. Dessa forma, serão recuperados os efeitos de localização “brutos”. A decomposição dos efeitos de localização do segundo estágio é dada por:

$$\hat{\kappa}_a = \gamma_0 + \ln(\text{employment density}_{a,t})\gamma_1 + \ln(\text{land area}_a)\gamma_2 + \text{touristic}_a\gamma_3 + \text{frontier}_a\gamma_4 + \text{region}_{r(a)}\gamma_5 + \bar{\theta}_{a,t}\gamma_6 + \bar{\psi}_{a,t}\gamma_7 + \rho_t + v_{a,t}. \quad (4)$$

A variável de interesse na estimação do segundo estágio é a densidade do emprego no arranjo ( $\text{employment density}_{a,t}$ ), a qual capta os possíveis efeitos das economias de aglomeração sobre os salários locais (produtividade). A área total do arranjo (quilômetros quadrados) é incluída para distinguir os efeitos de densidade dos efeitos de escala. A especificação incluiu como variável de dotação ou atributo local uma *dummy* indicando se o arranjo litorâneo possui uma função turística e uma *dummy* indicando se o arranjo localizado na fronteira do país possui integração com algum município do país vizinho. A proximidade ao litoral pode afetar o salário local como uma amenidade (dotação não produtiva) ou como uma dotação produtiva se indicar a proximidade a portos de escoamento. Por sua vez, a proximidade à fronteira pode ser uma dotação produtiva, se a distância aos potenciais mercados consumidores é reduzida. As macrorregiões foram incluídas para considerar as significantes heterogeneidades existentes entre as mesmas, a exemplo da estrutura produtiva. A região Sudeste foi assumida como a categoria de referência. A especificação incluiu as médias locais dos efeitos fixos de trabalhador ( $\bar{\theta}_{a,t}$ ) e firma ( $\bar{\psi}_{a,t}$ ) estimados no primeiro estágio.

A estratégia de identificação utiliza dados pareados para controlar a heterogeneidade não observada de firma e trabalhador e o respectivo *sorting* sobre os salários e efeitos de aglomeração. A inclusão do efeito fixo do indivíduo permite controlar o viés da habilidade não observada na equação salarial e a auto seleção, associada à atração dos trabalhadores mais habilidosos para os grandes centros urbanos. A inclusão dos efeitos fixos de localização e de firma permite controlar o *sorting* ou auto seleção decorrente da atração e manutenção das firmas mais produtivas nas áreas mais desenvolvidas e lidar com a seleção dos trabalhadores mais qualificados para as firmas mais produtivas. A endogeneidade na densidade do emprego pode surgir se existem choques locais contemporâneos (tecnológicos ou de outros tipos) que podem ampliar os salários e atrair migrantes, gerando uma causalidade reversa da densidade do emprego com os salários locais.

O trabalho propõe utilizar um instrumento tradicional da Economia do Trabalho (BARTIK, 1991; BLANCHARD; KATZ, 1992; CARD, 2001; AUTOR; DUGGAN, 2003; DIAMOND, 2016), o *Bartik IV*, para estimar o crescimento exógeno do emprego em cada arranjo e lidar com a endogeneidade da densidade. O instrumento estima o crescimento no emprego local induzido pelos choques na demanda nacional (emprego nacional), tendo em vista que esses choques são exógenos aos choques locais, se for excluído o próprio emprego local. O principal canal através do qual o instrumento explica a variação na demanda local por trabalho é a mudança na demanda nacional pelos bens produzidos, representada pelo emprego nacional. Nesse sentido, o emprego endógeno local, utilizado na medida de densidade, será instrumentalizado por sua variação exógena, derivada dos choques na demanda nacional.

Na estimação do segundo estágio do modelo serão adotadas duas versões do instrumento *Bartik* para a densidade do emprego local: o instrumento *Bartik* do crescimento setorial do emprego e o *Bartik* do crescimento do emprego no setor público, utilizado em Faggio e Overman (2014). Para o cálculo do crescimento do emprego, o ano de 1991 será assumido

como referência. O ano base busca garantir uma maior exogeneidade do instrumento em relação aos choques contemporâneos nos mercados locais. O instrumento *Bartik* utilizado neste estudo para a densidade do emprego local é definido pela seguinte expressão:

$$Bartik_{lma,t} = \sum_s E_{lma,s,t_0} \left( \frac{E_{(n-lma),s,t} - E_{(n-lma),s,t_0}}{E_{(n-lma),s,t_0}} \right) \quad (5)$$

onde  $E_{lma,s,t_0}$  é o nível de emprego do arranjo populacional (LMA) no setor<sup>4</sup>  $s$  no ano base  $t_0$  (1991) e o termo  $(E_{(n-lma),s,t} - E_{(n-lma),s,t_0})/E_{(n-lma),s,t_0}$  representa o crescimento do emprego nacional ( $n$ ) do setor  $s$  em cada ano  $t$  (2002-2014), a partir do ano base  $t_0$ , excluindo o emprego do local de referência ( $n-lma$ ).

Por sua vez, os efeitos da densidade podem ser subestimados se a diferença entre a área efetivamente ocupada e a extensão territorial oficial for muito alta. Este trabalho propõe adotar uma nova medida para definir a área efetivamente ocupada, a partir da área iluminada de cada arranjo populacional, obtida a partir das imagens de satélites extraídas do *Defense Meteorological Satellite Program - Operational Linescan System* (DMSP-OLS). As imagens são processadas e distribuídas pelo *National Centers for Environmental Information* – NCEI, em uma série anual para o período de 1992 a 2013. O ano de 2013 foi adotado como referência para a extração da luminosidade noturna utilizada na definição da densidade do emprego em cada arranjo populacional, ao longo do período de 2002 a 2014. A área iluminada de cada arranjo será definida pela seguinte expressão:

$$lighted\_area_{a,t} = \delta_{a,2013} \times area_{a,t}, \quad (6)$$

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

onde  $\delta_{a,2013}$  é a razão entre o número de *pixels* iluminados e o número total de *pixels* do arranjo  $a$  no ano de 2013, obtidos a partir das imagens de satélite, e  $area_{a,t}$  é a extensão territorial oficial. Portanto, se todos os *pixels* em um arranjo apresentam luminosidade noturna ( $\delta_{a,2013} = 1$ ), a sua área iluminada será equivalente à área territorial oficial. A área iluminada será usada neste estudo para atenuar a possível subestimação da densidade do emprego, particularmente nos arranjos cuja ocupação humana é concentrada em uma parcela relativamente pequena da sua área oficial total.

### 3 Resultados

Esta seção apresenta as características da amostra e os resultados da estimação do primeiro e segundo estágios do modelo.

#### 3.1 Características da Amostra

A Tabela 3.1 apresenta as estatísticas descritivas das variáveis utilizadas na estimação da equação de salários, correspondente ao primeiro estágio do modelo. A média e o valor máximo do logaritmo do salário real hora ( $lwreal\_hr$ ) é maior nos arranjos do que nos municípios que não pertencem a um arranjo. No entanto, os arranjos possuem maior dispersão salarial (desvio), sugerindo a concentração de trabalhadores de alta e baixa remuneração nessas áreas. Os arranjos possuem maior parcela de trabalhadores com escolaridade igual ou superior ao ensino médio completo (67,7%) do que os municípios que não pertencem a um arranjo (50,5%). Portanto, os arranjos possuem maior salário real médio, ao mesmo tempo

<sup>4</sup> Os setores foram definidos a partir da classificação de subsetores do IBGE, disponível na RAIS desde 1985. Utilizou-se a divisão dos setores em 2 dígitos, correspondendo a 25 subsetores de atividade econômica.



em que concentram os trabalhadores mais qualificados. Essa relação sugere a necessidade de explicar a maior produtividade dos arranjos, condicional às heterogeneidades individuais. As firmas dos arranjos populacionais apresentam um tamanho 4% maior em relação à média do seu setor, enquanto as firmas fora dos arranjos possuem tamanho 26% menor. Essa última evidência sugere a importância de investigar o prêmio urbano condicional às heterogeneidades das firmas, particularmente se as firmas maiores são as mais produtivas.

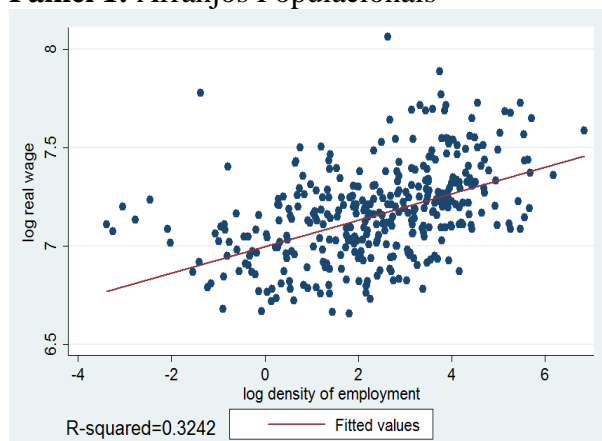
**Tabela 3.1 – Estatísticas descritivas das variáveis do primeiro estágio (2002-2014)**

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

Fonte: Elaboração própria, a partir da amostra da RAIS (2002-2014): 6.048.393 observações.

A Figura 3.1 apresenta a relação entre os salários médios reais de 2002 a 2014, a preços de 2014, e a densidade do emprego nos arranjos populacionais. O ajuste linear sugere salários crescentes quanto maior a densidade.

**Painel 1: Arranjos Populacionais**



**Painel 2: Grandes Arranjos Populacionais**



Fonte: Elaboração própria a partir da RAIS (2002-2014). Nota: Salário real deflacionado pelo IPCA (base: dezembro/2014).

Figura 3.1 – Salário real e densidade do emprego nos arranjos (2002–2014)

A variação salarial dos arranjos explicada pela densidade do emprego foi de 32%, sendo 30% apenas nos grandes arranjos. Por outro lado, a estimativa da elasticidade dos salários em relação à densidade do emprego nos arranjos populacionais do Brasil foi estimada em 6,7%, para o período de 2002 a 2014. Combes, Duranton e Gobillon (2008) encontraram uma elasticidade de 4,9% para a França.

A Tabela 3.2 apresenta a parcela de trabalhadores da amostra que mudam de firma e de arranjo, a parcela de firmas que mudam de arranjo e a distribuição dos trabalhadores na amostra por tamanho do arranjo e setores produtivos, excluindo o setor público.

**Tabela 3.2 – Distribuição dos trabalhadores por tamanho do arranjo populacional e setor e mobilidade dos trabalhadores e firmas**

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

Fonte: Elaboração própria, a partir da amostra da RAIS (2002-2014). Nota: *Labor Market Areas* – LMA.

Os arranjos concentram 86,6% dos trabalhadores da amostra, dos quais 59% estão nos grandes arranjos. Essa proporção é compatível com a distribuição de empregos da RAIS. Os municípios que não pertencem a um arranjo foram assumidos como a categoria de referência na estimação dos efeitos de localização. A indústria de transformação concentra a maior parcela de trabalhadores da amostra (30,7%) e será a categoria de referência setorial. A estratégia de identificação deste estudo requer mobilidade dos trabalhadores entre firmas e arranjos e das firmas entre arranjos, ao longo do painel. A tabela mostra que as condições de identificação são satisfeitas para o estudo. Portanto, a amostra permite identificar os três efeitos fixos simultaneamente, do trabalhador, da firma e do arranjo.

### 3.2 Prêmio Salarial Urbano dos Arranjos Populacionais

Essa seção apresenta os resultados da estimação econométrica da equação de salários correspondente ao primeiro estágio do modelo. As evidências encontradas mostram um prêmio salarial urbano médio de 24% (POLS) para os trabalhadores dos arranjos

populacionais, ao considerar as características observadas dos trabalhadores e do emprego. Ao considerar simultaneamente as heterogeneidades não observadas dos trabalhadores e firmas o prêmio urbano dos arranjos desaparece. Esse resultado preliminar sugere que uma grande parcela do prêmio urbano seria explicada pelas características não observadas específicas dos indivíduos e das firmas localizados nos arranjos. Como robustez, os resultados foram estimados com uma amostra que não realizou qualquer seleção entre as firmas que se movem entre arranjos e com uma amostra que exclui os *outliers* de salário<sup>5</sup>. Novamente, o prêmio desapareceu após o controle dos efeitos individuais e de firma, variando entre 26% (sem seleção de firmas) e 22% (sem *outliers*), sugerindo a robustez dos resultados. Foi encontrado um prêmio salarial urbano de 29,4% para os grandes arranjos, de 17% para os médios e de 7,4% para os arranjos pequenos (POLS).

O prêmio desaparece para os três grupos de arranjos após o controle da heterogeneidade não observada das firmas e trabalhadores simultaneamente. A inclusão dos efeitos fixos de cada um dos 365 arranjos separadamente em substituição à *dummy* de arranjo ou de tamanho propõe considerar a substancial heterogeneidade existente entre os diversos mercados de trabalho locais do Brasil na identificação dos respectivos prêmios urbanos. A Tabela 3.3 apresenta as estimativas dos efeitos de localização nos grandes arranjos, por ordem de tamanho populacional (2010), obtidas para as diferentes especificações da equação (3). A coluna I mostra os resultados da estimação da equação salarial pelo método *POLS*, que inclui as características observadas dos trabalhadores e do emprego, os efeitos de localização em cada um dos 365 arranjos e as *dummies* de tempo.

A coluna IV (FE<sub>ij</sub>) apresenta as estimativas de Efeitos Fixos obtidas através da transformação *within group* para cada combinação (*spell*) existentes entre as firmas e o trabalhador (*i, j*), a qual representa o termo de *matching*. Dentro de cada *spell*, os efeitos fixos de firma e indivíduo não variam, de modo que a transformação vai eliminar tais efeitos. Neste sentido, quaisquer variáveis que sejam constantes dentro de cada *spell* não serão identificadas (ANDREWS; SCHANK; UPWARD, 2006). As especificações condicionais propostas no modelo AKM adotam a primeira diferença ao invés de desvios da média para eliminar os efeitos fixos na estimação. A coluna II apresenta os resultados da estimação condicional (COND), quando são controlados (*absorbed*) apenas os efeitos fixos dos indivíduos (*i*). A coluna III apresenta o modelo que inclui apenas os efeitos de firmas (*j*). Ambos os modelos, incluem os efeitos fixos de localização e as características observadas dos trabalhadores. Por fim, a coluna V apresenta a estimação condicional, que inclui simultaneamente os efeitos fixos da firma e do trabalhador (*absorbed*), além dos efeitos de localização. Essa coluna representa o modelo condicional com efeito fixo de trabalhador sendo estimado primeiro.

A variação salarial explicada pelo modelo ( $R^2$  ajustado) foi de 60,6%, considerando somente as características observadas do trabalhador e do emprego (POLS). A inclusão da heterogeneidade individual não observada aumenta substancialmente a parcela da variação de salários individuais explicada para 91,7% (COND<sub>i</sub>). Ao incluir apenas os efeitos de firma o modelo explica 78% da variação salarial (COND<sub>j</sub>). Os resultados sugerem que o efeito dos indivíduos é mais relevante para explicar a variação dos salários individuais do que o efeito das firmas. A inclusão de ambos os efeitos responde por 93% da variação dos salários reais (COND<sub>ij</sub>), o que sugere a importância das heterogeneidades não observadas.

---

<sup>5</sup> Foram excluídos os extremos da distribuição do salário real hora, correspondendo aos valores 1% maiores e menores.

**Tabela 3.3 – Efeitos de localização nos grandes arranjos sobre o salário-hora real: primeiro estágio**

<i>ln (w real hora)</i>	(I) POLS	(II) COND i	(III) COND j	(IV) FE (i, j)	(V) COND(i,j)
Sorocaba	0,3381***	0,0211	-0,0052	-0,0697***	-0,0649**
Ribeirão Preto	0,2468***	0,0302**	-0,0038	-0,0142	-0,0126
Campo Grande	0,0853***	0,0053	0,0307*	-0,0078	-0,0118
Cuiabá	0,1343***	0,0197	0,0139	-0,0033	0,0115
Florianópolis	0,2888***	0,0197	0,1717***	0,0852***	0,0882***
Aracaju	-0,0605***	-0,0069	-0,0020	0,0166	0,0253
São José dos Campos	0,4181***	0,0255**	0,0361*	-0,0322***	-0,0276
João Pessoa	-0,1755***	-0,0157	0,0689	0,0448***	0,0376
Teresina	-0,1341***	0,0359*	-0,0098	-0,0059	0,0036
Maceió	-0,0579***	-0,0029	0,0415	0,0538***	0,0504
Natal	-0,1173***	-0,0236	0,0071	-0,0055	-0,0084
São Luís	-0,0350***	0,0325	0,0424	-0,0213	0,0033
Vitória	0,2162***	0,0468***	0,0281	0,0261***	0,0378
Baixada Santista	0,4178***	0,0574***	0,0156	-0,0123**	-0,0072
Manaus	0,2714***	0,0711**	-0,1198*	-0,0245***	-0,0198
Campinas	0,4381***	0,0317***	0,0275	-0,0430***	-0,0334**
Belém	0,1078**	0,0478***	0,0539	0,0956***	0,0912***
Goiânia	0,1203***	0,0277	0,0331	-0,0091	-0,0089
Curitiba	0,2859***	0,0535***	0,0326	-0,0063	0,0055
Porto Alegre	0,3491***	0,0258***	0,0359	-0,0170***	-0,0150
Fortaleza	-0,0538***	0,0020	-0,1061***	-0,0190***	-0,0174
Brasília	0,3922***	0,0916***	0,0759***	-0,0121***	0,0020
Salvador	0,1137***	0,0054	0,0591**	-0,0196***	-0,0053
Recife	-0,0800***	0,0252**	-0,0217	-0,0265***	-0,0289*
Belo Horizonte	0,2344***	0,0112	0,0071	-0,0026	0,0034
Rio de Janeiro	0,2778***	0,0829***	0,0540**	-0,0045	-0,0007
São Paulo	0,4239***	0,0427***	0,0331**	-0,0238***	-0,0160
Observações (N x T)	6.048.393	6.048.393	6.048.393	6.048.393	6.048.393
R <sup>2</sup>	0,606	0,923	0,794	0,113	0,943
R <sup>2</sup> Ajustado	0,606	0,917	0,784		0,935
Within R <sup>2</sup>		0,437	0,311	0,379	0,360

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

Notas: Erro padrão entre parênteses; \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

O prêmio urbano é estimado pelo método POLS a partir dos efeitos fixos de cada um dos 362 arranjos (LMAs)<sup>6</sup>. Quando o efeito fixo é positivo e estatisticamente significativo, o arranjo apresenta um prêmio urbano, ou seja, existem ganhos de aglomeração sobre os salários dos trabalhadores locais. Os resultados sugerem a existência de um prêmio urbano sobre os salários dos 19 dentre os 27 grandes arranjos populacionais (POLS), cujas maiores magnitudes estão concentradas nas macrorregiões do Sul e Sudeste. O maior prêmio foi encontrado no arranjo de Campinas (43,8%) e o menor no arranjo de Campo Grande (8,5%). Na região Nordeste, uma das mais pobres do país, apenas o arranjo de Salvador apresentou um prêmio salarial urbano (11,4%).

Por sua vez, quando os atributos não observados dos trabalhadores e firmas são considerados (CONDij), apenas 2 arranjos apresentaram um prêmio salarial urbano, embora com magnitudes menores: Belém (9,1%) e Florianópolis (8,8%). Nesse sentido, o *sorting* espacial dos trabalhadores e firmas parece explicar uma grande parcela dos ganhos atribuídos diretamente às economias de aglomeração nos mercados de trabalho locais. Ao controlar apenas os efeitos fixos individuais na equação de salários (CONDi), ainda permaneceu um prêmio urbano, em relação ao POLS, em 12 grandes arranjos, porém em menor magnitude. Se

<sup>6</sup> A amostragem não foi representativa para três arranjos populacionais pequenos.

o efeito de localização torna-se positivo (prêmio) ao eliminar os efeitos individuais, há indícios de que os atributos não observados dos indivíduos seriam negativamente correlacionados com os salários locais. Essa evidência pode sugerir que os trabalhadores do local não possuem as características produtivas melhor remuneradas no mercado de trabalho local, ou seja, o local pode concentrar trabalhadores de baixa habilidade produtiva.

Quando apenas os efeitos das firmas são controlados (CONDj), 7 arranjos ainda mantiveram o prêmio urbano, em menor magnitude em relação aos POLS. Os resultados podem sugerir que alguns mercados de trabalho locais concentram trabalhadores de maior habilidade enquanto outros concentram firmas mais produtivas. Por outro lado, existem arranjos que apresentam um prêmio urbano ao controlar os efeitos da firma ou do trabalhador, mas o prêmio desaparece após o controle de ambos. Nesses casos, é possível que o prêmio resulte da melhor correspondência (*matching*) entre as habilidades individuais e as necessidades produtivas das firmas locais. A figura 3.2 resume os resultados da tabela 3.3.

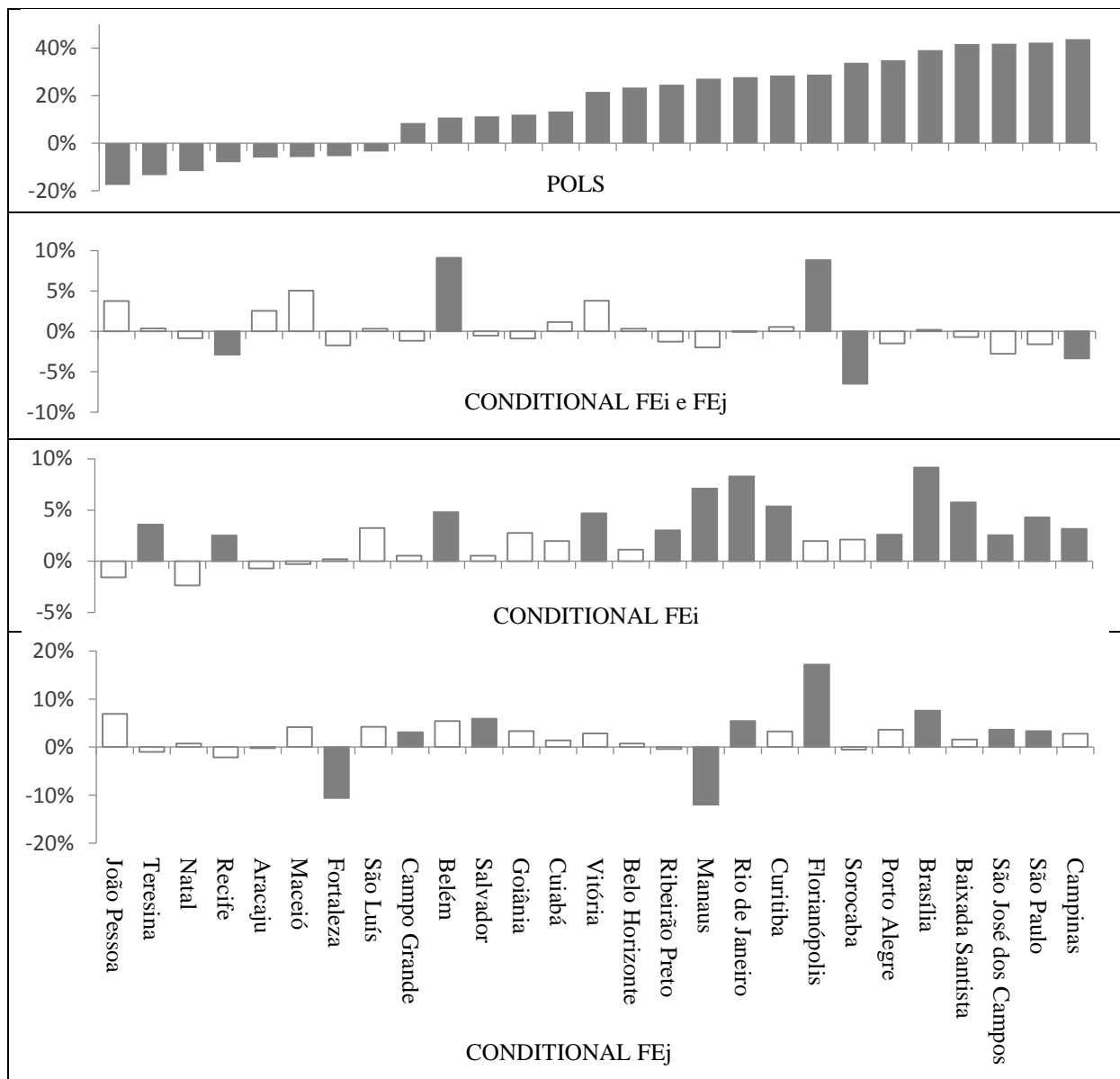


Figura 3.2 – Prêmio salarial urbano para os grandes arranjos populacionais do Brasil

Fonte: Elaboração própria a partir da RAIS (2002-2014). Colunas preenchidas na cor cinza representam os efeitos de localização estatisticamente significantes (categoria de referência: municípios fora dos arranjos).

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

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

### **3.3 Decomposição dos Efeitos de Localização**

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

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

O modelo POLS foi estimado sem a inclusão dos componentes não observados dos trabalhadores e firmas apenas para a comparação com os resultados da literatura. A primeira modificação deste estudo é a inclusão dos componentes não observados médios dos trabalhadores e firmas na decomposição dos salários locais. A segunda modificação é à estimação de um modelo de variável instrumental conjuntamente com os componentes não observados médios. O terceiro modelo utiliza a área iluminada dos arranjos para medir a densidade. Por fim, a especificação principal deste estudo estima o modelo de decomposição utilizando uma variável instrumental para a densidade, a área iluminada em substituição à área total e os componentes não observados.

A Tabela 3.4 apresenta os resultados da decomposição dos efeitos fixos dos 362 arranjos, estimados a partir da equação (4) do segundo estágio do modelo. A decomposição dos efeitos fixos dos arranjos inclui os componentes associados às características locais e às heterogeneidades não observadas dos trabalhadores e firmas, ao longo do painel (2002–2014). A coluna 1 apresenta os resultados do modelo estimado apenas com as variáveis observadas dos arranjos (*baseline*). As colunas 2 e 3 incluem as médias dos efeitos não observados das firmas e dos trabalhadores, respectivamente, enquanto a coluna 4 inclui ambos os efeitos. As colunas de 1 a 4 foram estimadas através do método POLS. O método FGLS foi utilizado na coluna 5 para realizar a correção da possível heterocedasticidade existente na variância dos efeitos de localização, estimados no primeiro estágio. A coluna 6 permite corrigir os erros-padrão, através do método de *Bootstrap* (POLS-BS), para lidar com eventual inclusão de resíduos da estimação dos efeitos individuais e das firmas no primeiro estágio. A parcela da variação dos salários locais explicada pelos componentes observados foi de 50,2%. Após incluir a média dos atributos não observados das firmas, o modelo ampliou marginalmente a variação salarial explicada para 50,9% (Coluna 2). A inclusão dos atributos não observados médios dos indivíduos elevou a parcela da variação explicada para 60% (Coluna 3). A inclusão dos efeitos médios do trabalhador e da firma, simultaneamente, manteve a variação explicada em 60% (Coluna 4).

**Tabela 3.4 – Decomposição dos efeitos de localização dos arranjos: segundo estágio**

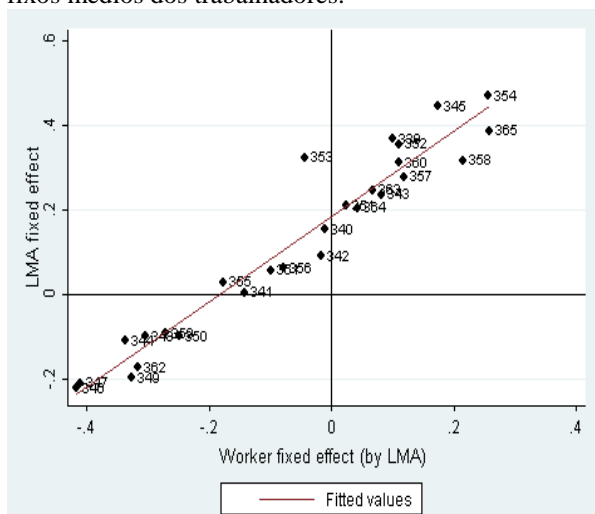
FE_LMA	(1)	(2)	(3)	(4)	(5)	(6)
	POLS	POLS	POLS	POLS	FGLS	POLS-BS
<i>ldensemp_lma</i>	0,0421*** (0,0105)	0,0417*** (0,0104)	0,0258* (0,0156)	0,0259* (0,0155)	0,0256*** (0,0006)	0,0365** (0,0147)
<i>larea_lma</i>	0,0214 (0,0152)	0,0204 (0,0150)	0,0084 (0,0154)	0,0083 (0,0154)	0,0057*** (0,0008)	0,0656*** (0,0188)
<i>(mean) i</i>			0,2536 (0,1800)	0,2503 (0,1793)	0,2825*** (0,0045)	0,0053 (0,1300)
<i>(mean) f</i>		0,0336* (0,0196)		0,0079 (0,0063)	0,0039*** (0,0011)	0,0012 (0,0426)
<i>Frontier</i>	-0,0764* (0,0455)	-0,0770* (0,0449)	-0,0820** (0,0413)	-0,0821** (0,0412)	-0,0712*** (0,0023)	-0,0983* (0,0593)
<i>Touristic</i>	-0,0072 (0,0389)	-0,0074 (0,0390)	0,0053 (0,0267)	0,0051 (0,0269)	0,0233*** (0,0031)	0,0364 (0,0465)
<i>dummies de região</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>Constante</i>	-0,0991 (0,1159)	-0,0884 (0,1159)	0,0637 (0,1569)	0,0641 (0,1571)	0,0887*** (0,0075)	-0,5040*** (0,1845)
Observações ( <i>LMA x T</i> )	4681	4681	4681	4681	4681	4681
<i>R</i> <sup>2</sup> Ajustado	0,502	0,509	0,600	0,600		0,225

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

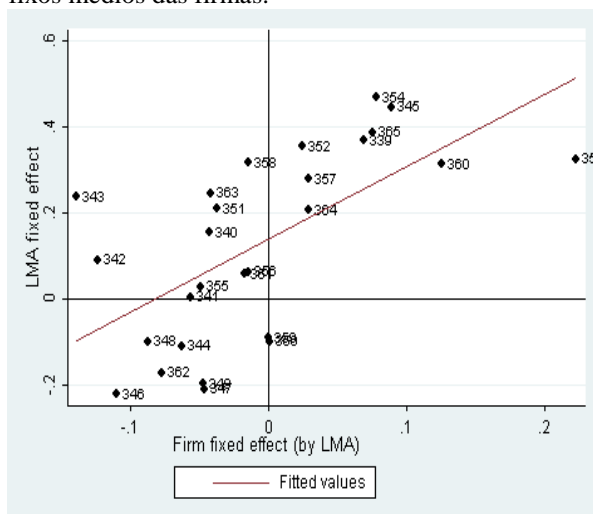
Notas: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ ; Erro-padrão robusto com *clusters* de LMAs; Colunas 1-5 assumem peso analítico proporcional ao número de indivíduos em cada LMA.

Assim como no modelo de determinação de salários individuais, os atributos individuais não observados se mostraram mais importantes para explicar o diferencial salarial local do que os atributos não observados das firmas. A Figura 3.3 mostra graficamente o poder explicativo de cada componente. Ambos os efeitos apresentam uma correlação positiva e crescente com os efeitos de localização ou salários locais, mas os efeitos dos trabalhadores apresentam uma relação um pouco mais forte, com base na inclinação da reta.

**Painel 1:** Efeitos fixos dos grandes arranjos e efeitos fixos médios dos trabalhadores.



**Painel 2:** Efeitos fixos dos grandes arranjos e efeitos fixos médios das firmas.



Fonte: Elaboração própria.

Figura 3.3 – Correlação entre os efeitos de localização dos grandes arranjos e os efeitos da firma e do trabalhador

Para considerar características observadas dos arranjos, o modelo incluiu *dummies* para arranjos litorâneos do tipo turístico (*touristic*) ou para arranjos com municípios na fronteira (*frontier*). Os resultados mostraram um efeito negativo nos salários médios dos arranjos com integração fronteiriça, sugerindo que não existem vantagens produtivas de facilidade de acesso aos mercados vizinhos, ou tais vantagens não são aproveitadas. Por sua vez, os arranjos litorâneos com função turística apresentaram um ganho salarial (0,023), somente no modelo FGLS. Os ganhos salariais nos arranjos turísticos podem ser gerados não somente da renda do setor turístico, mas também da localização privilegiada em relação a rios navegáveis e portos atuando como dotação produtiva, a exemplo do arranjo de Itajaí em Santa Catarina.

Os efeitos de área (efeitos de escala) variaram entre 0,6% e 6,6%, mas somente se mostraram estatisticamente significantes nos modelos FGLS e *Bootstrap*. A magnitude do efeito de área corresponde, aproximadamente, à metade do efeito da densidade, exceto no modelo *Bootstrap* cujo efeito passa a ser quase o dobro da densidade. Combes, Duranton e Gobillon (2008) encontraram um efeito de área que correspondeu a quase um terço do efeito da densidade, variando entre 1,1% e 2,2%, nas áreas de emprego da França. Os resultados sugerem que um aumento na população através do aumento da densidade tem um efeito maior sobre a produtividade local do que o aumento da população decorrente da maior área disponível, mantendo a densidade constante. Os efeitos da área representam um limite à expansão urbana e, portanto, podem refletir em alguma medida o preço da terra. Nesse caso, o efeito positivo sobre os salários podem sugerir que a maior disponibilidade de terra para a produção implica em maior produtividade local, em parte porque o custo da terra será menor.

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



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

A área iluminada corresponde à proporção de *pixels* iluminados, em cada arranjo, multiplicada por sua área oficial do ano de 2013. A Tabela 3.5 apresenta as estatísticas de iluminação do Brasil e dos arranjos. A partir das imagens de satélite (NCEI, 2013), é possível constatar que apenas 8,1% da área total do Brasil apresentou alguma luminosidade noturna no ano de 2013. A parcela iluminada da área total dos arranjos em 2013 foi de 15,2%, correspondendo a quase o dobro da parcela iluminada da área total do Brasil. A parcela de luminosidade noturna variou de 9,5%, entre os pequenos, para 55,2% entre os grandes arranjos. Os dados são consistentes com a distribuição da luminosidade no Brasil e em outros países, encontrada em estudos da literatura internacional (HENDERSON; STOREYGARD; WEIL, 2012).

**Tabela 3.5 – Distribuição da área oficial e da área iluminada no Brasil: 2007 e 2013**

<b>Categoria</b>	<b>Área (km<sup>2</sup>)</b>	<b>Sharelight 2007</b>	<b>Sharelight 2013</b>	<b>Light 2007</b>	<b>Light 2013</b>
Grandes	136.496,79	0,5442	0,5524	74.108,66	75.227,20
Médios	1.040.496,44	0,1205	0,1292	125.207,87	134.332,89
Pequenos	534.807,57	0,0900	0,0954	48.170,43	51.062,52
<b>Arranjos</b>	<b>1.711.800,80</b>	<b>0,1448</b>	<b>0,1524</b>	<b>247.486,96</b>	<b>260.622,62</b>
<b>Outros municípios</b>	<b>6.791.005,35</b>	<b>0,0568</b>	<b>0,0634</b>	<b>385.737,78</b>	<b>430.677,42</b>
<b>Brasil</b>	<b>8.502.806,15</b>	<b>0,0745</b>	<b>0,0813</b>	<b>633.224,74</b>	<b>691.300,04</b>

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

Notas: *Sharelight* é a parcela de *pixels* luminosos em cada área e *light* é a área iluminada (km<sup>2</sup>) obtida a partir do *share*.

Os dados mostram que existe uma grande diferença entre as extensões territoriais oficiais e a distribuição populacional e urbana, a partir das imagens de luminosidade noturna, o que motiva o uso da área iluminada como medida mais adequada para calcular a densidade urbana. Nesse sentido, a área com luminosidade noturna foi utilizada para calcular a densidade do emprego nos arranjos e estimar os respectivos efeitos das economias de aglomeração urbana. Se a área administrativa oficial subestima a densidade e possivelmente os seus efeitos sobre os salários locais, é esperado um aumento nos efeitos das economias de aglomeração com o uso da nova medida de área. A Tabela 3.6 apresenta os resultados da decomposição dos efeitos de localização após substituir a área oficial pela área iluminada dos arranjos (2013) no cálculo da densidade do emprego.

Os resultados mostraram uma elevação na magnitude dos efeitos da densidade sobre os salários locais. Os efeitos puros da aglomeração (densidade) variaram entre 2,9% e 5,4%, com o uso da área iluminada, enquanto os resultados anteriores, calculados a partir da área oficial total, variaram entre 2,6% e 4,2%. Portanto, como era previsto, o uso da área oficial parece subestimar os efeitos da aglomeração urbana. Por sua vez, a densidade do emprego é endógena aos salários locais, já que existe uma relação de causalidade reversa entre os salários locais e a densidade urbana. A endogeneidade resulta dos choques locais que afetam os salários reais, os quais motivam as decisões de localização e, portanto, a densidade. Se existem choques sobre a produtividade local, a exemplo dos choques tecnológicos, é possível que a densidade esteja capturando parte dos efeitos desses choques. A literatura em Economia

Urbana utiliza variáveis instrumentais para lidar com a endogeneidade da densidade, a exemplo da defasagem temporal da população.

**Tabela 3.6 – Decomposição dos efeitos de localização dos arranjos:  
segundo estágio com a área iluminada (2013)**

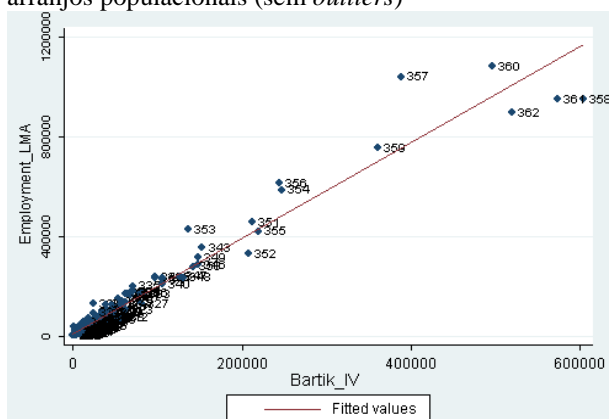
FE_LMA	(1) POLS	(2) POLS	(3) POLS	(4) POLS	(5) FGLS	(6) POLS-BS
<i>ldensemplight13</i>	0,0543*** (0,0182)	0,0539*** (0,0179)	0,0306 (0,0244)	0,0308 (0,0243)	0,0290*** (0,0009)	0,0533** (0,0211)
<i>larea_lma</i>	0,0095 (0,0178)	0,0085 (0,0175)	0,0023 (0,0148)	0,0021 (0,0148)	-0,0007 (0,0008)	0,0445*** (0,0150)
<i>(mean) f</i>		0,0345* (0,0202)		0,0077 (0,0060)	0,0040*** (0,0011)	0,0013 (0,0419)
<i>(mean) i</i>			0,2612 (0,1840)	0,2580 (0,1833)	0,2998*** (0,0046)	0,0053 (0,1285)
<i>Frontier</i>	-0,1298** (0,0517)	-0,1298** (0,0508)	-0,1163*** (0,0447)	-0,1164*** (0,0446)	-0,1104*** (0,0023)	-0,1353** (0,0591)
<i>Touristic</i>	0,0078 (0,0410)	0,0074 (0,0410)	0,0142 (0,0268)	0,0140 (0,0269)	0,0285*** (0,0031)	0,0542 (0,0456)
<i>dummies de região</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim
<i>Constante</i>	-0,0760 (0,1126)	-0,0653 (0,1124)	0,0819 (0,1505)	0,0824 (0,1507)	0,1151*** (0,0073)	-0,4442*** (0,1666)
Observações	4681	4681	4681	4681	4681	4681
$R^2$	0,491	0,498	0,595	0,595		0,220
$R^2$ Ajustado	0,489	0,496	0,593	0,593		0,216

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

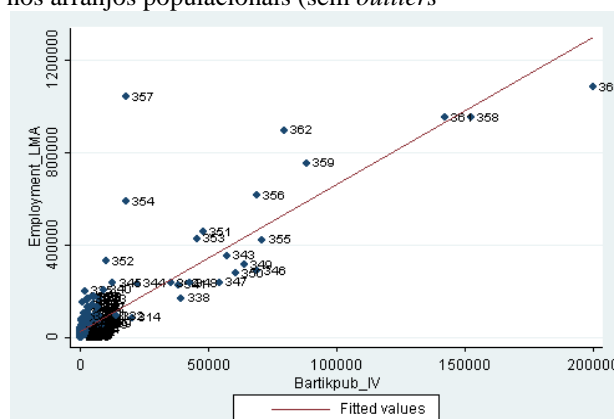
Notas: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ ; Erro-padrão robusto com *clusters* de LMAs; Colunas 1-5 assumem peso analítico proporcional ao número de indivíduos em cada LMA.

Este trabalho fará uso de uma variável instrumental tradicional na Economia do Trabalho, o *Bartik IV* setorial, originalmente concebido por Bartik (1991). O *Bartik IV* estima qual seria o crescimento setorial previsto a partir do crescimento nacional do respectivo setor, baseado na parcela inicial do emprego setorial local. Nesse sentido, o crescimento adicional do emprego local em relação ao crescimento previsto, a partir da sua estrutura setorial, seria explicado pelos choques locais. O crescimento exógeno da demanda nacional é utilizado para estimar a demanda por emprego local, após a exclusão do emprego do próprio local no cálculo do crescimento do emprego nacional. O estudo utilizará duas versões do instrumento *Bartik*. A primeira versão do instrumento estima o crescimento do emprego local a partir do crescimento do emprego setorial nacional e a segunda estima o crescimento do emprego local a partir do crescimento nacional do emprego público. A Figura 3.4 apresenta um ajuste linear entre o emprego médio observado nos arranjos e o emprego estimado pelo *Bartik IV*. É possível perceber um melhor ajuste linear do emprego estimado pelo *Bartik* setorial (Painel 1) do que o emprego previsto pelo *Bartik* do setor público (Painel 2). Portanto, a análise dos resultados do modelo de variável instrumental deve focar sobre o *Bartik* setorial.

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



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



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

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

Figura 3.4 – Emprego médio observado e estimado (*Bartik IV*) dos arranjos: 2002-2014

A Tabela 3.7 apresenta os resultados da decomposição utilizando o instrumento *Bartik* para a densidade do emprego. A equação de decomposição (4) foi estimada em dois estágios simultâneos pelo método 2SLS.

**Tabela 3.7 – Decomposição dos efeitos de localização dos arranjos (segundo estágio):**  
*Bartik IV* para a densidade do emprego

FE_LMA	(1) POLS	(2) 2SLS ( <i>lbartik</i> )	(3) POLS	(4) 2SLS ( <i>lbartik</i> )	(5) POLS	(6) POLS	(7) 2SLS ( <i>lbartikpub</i> )
<i>ldensemp_lma</i>	0,0423*** (0,0105)	0,0249** (0,0119)	0,0261* (0,0156)	0,0264** (0,0120)	0,0436*** (0,0105)	0,0273* (0,0158)	0,0285** (0,0124)
<i>larea_lma</i>	0,0212 (0,0153)	0,0571*** (0,0140)	0,0082 (0,0154)	0,0603*** (0,0138)	0,0197 (0,0152)	0,0071 (0,0152)	0,0544*** (0,0137)
( <i>mean</i> ) <i>i</i>			0,2499 (0,1792)	0,0052*** (0,0009)		0,2478 (0,1784)	0,0052*** (0,0009)
( <i>mean</i> ) <i>f</i>			0,0079 (0,0063)	0,0012* (0,0007)		0,0081 (0,0063)	0,0012* (0,0007)
<i>Frontier</i>	-0,0754* (0,0455)	-0,1214** (0,0504)	-0,0816** (0,0413)	-0,1224** (0,0511)	-0,0763* (0,0454)	-0,0829** (0,0414)	-0,1457*** (0,0452)
<i>touristic</i>	-0,0069 (0,0389)	0,0303 (0,0484)	0,0053 (0,0269)	0,0336 (0,0473)	-0,0112 (0,0393)	0,0014 (0,0272)	0,0118 (0,0469)
<i>dummies região</i>	Sim	Sim	Sim	Sim	Sim	Sim	Sim
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim	Sim
<i>constante</i>	-0,0990 (0,1162)	-0,4160*** (0,1191)	0,0641 (0,1573)	-0,4401*** (0,1171)	-0,0939 (0,1160)	0,0668 (0,1562)	-0,4027*** (0,1189)
Observações	4623	4623	4623	4623	4511	4511	4511
<i>R</i> <sup>2</sup>	0,505	0,213	0,602	0,235	0,508	0,604	0,251
<i>R</i> <sup>2</sup> Ajustado	0,503	0,210	0,601	0,231	0,506	0,602	0,247
<b>Estatísticas do primeiro estágio - 2SLS</b>							
<i>Bartik IV</i>		0,7923***		0,7911***			0,6557***
F(20,356)		218,73					
F(22,356)				200,71			
F(22,347)							186,85
Prob. > F		0,0000		0,0000			0,0000
<i>R</i> <sup>2</sup> Ajustado		0,915		0,916			0,736

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

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

Para permitir a comparação entre os modelos, foi necessário excluir os arranjos que apresentaram uma estimativa negativa para o número de empregos (2), os arranjos correspondentes aos municípios isolados recém-criados (4), os quais não possuíam emprego em 1991, e os arranjos que não apresentaram emprego no setor público (8) em 1991. Esses arranjos geram estimativas negativas ou de valor zero para o emprego e, portanto, não permitem a aplicação do logaritmo. As estatísticas do primeiro estágio, no final da tabela, mostram que ambas as versões do *Bartik IV* apresentam correlação estatisticamente significativa com a densidade do emprego. O instrumento, conjuntamente com as demais covariadas exógenas do modelo, apresenta um alto poder explicativo para a variação na densidade do emprego ( $R^2$ ). Portanto, os supostos da existência de covariância significativa com a variável endógena e da exogeneidade, com base no crescimento nacional, garantem a validade da identificação do modelo. As colunas 1 e 2 mostram os modelos POLS e 2SLS, respectivamente, sem a inclusão dos componentes não observados de indivíduos e firmas. A estimativa do efeito de aglomeração foi reduzida de 4,2% para 2,5%, sugerindo que a endogeneidade tem forte impacto sobre a identificação do modelo.

Por sua vez, ao incluir os efeitos não observados das firmas e indivíduos, o efeito da densidade muda apenas marginalmente, em ambas as versões do *Bartik IV* (Colunas 4 e 7), em relação ao modelo POLS (Colunas 3 e 6). O *Bartik* setorial encontrou um efeito de 2,64% para a densidade do emprego, enquanto o modelo endógeno mostrou um efeito de 2,61%. Portanto, a endogeneidade não comprometeu a identificação dos efeitos de aglomeração sobre a produtividade local nos arranjos populacionais, uma vez que os componentes não observados são incluídos no modelo. Se os efeitos da firma e do trabalhador já estiverem controlando, em alguma medida, os possíveis efeitos dos choques locais, isso poderia justificar os aumentos apenas marginais no efeito da densidade com o uso do instrumento. Esses resultados sugerem que a endogeneidade da densidade pode ser explicada pelo *sorting* espacial dos trabalhadores e das firmas com heterogêneas distribuições de produtividades.

A Tabela 3.8 apresenta os resultados da decomposição salarial com a densidade da área iluminada sendo instrumentalizada pelo *Bartik IV*. As estatísticas do primeiro estágio, apresentadas no final da tabela, mostram que o *Bartik IV* atende ao suposto de covariância parcial com a densidade da área iluminada. As colunas 1 e 2 da tabela mostram os resultados da estimação dos modelos POLS e 2SLS sem a inclusão dos componentes não observados das firmas e trabalhadores. Os resultados mostram uma redução de 5,4% para 4,7% na estimativa do efeito da densidade da área iluminada, entre os modelos POLS e 2SLS. Ao incluir os componentes não observados, as estimativas apresentam um crescimento de 3,1% para 4,9%, entre os modelos POLS e 2SLS (Colunas 3 e 4).

Por sua vez, comparando os modelos 2SLS com e sem a inclusão dos componentes não observados, é possível perceber estimativas muito próximas do efeito da densidade da área iluminada (4,9% e 4,7%), sugerindo novamente que uma parte da endogeneidade resulta do *sorting* espacial das heterogeneidades não observadas. Combes, Duranton e Gobillon (2008) utilizaram defasagens temporais da densidade e do mercado potencial, além de outras variáveis históricas, como instrumentos para a densidade do emprego nos mercados de trabalho locais da França. Os resultados mostraram uma mudança marginal nos efeitos da densidade, os quais foram reduzidos de 3,2% para 3% após o uso dos instrumentos. Ao controlar os efeitos individuais na equação salarial do primeiro estágio, os efeitos da densidade foram reduzidos para 2,9% no modelo de variável instrumental.

**Tabela 3.8 – Decomposição dos efeitos de localização dos arranjos (segundo estágio):  
Bartik IV e densidade do emprego da área iluminada (2013)**

FE_LMA	(1) POLS	(2) 2SLS ( <i>lbartik</i> )	(3) POLS	(4) 2SLS ( <i>lbartik</i> )	(5) POLS	(6) POLS	(7) 2SLS ( <i>lbartikpub</i> )
<i>ldensemplight13</i>	0,0545*** (0,0182)	0,0470** (0,0222)	0,0310 (0,0243)	0,0499** (0,0222)	0,0560*** (0,0183)	0,0323 (0,0246)	0,0530** (0,0235)
<i>larea_lma</i>	0,0094 (0,0179)	0,0408*** (0,0137)	0,0020 (0,0148)	0,0431*** (0,0136)	0,0078 (0,0180)	0,0008 (0,0148)	0,0364*** (0,0131)
(mean) <i>i</i>			0,2578 (0,1832)	0,0053*** (0,0009)		0,2561 (0,1827)	0,0053*** (0,0009)
(mean) <i>f</i>			0,0077 (0,0060)	0,0013* (0,0007)		0,0079 (0,0061)	0,0013* (0,0007)
<i>Frontier</i>	-0,1290** (0,0517)	-0,1371*** (0,0477)	-0,1161*** (0,0447)	-0,1390*** (0,0485)	-0,1316** (0,0519)	-0,1189*** (0,0450)	-0,1627*** (0,0422)
<i>touristic</i>	0,0081 (0,0410)	0,0473 (0,0498)	0,0142 (0,0269)	0,0518 (0,0487)	0,0044 (0,0414)	0,0109 (0,0271)	0,0297 (0,0481)
<i>dummies de região</i>	Sim	Sim	Sim	Sim	Sim	Sim	Sim
<i>dummies de tempo</i>	Sim	Sim	Sim	Sim	Sim	Sim	Sim
<i>Constante</i>	-0,0759 (0,1129)	-0,3995*** (0,1150)	0,0825 (0,1509)	-0,4235*** (0,1130)	-0,0707 (0,1127)	0,0855 (0,1499)	-0,3866*** (0,1160)
Observações	4623	4623	4623	4623	4511	4511	4511
$R^2$	0,491	0,204	0,596	0,227	0,494	0,597	0,237
$R^2$ Ajustado	0,489	0,201	0,594	0,223	0,491	0,595	0,233
<b>Estatísticas do primeiro estágio 2SLS</b>							
<i>Bartik IV</i>		0,4202***		0,4183***			0,3517***
F(20,356)		147,20					
F(22,356)				158,00			
F(22,347)							153,11
Prob. > F		0,000		0,000			0,000
$R^2$ Ajustado		0,737		0,741			0,575

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

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

Os modelos que utilizam a variável instrumental e os componentes não observados encontram um efeito de 2,6% para a densidade do emprego sobre a produtividade local e de 4,9% para a densidade da área iluminada. Nesse sentido, o uso da área iluminada praticamente duplica o efeito da aglomeração urbana sobre a produtividade local, mesmo nos modelos que utilizam variável instrumental. Portanto, os resultados permitem concluir que existe um efeito puro da aglomeração urbana sobre o diferencial salarial nas áreas de emprego do Brasil.

#### 4 Conclusão

O objetivo deste trabalho foi avaliar a contribuição da heterogeneidade não observada das firmas e dos trabalhadores para a determinação de salários individuais e dos salários locais (efeitos de localização) no mercado de trabalho formal do Brasil. As heterogeneidades foram definidas pelos atributos não observados e produtivos, os quais podem apresentar maior remuneração nas áreas mais densas. As áreas de emprego foram definidas a partir dos arranjos populacionais do IBGE (2015). O referencial teórico mostrou que os efeitos positivos da aglomeração urbana sobre a produtividade local (salários reais) podem ser confundidos com a concentração dos trabalhadores e firmas mais produtivos nos mercados de trabalho locais densos, caracterizando um processo de *sorting* espacial entre localizações.

O modelo de Combes, Duranton e Gobillon (2008) foi combinado com o modelo AKM de decomposição salarial para estimar os resultados de interesse. Ambos os modelos sugerem uma estimação em dois estágios. O primeiro estágio estimou equações de salários, utilizando

um painel de micro dados com informações combinadas de trabalhadores e firmas (RAIS, 2002–2014), e identificou os efeitos de localização dos arranjos. Os resultados do primeiro estágio mostraram que as heterogeneidades não observadas individuais exerceram maior poder explicativo para a variação salarial do que os atributos não observados das firmas. A inclusão simultânea dos efeitos não observados dos trabalhadores e das firmas, além das demais covariadas observadas, explicou 93,5% da variação dos salários reais individuais.

No segundo estágio, o modelo de decomposição salarial foi estimado para identificar os efeitos de aglomeração sobre a produtividade local (salário local), a partir das estimativas dos efeitos de localização dos arranjos. A inclusão dos componentes não observados reduziu a estimativa dos efeitos de densidade de 4,2% para 2,6%. O *sorting* se mostrou responsável por quase metade dos efeitos de aglomeração, sugerindo a importância da distribuição espacial de trabalhadores e firmas para a explicação dos diferenciais salariais entre os mercados de trabalho locais. O uso da densidade calculada a partir da área iluminada ampliou o efeito para 3,1%. Por fim, o modelo principal deste estudo, que utiliza IV com a área iluminada e os componentes não observados, gerou um efeito de 4,9%.

Este trabalho contribui com as literaturas nacional e internacional da Economia Urbana e da Economia do Trabalho ao estimar modelos que controlam simultaneamente os efeitos das firmas, dos trabalhadores e da localização. As evidências sugerem a importância do *sorting* espacial baseado nos atributos não observados dos indivíduos e firmas para explicar os diferenciais salariais entre as áreas de emprego e os efeitos de aglomeração. Por sua vez, as evidências encontradas com a densidade do emprego, calculada a partir da área iluminada, mostram que os efeitos de aglomeração podem ser subestimados se as áreas administrativas oficiais de cada localidade são adotadas. Se a diferença entre as áreas é muito grande, como mostram as imagens de satélite da iluminação noturna, as estimativas existentes subestimam os efeitos da aglomeração.

## REFERÊNCIAS

ABOWD, J. M.; CREECY, R. H.; KRAMARZ, F. Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data. **Longitudinal Employer-Household Dynamics Technical Papers**, Center for Economic Studies, US Census Bureau, n. 6. 2002. Disponível em: <<ftp://ftp2.census.gov/ces/tp/tp-2002-06.pdf>>. Acesso em: 6, Jan. 2015.

ABOWD, J. M.; KRAMARZ, F.; MARGOLIS, D. N. High Wage Workers and High Wage Firms. **Econometrica**, v. 67, n. 2, p. 251-333, 1999.

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP. **Anuário estatístico brasileiro do petróleo, gás natural e biocombustíveis**. Rio de Janeiro: ANP, 2016. Disponível em: <[http://www.anp.gov.br/wwwanp/images/publicacoes/Anuario\\_Estatistico\\_ANP\\_2016.pdf](http://www.anp.gov.br/wwwanp/images/publicacoes/Anuario_Estatistico_ANP_2016.pdf)>. Acesso em: 20 Jan. 2017.

ANDREWS, M.; SCHANK, T.; UPWARD, R. Practical fixed-effects estimation methods for the three-way error-components model. **The Stata Journal**, v. 6, n. 4, p. 461-481, 2006.

AUTOR, D. H.; DUGGAN, M. G. The Rise in the Disability Rolls and the Decline in Unemployment. **The Quarterly Journal of Economics**, v. 118, n. 1, p. 157-206, 2003.

AZZONI, C. R. Personal income distribution within states and income inequality between states in Brazil: 1960, 70, 80 and 91. In: HEWINGS, Geoffrey et al. (Eds.). **Understanding and Interpreting Economic Structure**. Springer-Verlag Berlin Heidelberg, 1999. 372 p.

BARROS, R. P.; FRANCO, S.; MENDONÇA, R. A Recente Queda na Desigualdade de Renda e o Acelerado Progresso Educacional Brasileiro da Última Década. In: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Orgs.). **Desigualdade de Renda no Brasil: uma análise da queda recente**. Brasília: IPEA, v. 2, 2007. 552 p.

BARROS, R. P.; HENRIQUES, R.; MENDONÇA, R. Pelo fim das décadas perdidas: educação e desenvolvimento sustentado no Brasil. In: HENRIQUES, Ricardo (Org.). **Desigualdade e pobreza no Brasil**. Rio de Janeiro: IPEA, 2000. 740 p. Disponível em: <<http://www.ipea.gov.br/sites/000/2/livros/desigualdadepobrezabrasil/capitulo14.pdf>>.

Acesso em: 06 Ago. 2011.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

MENEZES-FILHO, N. A.; FERNANDES, R.; PICCHETTI, P. Educação e Queda Recente da Desigualdade no Brasil. In: BARROS, R. P.; FOGUEL, M. N.; ULYSSEA, G. (Orgs.). **Desigualdade de Renda no Brasil: uma análise da queda recente**. Brasília: IPEA, v. 2, 552p. 2007. Disponível em: <[http://www.ipea.gov.br/sites/000/2/livros/desigualdaderenda\\_nobrasilv2/Livrocompleto.pdf](http://www.ipea.gov.br/sites/000/2/livros/desigualdaderenda_nobrasilv2/Livrocompleto.pdf)> Acesso em: 26 Abr. 2012.

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

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

MITTAG, N. New methods to estimate models with large sets of fixed effects with an application to matched employer-employee data from Germany. **Research Data Centre (FDZ) Method Report**, Institute for Employment Research (IAB), Germany, n. 1. 2012. Disponível em: <[http://doku.iab.de/fdz/raporte/2012/MR\\_01-12\\_EN.pdf](http://doku.iab.de/fdz/raporte/2012/MR_01-12_EN.pdf)>. Acesso em: 3 Jan. 2015.

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

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. In: ENCONTRO REGIONAL DE ECONOMIA – ANPEC NORDESTE, 16, 2011, Fortaleza. **Anais...** Fortaleza: ANPEC, 2011. Disponível em: <[http://www.bnb.gov.br/content/aplicacao/eventos/forumbnb2011/docs/2011\\_maiores\\_cidade\\_s.pdf](http://www.bnb.gov.br/content/aplicacao/eventos/forumbnb2011/docs/2011_maiores_cidade_s.pdf)>. Acesso em: 15 Dez. 2011.

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

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

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