

CONTRIBUIÇÃO DOS EFEITOS DE FIRMA E DE INDIVÍDUO PARA OS EFEITOS DE LOCALIZAÇÃO SOBRE OS SALÁRIOS E PARA A VARIAÇÃO SALARIAL DO TRABALHADOR FORMAL DO BRASIL¹

Diana Lúcia Gonzaga da Silva²

O objetivo deste artigo é avaliar a contribuição da heterogeneidade não observada de firma e indivíduo para os efeitos de localização sobre os salários e para a variação salarial do trabalhador formal do Brasil. Este trabalho estima uma equação salarial com as características observadas dos trabalhadores, a heterogeneidade não observada dos trabalhadores e das firmas e os efeitos de localização, usando os microdados da Relação Anual de Informações Sociais e Migração (Rais-Migra) (1995-2008). O modelo de decomposição salarial de Abowd, Kramarz e Margolis (1999) foi utilizado para lidar com múltiplos efeitos fixos em um grande banco de microdados pareados de trabalhador e firma. Os resultados mostram que os efeitos de firma e trabalhador respondem por substancial variação dos salários reais e dos efeitos de localização sobre os salários dos trabalhadores formais.

Palavras-chave: decomposição salarial; efeito individual; efeito de firma; efeito de localização; heterogeneidade.

THE CONTRIBUTION OF THE FIRM AND WORKER EFFECTS FOR THE LOCATION EFFECTS ON WAGES AND FOR THE FORMAL WORKER'S WAGE VARIATION FROM BRAZIL

The objective of this paper is to assess the contribution of unobserved heterogeneity of firm and individual to the effects of location on wages and the wage variation of the formal worker in Brazil. This paper estimates a wage equation with observable characteristics of workers, unobserved heterogeneity of worker and firm and the effects of location, using the Rais-Migra microdata (1995-2008). The wage decomposition model of Abowd, Kramarz and Margolis (1999) is used to deal with multiple fixed effects in a large database which is composed of worker and firm combinations. The findings show that the firm and worker effects account for a substantial variation of the real wages and effects of location on wages of formal workers from Brazil.

Keywords: wage decomposition; individual effect; firm effect; location effect; heterogeneity.

JEL: J24; J31; R23; C23.

1 INTRODUÇÃO

O entendimento dos determinantes e das disparidades salariais tem sido alvo de diversos estudos em economia. No Brasil, muitas evidências apontaram para as características observadas, a exemplo da educação, como principais determinantes salariais.

1. Este trabalho foi desenvolvido com o suporte financeiro de bolsa de doutorado da Fundação de Amparo à Pesquisa do Estado de São Paulo (Fapesp), no âmbito do convênio entre Fapesp e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), vinculado ao processo nº 2014/03077-7.

2. Doutora em economia pela Universidade de São Paulo (USP) e professora adjunta no Departamento de Economia da Universidade Federal da Bahia (UFBA). *E-mail:* <diana.gonzaga@gmail.com>.

A partir da equação minceriana (Mincer, 1974), os estudos sobre os determinantes salariais avançaram no controle de características observáveis e não observáveis dos trabalhadores. Entretanto, permanecia um diferencial salarial favorável aos grandes centros urbanos, que motivou o surgimento de estudos integrando a economia urbana e a economia do trabalho para explicar o diferencial salarial, a partir das economias de aglomeração urbana. A localização passou, então, a ser incluída como um determinante salarial.

A distribuição espacial de salários no Brasil é bastante desigual. Os dados da Relação Anual de Informações Sociais e Migração (Rais-Migra), do Ministério do Trabalho e Emprego (MTE), mostram que, entre as regiões metropolitanas (RMs) do Brasil, o salário real de João Pessoa correspondeu a 65% do salário de São Paulo em 2008. Em geral, os salários das RMs são maiores do que os dos estados. As RMs de Florianópolis e São Luís apresentaram salários reais 45% e 34% maiores do que seus respectivos estados, em 2008. Isso sugere a existência de ganhos de aglomeração nessas áreas. No entanto, ainda persiste um diferencial salarial no Brasil em benefício das macrorregiões mais dinâmicas do Centro-Sul do país, mesmo entre as áreas metropolitanas. Diversos fatores podem explicar as diferenças observadas de salários, entre eles os custos 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, por tamanho de empresas, entre outros (Servo e Azzoni, 2002; Silveira Neto e Azzoni, 2004; Rocha, Silveira Neto e Gomes, 2011).

Entretanto, o controle das características observadas dos trabalhadores não é suficiente para dissipar as desigualdades salariais. Estudos mais recentes consideram a presença de habilidades não observadas dos trabalhadores na diferenciação salarial. Considerar as habilidades não observadas possibilita corrigir o viés de seleção associado à atração de trabalhadores mais habilidosos pelos grandes centros urbanos, reduzindo ou eliminando grande parte da vantagem salarial aparente associada às áreas urbanas densas. Combes, Duranton e Gobillon (2008) mostraram que as diferenças na composição de habilidades foram responsáveis por 40% a 50% das disparidades espaciais de salários de trabalhadores da França, sugerindo evidências de *sorting* por habilidades. Combes *et al.* (2012) mostraram uma seleção negativa por habilidades na migração de trabalhadores para as áreas menos densas, e positiva para as mais densas da França.

No Brasil, Freguglia e Menezes-Filho (2011) mostraram que 63% do diferencial salarial, entre 1995 e 2002, era explicado pelas características não observadas dos trabalhadores. Contudo, o diferencial salarial real persistiu mesmo após considerar as habilidades observadas e não observadas e as características ocupacionais. A persistência do diferencial sugere a existência de efeitos específicos de localização sobre os salários. A literatura sobre as economias de aglomeração prevê maiores salários e produtividade em áreas urbanas densas. Os ganhos de aglomeração no

mercado de trabalho podem emergir do aumento da produtividade dos trabalhadores na presença de externalidades positivas, tais como o acesso à maior diversidade de ocupações, o acúmulo de capital humano e *spillovers* de conhecimento, as interações sociais, entre outras. Um mercado mais amplo permite o compartilhamento (*sharing*) de fornecedores e de trabalhadores com habilidades similares, além de possibilitar melhor *matching* entre trabalhadores e firmas. As áreas mais densas facilitam a aprendizagem (*learning*) e o desenvolvimento de novas tecnologias (Puga, 2010).

Os maiores salários recebidos pelos trabalhadores nas cidades densas representam um prêmio salarial urbano, com ganhos variando entre 5% e 10% dos salários reais, na literatura internacional (Halfdanarson, Heuermann e Südekum, 2008). Os trabalhos de Glaeser e Maré (1994; 2001) identificam uma relação positiva entre os salários e o tamanho das cidades, com salários 32%-33% maiores para os trabalhadores de áreas metropolitanas dos Estados Unidos. A partir destes estudos, tem surgido um ramo de pesquisa que incorpora as economias de aglomeração urbana como fatores relevantes para a determinação salarial.

No Brasil, ainda existem poucos estudos nesta linha, mas já há evidências da existência de ganhos de aglomeração no mercado de trabalho. Rocha, Silveira Neto e Gomes (2011) encontraram um diferencial de 0,12 salário mínimo para os trabalhadores de RMs, após o controle por características observáveis e não observáveis, entre 2000 e 2008. Campos e Silveira Neto (2009) encontraram um ganho salarial de 16% para os trabalhadores de RMs, em 2000. Silva, Santos e Freguglia (2016) encontraram um prêmio salarial urbano de 3,4% nas RMs utilizando os microdados da Relação Anual de Informações Sociais (Rais), para o período de 1995 a 2008, controlando as características observadas e não observadas dos trabalhadores. As evidências sugerem um ganho específico associado à localização e reforçam a necessidade de compreender os diferenciais salariais persistentes, a partir das economias de aglomeração.

Os ganhos associados às características de trabalhadores e firmas localizados em determinadas áreas podem ser confundidos com os ganhos de aglomeração. Combes *et al.* (2012) sugeriram o *sorting* por habilidades dos trabalhadores e a atuação diferenciada dos benefícios de aglomeração sobre as habilidades como explicações para os diferenciais salariais espaciais. Se as escolhas de localização não são exógenas, atributos não observados de firmas e trabalhadores em áreas densas podem enviesar as estimativas do prêmio salarial urbano. O *sorting* por habilidades surge porque trabalhadores com melhores atributos tendem a aglomerar-se nos mercados de trabalho densos. Combes, Duranton e Gobillon (2008) destacaram que a identificação de estimativas não enviesadas na equação salarial requer a inclusão dos efeitos fixos de firmas, uma vez que estes estariam correlacionados com os efeitos de setores. A seleção de firma pode surgir se apenas as mais produtivas sobreviverem nas áreas densas.

Combes, Duranton e Gobillon (2008) e Combes *et al.* (2012) desenvolveram uma abordagem unificada dos determinantes salariais, incluindo os efeitos de localização, de setores e as habilidades não observadas. No entanto, por limitações de dados, não adotaram controles suficientes associados às características observadas dos trabalhadores, tais como a educação. Os estudos no Brasil controlam somente uma parte do *sorting* espacial, associado às heterogeneidades não observadas dos trabalhadores. Entretanto, o entendimento das questões associadas ao *sorting* de trabalhadores e firmas, em que os trabalhadores mais produtivos seriam selecionados para as firmas que pagam os maiores salários, requer a disponibilidade de dados longitudinais pareados (*matched*) de trabalhador e firma.

A limitação empírica para estimar modelos com heterogeneidade não observada de trabalhador e firma por mínimos quadrados (*full least square*) surge da complexidade computacional para incluir uma dimensão muito grande de identificadores de firma e trabalhador com dados longitudinais. Desde o estudo de Abowd, Kramarz e Margolis (1999), a literatura internacional que trata dos diferenciais de salários tem buscado soluções estruturais e estatísticas, factíveis computacionalmente, para estimar os parâmetros dos modelos que lidam com heterogeneidade de firma e trabalhador e encontrar resultados próximos à solução geral de mínimos quadrados (*full least square*). As abordagens recentes buscam técnicas para lidar com os modelos de múltiplos efeitos fixos (Abowd, Kramarz e Margolis, 1999; Abowd, Creecy e Kramarz, 2002; Andrews, Schank e Upward, 2006; Guimarães e Portugal, 2010; McCaffrey *et al.*, 2012; Mittag, 2012; Gaure, 2013).

Abowd, Kramarz e Margolis (1999) propuseram um modelo estatístico de decomposição salarial buscando explicar o diferencial setorial de salários da França, a partir de componentes associados às características observadas dos trabalhadores e à heterogeneidade não observada de indivíduos e firmas. O modelo foi estimado com dados longitudinais para o período 1976-1987. O estudo desenvolveu abordagens alternativas de estimação da equação de salários, dada a limitação computacional da solução geral de mínimos quadrados. Os resultados encontrados mostraram que a heterogeneidade individual não observada era uma fonte muito importante da variação salarial na França, mais relevante do que os efeitos de firma, e ambos os efeitos eram pouco correlacionados. As especificações que controlavam os efeitos fixos de trabalhador explicaram entre 77% e 83% da variação salarial da França, enquanto as demais especificações explicaram entre 32% e 55%. Ao incluir somente o efeito de firma, apenas 55% da variação salarial foi explicada. Os efeitos individuais explicaram cerca de 90% do diferencial salarial intersetorial e os efeitos de firma entre 7% e 25% do diferencial. Ambos responderam por 85%-96% da variação setorial de salários da França.

Menezes-Filho, Muendler e Ramey (2008) buscaram replicar para São Paulo, estado mais desenvolvido do Brasil, o modelo de decomposição salarial de Abowd *et al.* (2001) estimado com dados da França e dos Estados Unidos. O estudo do Brasil controlou um efeito combinado de estabelecimento, correspondente ao efeito de firma somado a uma média dos efeitos dos trabalhadores em cada firma. Foram utilizados os microdados da Rais (Brasil, 1990; 1997). Os resultados do Brasil foram compatíveis com os Estados Unidos e a França, mostrando que as características observadas dos trabalhadores foram os fatores mais relevantes para a estrutura salarial da indústria, particularmente no Brasil. O efeito de estabelecimento mostrou-se limitado em São Paulo. Os componentes observados dos trabalhadores e o efeito de estabelecimento explicaram entre 51% e 47% da variação salarial do Brasil. A variabilidade salarial residual foi maior no Brasil, sugerindo a possível relevância dos componentes não observados individuais para o país.

A desigualdade salarial e os retornos à educação foram maiores no Brasil do que na França e nos Estados Unidos, de acordo com Menezes-Filho, Muendler e Ramey (2008). Os diferenciais salariais por ocupação e gênero eram muito similares entre os países, embora o prêmio ocupacional do Brasil fosse mais parecido com o da França. A estrutura salarial do Brasil mostrou-se mais parecida com a França. A similaridade pode ser justificada pela estrutura institucional e pela regulação mais intensa do mercado de trabalho de ambos os países, assim como por seus retornos salariais associados às características da firma (tamanho, capital e intensidade de qualificação).

Portanto, este artigo busca contribuir para a literatura nacional ao investigar o *sorting* associado aos salários dos trabalhadores formais e aos efeitos de localização, a partir do controle dos efeitos de firma e trabalhador, recorrendo aos avanços metodológicos na literatura internacional. A pesquisa propõe o uso de um painel de microdados pareados de trabalhador e firma, a Rais-Migra, do MTE, para estimar equações de salário com as características observadas dos trabalhadores, a heterogeneidade não observada dos trabalhadores e das firmas e os efeitos de localização nas RMs, entre os anos de 1995 a 2008.

Desse modo, além desta introdução, este trabalho é composto por mais quatro seções. A seção 2 apresenta a abordagem proposta por Abowd, Kramarz e Margolis (1999), na qual este artigo é baseado. Na seção 3 são apresentados o modelo econométrico, a base de dados e a estratégia empírica adotada para a estimação dos resultados. A seção 4 descreve os resultados encontrados. Por fim, a seção 5 apresenta as considerações finais.

2 HETEROGENEIDADE NO MERCADO DE TRABALHO E DIFERENCIAL SALARIAL: MODELO DE DECOMPOSIÇÃO SALARIAL

Os resultados subjacentes aos mercados de trabalho são bastante heterogêneos, o que tem motivado uma extensa literatura para isolar as fontes de tais heterogeneidades. Indivíduos observacionalmente similares podem ganhar salários diferenciados e possuir trajetórias distintas de emprego. A ausência de microdados relacionando características de trabalhadores e firmas tem sido um fator limitador nas pesquisas que buscam separar os efeitos decorrentes de decisões de firmas daqueles resultantes de decisões do trabalhador. Uma abordagem desta literatura tem focado sobre a heterogeneidade individual não observada (efeito de indivíduo), enquanto outra abordagem tem avaliado a heterogeneidade não observada de firmas (efeito firma) como explicação para os diferenciais salariais (Abowd, Kramarz e Margolis, 1999).

Para abordar os fatores relacionados à heterogeneidade salarial será apresentado um modelo de decomposição estatística de salários, desenvolvido por Abowd, Kramarz e Margolis (1999) para tratar os diferenciais salariais intersetoriais. Ainda é controverso o papel da heterogeneidade não observada como fonte dos diferenciais salariais entre setores. Em Abowd, Kramarz e Margolis (*op. cit.*), o efeito puro de setor é definido como a agregação dos efeitos puros de firmas dentro do respectivo setor, correspondendo às variáveis indicadoras de setor na equação salarial. O efeito residual de firma é definido como um desvio do efeito de setor. Portanto, o modelo considera a seguinte equação de salários:

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

onde X é uma matriz ($N^* \times P$) de características observáveis (em desvios da média) variantes no tempo; D é uma matriz ($N^* \times N$) de indicadores para os indivíduos; F é uma matriz ($N^* \times m$) de indicadores para os efeitos da firma, em que i trabalha em t (J é o número total de firmas); a matriz A ($J \times K$) classifica cada uma das J firmas em um dos K setores, tal que $K(j)$ denota a classificação setorial da firma j ; y é o vetor ($N^* \times 1$) de salários (em desvios da média); ε é o vetor de resíduos; e $N^* = NT$. Os parâmetros são os vetores β ($P \times 1$), θ ($N \times 1$), ψ ($m \times 1$) e a variância do erro σ^2_ε . O vetor de parâmetros κ ($K \times 1$) é interpretado como uma média ponderada do efeito puro de firma. O efeito $(F\psi - FA\kappa)$ pode ser representado como $M_{FA}F\psi$. Os termos $FA\kappa$ e $M_{FA}F\psi$ resultam da decomposição de $F\psi$ em dois componentes ortogonais.

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

$$E[\varepsilon_{it} \mid i, t, J(i, t), x_{it}] = 0, \quad (1.1)$$

$$cov [\varepsilon_{it}, \varepsilon_{ns} \mid i, t, n, s, J(i, t), J(n, s), x_{it}, x_{ns}] = \begin{cases} \sigma_\varepsilon^2 & \text{para } i = n \text{ e } t = s \\ 0 & \text{caso contrário} \end{cases}, \quad (1.2)$$

onde $i = 1, \dots, N$ indica o indivíduo; $t = 1, \dots, T$ o período de tempo; e $J(i, t)$ indica a firma em que o trabalhador i está empregado no tempo t . A equação (1) pode ser interpretada como a esperança condicional dos salários, dados os indicadores de indivíduos, firmas e setores e as características observadas.

Embora alguns estudos já tenham avançado na solução *least squares* para a equação (1), com amostras pequenas (Leonard e Van Audenrode, 1996; Entorf, Gollac e Kramarz, 1999; Goux e Maurin, 1999), o esforço da literatura recente tem sido encontrar soluções factíveis para grandes bancos de dados pareados de firma e trabalhador. Muitos estudos estimam versões incompletas dessa equação, o que gera interpretações ambíguas sobre os seus parâmetros. A estimação da equação (1) sem algum componente, ou com combinações lineares deles (viés de agregação), pode modificar o significado dos efeitos subjacentes. Nesse contexto, Abowd, Kramarz e Margolis (1999) descrevem os casos de variações na parametrização da equação (1), derivados da literatura.

Uma versão incompleta da equação (1), sem o efeito firma residual ($M_{FA}F\psi$), implicará κ^* e β^* , os efeitos verdadeiros somados a um termo de viés de variável omitida:

$$\kappa^* = \kappa + (A'F'M_{[D \ X]}FA)^{-1} A'F'M_{[D \ X]}M_{FA}F\psi, \quad (2)$$

$$\beta^* = \beta + (X'M_{[D \ FA]}X)^{-1} X'M_{[D \ FA]}M_{FA}F\psi, \quad (3)$$

onde $M_{[D \ X]}$ é a matriz M_Z com $Z \equiv [D \ X]$, assumindo $M_A \equiv I - A(A'A)^{-1}A'$ para uma matriz A arbitrária. Por sua vez, uma versão da equação (1) sem o remanescente efeito de firma ($M_{FA}F\psi$) e o efeito de indivíduo (θ) implicará κ^{**} .

$$\begin{aligned} \kappa^{**} &= \kappa + (A'F'M_XFA)^{-1} A'F'M_X(M_{FA}F\psi + D\theta) \\ &\equiv (A'F'M_XFA)^{-1} A'F'M_XF\psi + (A'F'M_XFA)^{-1} A'F'M_XD\theta. \end{aligned} \quad (4)$$

Esta última expressão representa a decomposição dos efeitos setoriais realizada por Abowd, Kramarz e Margolis (1999) em componentes de firma e indivíduo. Portanto, as estimativas de versões incompletas da equação salarial devem resultar em estimativas e interpretações enviesadas dos efeitos de firma e indivíduo e em efeitos setoriais inconsistentes.

2.1 Métodos de estimação do modelo de decomposição salarial

Com base na equação de salários (1), Abowd, Kramarz e Margolis (1999) desenvolveram abordagens estatísticas alternativas à de mínimos quadrados para estimar o modelo completo, incluindo os efeitos de firma e de indivíduo. A matriz (*cross-product*) contendo as variáveis da equação salarial completa, sem os efeitos setoriais, a qual deve ser utilizada para a identificação dos parâmetros, é dada por:

$$\begin{bmatrix} X'X & X'D & X'F \\ D'X & D'D & D'F \\ F'X & F'D & F'F \end{bmatrix},$$

cujas dimensões dependem do tamanho de N (indivíduos) e J (firmas). Métodos computacionais usuais para a estimação dos parâmetros $[\beta' \theta' \psi']$ por mínimos quadrados não são, em geral, factíveis. Abowd, Kramarz e Margolis (1999) propõem métodos alternativos para preservar, em alguma medida, a estrutura geral do modelo completo. A mobilidade dos indivíduos entre firmas é uma condição necessária para a identificação estatística do modelo, ou seja, para encontrar os efeitos fixos separadamente, independentemente da abordagem computacional adotada. A primeira abordagem, denominada de método consistente, recorre à primeira diferença nos dados agrupados intrafirma-indivíduo, usando a definição:

$$\begin{aligned} \psi_j &= \phi_j + \gamma_j s_{it} \\ F\psi &= F_0\phi + F_1\gamma, \end{aligned} \quad (5)$$

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

$$\begin{aligned} y_{in_{it}} - y_{in_{it-1}} &= (x_{in_{it}} - x_{in_{it-1}})\beta + \gamma_{J(i, n_{it})} (s_{in_{it}} - s_{in_{it-1}}) + \varepsilon_{in_{it}} - \varepsilon_{in_{it-1}}, \\ \Delta y &= \Delta X\beta + \tilde{F}\gamma + \Delta \varepsilon, \end{aligned} \quad (6)$$

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

$$\tilde{\beta} = (\Delta X' M_{\tilde{F}} \Delta X)^{-1} \Delta X' M_{\tilde{F}} \Delta y, \quad (7)$$

$$\tilde{\gamma} = (\tilde{F}'\tilde{F})^{-1} \tilde{F}'(\Delta y - \Delta X\tilde{\beta}). \quad (8)$$

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

Por sua vez, a classe de estimadores definidos em Abowd, Kramarz e Margolis (1999) como métodos condicionais não restringe a amostra na estimação, sendo capaz de identificar os efeitos fixos separadamente. A denominação de condicional resulta da sua relação com as técnicas-padrão de modelos lineares e de dados em painel, associada aos modelos de efeitos fixos individuais. Supostos adicionais de ortogonalidade serão necessários: as interações entre X , D e F serão *proxy* para as correlações entre estas variáveis e a estimação assumirá ortogonalidade condicional, dadas as interações.

Neste contexto, define-se uma matriz Z ($N^* \times Q$) a partir de Q funções de informações em X , D e F . Para a construção de Z , em Abowd, Kramarz e Margolis (1999), foram incluídos o tamanho da firma e seu quadrado, o setor, a experiência do trabalhador e sua idade ao final do período de estudos. A solução de mínimos quadrados para (1), sem efeitos setoriais, é encontrada sob a hipótese de que X e D são ortogonais à projeção de F sobre o espaço nulo de Z :

$$y = X\beta + D\theta + Z\gamma + M_z F\psi + \varepsilon, \quad (9)$$

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

$$X'M_z F = 0, \quad (10)$$

$$D'M_z F = 0. \quad (11)$$

A estimação independente da ordem (*order-independent estimation*) é um método condicional realizado em dois estágios independentes. O primeiro estágio segue uma abordagem de estimação longitudinal, intraindivíduos (*within-D*), em que X e Z são projetados sobre D , obtendo:

$$\begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'M_D X & X'M_D Z \\ Z'M_D X & Z'M_D Z \end{bmatrix}^{-1} \begin{bmatrix} X'M_D y \\ Z'M_D y \end{bmatrix}, \quad (12)$$

$$\hat{\theta} = (D'D)^{-1} D'(y - X\hat{\beta} - Z\hat{\lambda}). \quad (13)$$

O segundo estágio, *intrafirma (within-F)*, computa a solução de mínimos quadrados para os parâmetros de F e Z , a partir da seguinte especificação:

$$y = F\psi + Z\pi + \nu, \quad (14)$$

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

$$\hat{\pi} = (Z'M_F Z)^{-1} Z'M_F y, \quad (15)$$

$$\hat{\psi} = (F'F)^{-1} F'(y - Z\hat{\pi}). \quad (16)$$

No método condicional dependente da ordem (*order-dependent estimation*), os estágios são dependentes, de modo que as estimativas dos parâmetros podem ser diferentes, a depender de quais efeitos são estimados primeiro. Na abordagem em que os efeitos de indivíduos são estimados, o primeiro estágio recupera os parâmetros β , θ e λ como no método de independência da ordem, de acordo com (12) e (13). No segundo estágio, os efeitos de firma são estimados utilizando as equações em (5). Para tanto, todas as observações sobre os indivíduos empregados na mesma firma são agrupadas no conjunto $\{j\} \equiv \{(i,t) \mid J(i,t) = j\}$, de N_j elementos, tal que:

$$\hat{y}_{\{j\}} \equiv y_{\{j\}} - x_{\{j\}}\hat{\beta} - \hat{\theta}_{\{j\}}, \quad (17)$$

onde $\hat{\beta}$ e $\hat{\theta}$ são as estimativas de primeiro estágio. A equação no nível de firma será dada por:

$$\hat{y}_{\{j\}} = F_{\{j\}} \begin{bmatrix} \phi_j \\ \gamma_j \end{bmatrix} + \zeta_{\{j\}} \quad (18)$$

$$\zeta_{\{j\}} \equiv \varepsilon_{\{j\}} + x_{\{j\}}(\beta - \hat{\beta}) + (\theta_{\{j\}} - \hat{\theta}_{\{j\}}).$$

Portanto, o estimador de mínimos quadrados de (23) será:

$$\begin{bmatrix} \hat{\phi}_j \\ \hat{\gamma}_j \end{bmatrix} = (F'_{\{j\}} F_{\{j\}})^{-1} F'_{\{j\}} \hat{y}_{\{j\}}, \text{ para } j = 1, \dots, J. \quad (19)$$

Por sua vez, na abordagem em que os efeitos de firma são estimados, o primeiro estágio utiliza o estimador de $\hat{\psi}$ dado por (16), enquanto o segundo estágio encontra β e θ a partir de:

$$\begin{aligned}y - F\hat{\psi} &= X\beta + D\theta + \zeta \\ \zeta &= \varepsilon + F(\psi - \hat{\psi}).\end{aligned}\tag{20}$$

Portanto, os estimadores de β e θ serão:

$$\hat{\beta} = (X'M_D X)^{-1} X'M_D (y - F\hat{\psi}),\tag{21}$$

$$\hat{\theta} = (D'D)^{-1} D'(y - X\hat{\beta} - F\hat{\psi}).\tag{22}$$

Os estimadores de β e θ obtidos pelo método independente da ordem são idênticos aos obtidos pelo método condicional dependente da ordem, em que os efeitos de indivíduos são estimados primeiro. O estimador independente da ordem de ψ é idêntico ao seu estimador condicional dependente da ordem, com o efeito de firma estimado primeiro.

3 BASE DE DADOS, MODELO ECONOMÉTRICO E ESTRATÉGIA EMPÍRICA

O banco de dados deste estudo é composto por uma amostra aleatória de 5% dos trabalhadores formais da base Rais-Migra, para o período de 1995 a 2008. A Rais-Migra é uma ampla base de dados longitudinal do Brasil pareada de trabalhador e firma, extraída dos registros administrativos da Rais do MTE. O banco de dados final utilizado é composto por 4.665.388 observações, que correspondem a um painel balanceado com 333.242 trabalhadores formais do Brasil, empregados em cada ano, com rendimento positivo e idade entre 18 e 65 anos. Na amostra, existem 140.850 firmas e 589.467 combinações de trabalhador e firma.

Para a construção dos salários nominais, a remuneração do trabalhador, em número de salários mínimos, foi multiplicada pelo salário mínimo vigente em cada ano. O Índice Nacional de Preços ao Consumidor Amplo (IPCA) do Instituto Brasileiro de Geografia e Estatística (IBGE) foi utilizado para deflacionar os salários, assumindo 2008 como o ano-base. O IPCA é calculado pelo IBGE para apenas onze RMs do Brasil: Belém, Fortaleza, Recife, Salvador, São Paulo, Rio de Janeiro, Goiânia, Distrito Federal, Belo Horizonte, Curitiba e Porto Alegre. Desse modo, para deflacionar os valores nominais para todas as RMs, os índices foram extrapolados para os estados correspondentes. O critério de maior aproximação foi utilizado para os estados que não possuíam alguma das RMs contempladas no cálculo do índice.³ Assim, o índice de preços de Belém foi extrapolado para os estados da região Norte; o de Fortaleza para Maranhão, Piauí e Rio Grande

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

do Norte; o de Recife para Paraíba, Alagoas e Sergipe; o do Rio de Janeiro para Espírito Santo; o de Curitiba para Santa Catarina; e o de Goiânia para os estados de Mato Grosso e de Mato Grosso do Sul.

Como características observadas dos trabalhadores foram extraídas a escolaridade, em ciclos de estudo, a idade (18-65 anos) e a experiência (em meses, no mesmo vínculo). Na Rais-Migra, a variável educação é definida por grau de instrução, correspondente aos seguintes ciclos: até o 5º ano do ensino fundamental incompleto; 5º ano completo do ensino fundamental; do 6º ao 9º anos do ensino fundamental incompleto; ensino fundamental completo; ensino médio incompleto; ensino médio completo; ensino superior incompleto; e ensino superior completo. Para a inclusão da educação no modelo serão definidas variáveis *dummies* para cada ciclo escolar: analfabeto (*educ1*); fundamental incompleto ou completo (*educ2*); ensino médio incompleto ou completo (*educ3*); e ensino superior incompleto ou completo (*educ4*).

A definição de RM adotada neste artigo segue a classificação do IBGE, a qual será assumida para definir as áreas urbanas densas. Foram identificadas 24 RMs no Brasil, na Rais-Migra: Belém, Macapá, São Luís, Fortaleza, Natal, Recife, Maceió, Salvador, Belo Horizonte, Vale do Aço, Vitória, Rio de Janeiro, São Paulo, Baixada Santista, Campinas, Curitiba, Londrina, Maringá, Florianópolis, Vale do Itajaí, Norte/Nordeste Catarinense, Porto Alegre, Goiânia e João Pessoa.

A identificação dos efeitos fixos de firma e localização requer mobilidade (*within*) dos trabalhadores entre firmas e RMs. A tabela 1 apresenta a variação *within* das variáveis. A variação *within* é a variação existente intraindivíduo, enquanto a variação *between* é a variação entre os indivíduos. Como se pode observar na tabela 1, a condição de mobilidade é atendida para a identificação dos estimadores.

TABELA 1
Mobilidade de indivíduos entre firmas e RMs (1995-2008)

Variável	Variação	Parcela de variação (%)
	<i>Overall</i>	-
RM	<i>Between</i>	0,95
	<i>Within</i>	0,05
	<i>Overall</i>	-
Firma	<i>Between</i>	0,91
	<i>Within</i>	0,09

Fonte: Rais-Migra/MTE.

Elaboração da autora.

Obs.: A parcela de variação é construída com o valor da variância, obtido a partir do desvio-padrão *within* (*between*), o qual foi dividido pela variância total (*overall*). A mobilidade entre RMs inclui, também, os migrantes que estão entrando em alguma RM.

A estratégia de identificação deste estudo consiste na adoção de modelagem de dados em painel, para controlar a heterogeneidade não observada individual e de firma e o respectivo *sorting* sobre os salários e os efeitos de localização. O controle dos efeitos fixos de indivíduos permite eliminar o viés das estimativas dos parâmetros na equação salarial. Permite, ainda, a correção do problema de autosseleção associado à atração dos trabalhadores mais habilidosos para os grandes centros urbanos, o que pode ser confundido com os ganhos de aglomeração. A inclusão dos efeitos de localização em áreas metropolitanas permite controlar o *sorting* ou a autosseleção entre localizações, associado à distribuição espacial heterogênea de firmas e trabalhadores, e lidar com os fatores específicos e as amenidades locais constantes no tempo.

O modelo econométrico adotado para as estimações assume a seguinte equação minceriana estendida:

$$\ln(w_{it}) = \beta_0 + \beta_1 educ_{it} + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + \beta_4 idade_{it} + \beta_5 idade_{it}^2 + \alpha RM_{\kappa(it)} + \gamma F_{j(it)} + \theta_i + T_t + \varepsilon_{it}. \quad (23)$$

Na equação (23), $\ln(w_{it})$ é o logaritmo natural do salário real do trabalhador i no tempo $t = 1995, \dots, 2008$; $educ_{it}$ é um vetor de *dummies* de escolaridade para quatro ciclos de estudo (analfabeto, ensino básico, ensino médio e ensino superior); exp_{it} é a experiência de trabalho em meses; exp_{it}^2 e $idade_{it}^2$ são os termos quadráticos da experiência e idade. Quando for possível, será incluída uma *dummy* de gênero (= 1 se feminino). O vetor de *dummies* de localização ($RM_{\kappa(it)}$) é a variável que vai captar os efeitos fixos de localização em cada RM ($\kappa = 1, 2, \dots, 24$), no tempo t . A categoria de referência serão os municípios que não pertencem a alguma área metropolitana. Se o efeito fixo de RM for positivo e significativo, o vetor α representará a magnitude do prêmio salarial urbano. As habilidades individuais não observáveis do trabalhador são representadas pelo termo θ_i . O vetor $F_{j(it)}$ denota os efeitos fixos de cada firma j no tempo t . Os efeitos temporais correspondem ao vetor T_t , sendo 1995 o ano-base, e o termo de erro do modelo é definido por ε_{it} .

A tabela 2 apresenta as estatísticas descritivas básicas das variáveis na amostra. A maior parte dos trabalhadores é do sexo masculino (57%), possui idade média de 40 anos e ensino fundamental ou médio (35%). Nas RMs, também foi revelada uma predominância dos trabalhadores do sexo masculino (55%), com idade média de 40 anos e ensino médio (35%). O salário real médio (w_{real}), a preços de 2008, nas áreas metropolitanas (R\$ 2.630,35), era maior do que fora delas (R\$ 1.926,45), um indício de prêmio salarial urbano para os trabalhadores nessas áreas. Cerca de 3% e 6% da amostra correspondem às observações que apresentaram mobilidade entre municípios (*migra_mun*) e firmas (*mover_firma*), respectivamente, no painel.

TABELA 2
Estatísticas descritivas de variáveis na amostra (1995-2008)

Variável	Total	Amostra das RMs				Amostra sem RMs			
	Média	Média	Desvio	Mínimo	Máximo	Média	Desvio	Mínimo	Máximo
<i>exp</i>	145,35	148,44	89,92	0	678,80	140,57	85,89	0	599,90
<i>idade</i>	40	40,38	8,60	18	65	39,41	8,61	18	65
<i>w_real</i>	2.353,62	2.630,35	3.001,02	66,37	153.731,30	1.926,45	2.586,12	69,50	130.625,50
<i>ln_wreal</i>	7,32	7,46	0,88	4,20	11,94	7,10	0,91	4,24	11,78
<i>educ1</i>	0,01	0,01	0,09	0	1	0,01	0,12	0	1
<i>educ2</i>	0,35	0,31	0,46	0	1	0,42	0,49	0	1
<i>educ3</i>	0,35	0,35	0,48	0	1	0,35	0,48	0	1
<i>educ4</i>	0,29	0,34	0,47	0	1	0,21	0,41	0	1
<i>dsexo</i>	0,43	0,45	0,50	0	1	0,41	0,49	0	1
<i>migra_mun</i>	0,03	0,03	0,17	0	1	0,03	0,18	0	1
<i>mover_firma</i>	0,06	0,07	0,25	0	1	0,05	0,21	0	1
Observações	4.665.388	2.831.248				1.834.140			

Fonte: Rais-Migra/MTE.
Elaboração da autora.

A estimação do modelo econométrico completo adotará um algoritmo de iteração de ponto fixo, proposto por Guimarães e Portugal (2010) e desenvolvido, em sua versão implementável para o *software* Stata,⁴ por Correia (2014), para estimar modelos com múltiplos efeitos fixos, tais como o de Abowd, Kramarz e Margolis (1999). Para entender o funcionamento do algoritmo, assumo o seguinte modelo, cujo interesse é obter estimativas de mínimos quadrados ordinários (MQO) para β :

$$y = X\beta + D_1\alpha_1 + D_2\alpha_2 + D_3\alpha_3 + \epsilon, \quad (24)$$

onde D_i são as variáveis indicadoras e α_i os efeitos fixos correspondentes. O algoritmo segue o teorema de Frisch-Waugh-Lovell (FWL), em que y e X são regredidos contra cada D_i , gerando os resíduos u_y e u_x , e, então, u_y será regredido contra u_x . De maneira prática, o algoritmo faz uma regressão linear absorvendo (*demean*) os efeitos fixos indicados. O apêndice apresenta a estratégia de iteração de ponto fixo utilizada para a estimação dos modelos com múltiplos efeitos fixos.

Portanto, esse será o método adotado para estimar o modelo completo (23) e encontrar a solução condicional de MQO, com a inclusão simultânea dos efeitos fixos. Além do método condicional, serão estimadas versões alternativas do modelo geral, por meio do método consistente intratrabalhador-firma e recorrendo aos métodos

4. O algoritmo na versão Stata é o *reghdfe* (*linear and instrumental-variable regression absorbing any number of high-dimensional fixed effects*) e pode ser encontrado em Correia (2014).

tradicionais de dados em painel, como o método de mínimos quadrados ordinários agrupado (POLS) e o método de efeitos fixos.

4 RESULTADOS

Esta seção apresenta os resultados das estimações econométricas do modelo geral e das versões reduzidas. A seção também traz as contribuições dos efeitos fixos de firma e trabalhador para os efeitos de localização em RMs do Brasil.

4.1 Resultados das equações salariais

A tabela 3 apresenta os resultados das estimações econométricas realizadas com base em diferentes métodos, para lidar com os efeitos fixos de firma, indivíduo e localização, a partir da equação salarial (23). Os resultados da coluna I mostram a estimação, pelo método POLS, da equação salarial com as características observadas dos trabalhadores e os efeitos de localização em RMs, sem controlar os efeitos fixos de firma e trabalhador. A coluna (II) apresenta os resultados de uma abordagem de estimação condicional, em que os efeitos fixos de firma e trabalhador são *absorbed*, enquanto os efeitos de localização são incluídos como variáveis na regressão. Essa coluna foi denominada de condicional, com efeitos fixos de trabalhador sendo estimados primeiro, pois o método de iteração adotado funciona como a abordagem condicional proposta em Abowd, Kramarz e Margolis (1999), como foi explicado na seção anterior.

Guimarães e Portugal (2010) apontaram as vantagens de subtrair (*demean*) a influência dos efeitos fixos sobre as variáveis e trabalhar somente com os resíduos (Frisch-Waugh-Lovell), em relação à estimação de regressões completas, que incluem todos os efeitos fixos diretamente. As vantagens estariam associadas ao aumento na velocidade de convergência das regressões do primeiro estágio, à possibilidade de testar especificações do modelo com as variáveis transformadas, sem precisar lidar com os múltiplos efeitos fixos, e à redução da capacidade computacional necessária para grandes bancos de dados. Tais vantagens estão subjacentes, neste artigo, aos métodos denominados de condicionais (*absorbed*).

A coluna (III) apresenta os resultados da estimação pelo método de efeitos fixos,⁵ tradicional para modelos em painel. O estimador de efeitos fixos reconhece a existência de correlação entre a heterogeneidade individual não observada e os regressores do modelo, e permite eliminar o viés associado à omissão de tais efeitos específicos individuais (Wooldridge, 2002). Por sua vez, a identificação dos coeficientes dos efeitos de localização, por meio deste estimador, requer que exista mobilidade suficiente dos trabalhadores entre as RMs ou em direção a elas.

5. As variáveis que são fixas ou possuem variação constante no tempo, tais como a idade e o gênero do trabalhador, serão retiradas das especificações de estimação que incluem os efeitos fixos individuais.

Na coluna (IV), são apresentados os resultados da estimação condicional, quando são controlados (*absorbed*) apenas os efeitos fixos de indivíduos. A coluna (V), por sua vez, apresenta os resultados para o modelo que inclui apenas os efeitos de firmas. Ambos os modelos incluem os efeitos fixos de localização e as características observadas dos trabalhadores.

A coluna (VI) apresenta as estimativas obtidas por meio da transformação intratrabalhador-firma, como no método consistente de Abowd, Kramarz e Margolis (1999). A diferença em relação a Abowd, Kramarz e Margolis (*op. cit.*) é que eles usam a primeira diferença em vez de desvios da média. Nesta coluna, as unidades de análise são as combinações (*spells*) existentes de firma e trabalhador (*i, j*), a partir das quais é realizada a transformação *within group*, com o estimador de efeitos fixos. Dentro de cada *spell*, os efeitos fixos de firma e indivíduo não variam, de modo que a transformação *within* vai eliminar tais efeitos. Neste sentido, quaisquer variáveis que sejam constantes dentro de cada *spell* não serão identificadas (Andrews, Schank e Upward, 2006).

TABELA 3
Resultados das estimações econométricas da equação salarial (1995-2008)

Variáveis	Variável dependente: logaritmo natural do salário real ($\ln(w)$)					
	(I)	(II)	(III)	(IV)	(V)	(VI)
	POLS (<i>pooled</i> OLS)	Condicional efeito fixo individual primeiro ¹	Efeitos fixos	Intraindivíduo	Intrafirma ²	Consistente intraindivíduo-firma
Constante	5,2116*** (0,0068)	6,9228*** (0,0004)	6,8862*** (0,0028)	6,8862*** (0,0004)	5,7878*** (0,0003)	6,9308*** (0,0030)
Características do trabalhador						
<i>educ2</i>	0,2708*** (0,0033)	0,0335*** (0,0026)	0,0340*** (0,0027)	0,0340*** (0,0027)	0,0671*** (0,0026)	0,0399*** (0,0027)
<i>educ3</i>	0,7604*** (0,0033)	0,0541*** (0,0026)	0,0532*** (0,0027)	0,0532*** (0,0027)	0,3972*** (0,0026)	0,0642*** (0,0027)
<i>educ4</i>	1,4324*** (0,0033)	0,1507*** (0,0027)	0,1601*** (0,0028)	0,1601*** (0,0028)	0,9292*** (0,0026)	0,1544*** (0,0028)
<i>idade</i>	0,0610*** (0,0003)	-	-	-	0,0414*** (0,0002)	-
<i>idade2</i>	-0,0007*** (0,0000)	-	-	-	-0,0004*** (0,0000)	-
<i>exp</i>	0,0002*** (0,0000)	0,0007*** (0,0000)	0,0010*** (0,0000)	0,0010*** (0,0000)	0,0010*** (0,0000)	0,0004*** (0,0000)
<i>exp2</i>	0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	-0,0000*** (0,0000)	0,0000*** (0,0000)	-0,0000*** (0,0000)
<i>dsexo</i>	-0,5249*** (0,0007)	-	-	-	-0,2079*** (0,0006)	-

(Continua)

(Continuação)

Variável dependente: logaritmo natural do salário real ($\ln(w)$)						
Variáveis	(I)	(II)	(III)	(IV)	(V)	(VI)
	POLS (<i>pooled</i> OLS)	Condicional efeito fixo individual primeiro ¹	Efeitos fixos	Intra-indivíduo	Intra-firma ²	Consistente intra-indivíduo-firma
Efeito de localização: RM						
1. Baixada Santista	0,3667*** (0,0034)	-0,0321*** (0,0123)	0,0170** (0,0075)	0,0170** (0,0075)	-0,0421** (0,0195)	-0,0310** (0,0121)
2. Belém	-0,1825*** (0,0027)	-0,0305* (0,0175)	-0,0161** (0,0075)	-0,0161** (0,0075)	-0,0658** (0,0278)	-0,0289* (0,0171)
3. Belo Horizonte	0,2832*** (0,0015)	0,0063 (0,0054)	0,0253*** (0,0033)	0,0253*** (0,0033)	-0,0064 (0,0085)	0,0065 (0,0053)
4. Campinas	0,5209*** (0,0025)	-0,0034 (0,0075)	0,0396*** (0,0042)	0,0396*** (0,0042)	-0,0183 (0,0117)	-0,0043 (0,0075)
5. Curitiba	0,2296*** (0,0018)	-0,0438*** (0,0097)	0,0749*** (0,0041)	0,0749*** (0,0041)	-0,0391*** (0,0148)	-0,0347*** (0,0096)
6. Florianópolis	0,1754*** (0,0027)	0,0040 (0,0191)	0,0733*** (0,0071)	0,0733*** (0,0071)	0,0068 (0,0300)	-0,0050 (0,0188)
7. Fortaleza	-0,2073*** (0,0023)	-0,0455*** (0,0100)	-0,0139** (0,0066)	-0,0139** (0,0066)	-0,0424*** (0,0159)	-0,0468*** (0,0098)
8. Goiânia	0,0334*** (0,0025)	-0,0255* (0,0131)	-0,0407*** (0,0072)	-0,0407*** (0,0072)	-0,0631*** (0,0210)	-0,0247* (0,0129)
9. João Pessoa	-0,7894*** (0,0032)	-0,0104 (0,0277)	0,0057 (0,0099)	0,0057 (0,0099)	0,0139 (0,0444)	-0,0123 (0,0269)
10. Londrina	0,0966*** (0,0043)	0,0112 (0,0233)	-0,0177* (0,0101)	-0,0177* (0,0101)	-0,0022 (0,0371)	0,0197 (0,0228)
11. Macapá	0,5833*** (0,0085)	0,0280 (0,0501)	0,1434*** (0,0216)	0,1434*** (0,0216)	0,0826 (0,0795)	-0,0024 (0,0523)
12. Maceió	-0,1372*** (0,0041)	0,0314** (0,0131)	0,0688*** (0,0101)	0,0688*** (0,0101)	-0,0047 (0,0212)	0,0360*** (0,0127)
13. Maringá	-0,0038 (0,0058)	-0,1170*** (0,0309)	-0,0770*** (0,0125)	-0,0770*** (0,0125)	-0,1042** (0,0491)	-0,0959*** (0,0304)
14. Natal	-0,1173*** (0,0050)	0,0188 (0,0199)	-0,0080 (0,0079)	-0,0080 (0,0079)	0,0365 (0,0319)	0,0096 (0,0195)
15. Norte/Nordeste Catarinense	0,4250*** (0,0056)	-0,0476** (0,0204)	0,0575*** (0,0112)	0,0575*** (0,0112)	-0,0501 (0,0313)	-0,0375* (0,0200)
16. Porto Alegre	0,2615*** (0,0017)	0,0744*** (0,0036)	0,0401*** (0,0030)	0,0401*** (0,0030)	0,0937*** (0,0057)	0,0738*** (0,0035)

(Continua)

(Continuação)

Variáveis	Variável dependente: logaritmo natural do salário real ($\ln(w)$)					
	(I)	(II)	(III)	(IV)	(V)	(VI)
	POLS (<i>pooled</i> OLS)	Condicional efeito fixo individual primeiro ¹	Efeitos fixos	Intraindividuo	Intrafirma ²	Consistente intraindividuo-firma
17. Recife	-0,0666*** (0,0026)	0,0230* (0,0119)	0,0235*** (0,0063)	0,0235*** (0,0063)	0,0369* (0,0189)	0,0245** (0,0117)
18. Rio de Janeiro	0,1914*** (0,0013)	-0,0210*** (0,0057)	0,0173*** (0,0029)	0,0173*** (0,0029)	-0,0008 (0,0088)	-0,0029 (0,0058)
19. Salvador	-0,0372*** (0,0020)	0,2384*** (0,0084)	0,1218*** (0,0049)	0,1218*** (0,0049)	0,2302*** (0,0130)	0,2426*** (0,0082)
20. São Luís	-0,1413*** (0,0035)	-0,1154*** (0,0087)	0,0143** (0,0064)	0,0143** (0,0064)	-0,3033*** (0,0140)	-0,0901*** (0,0087)
21. São Paulo	0,4239*** (0,0009)	0,0100*** (0,0031)	0,0419*** (0,0020)	0,0419*** (0,0020)	0,0192*** (0,0048)	0,0109*** (0,0031)
22. Vale do Aço	0,3838*** (0,0056)	0,0799*** (0,0303)	0,0014 (0,0144)	0,0014 (0,0144)	-0,0190 (0,0440)	0,0727** (0,0298)
23. Vale do Itajaí	0,2677*** (0,0049)	-0,0000 (0,0177)	0,0065 (0,0104)	0,0065 (0,0104)	0,0018 (0,0282)	0,0029 (0,0172)
24. Vitória	0,1288*** (0,0034)	0,0369** (0,0158)	-0,0015 (0,0076)	-0,0015 (0,0076)	0,0566** (0,0253)	0,0374** (0,0154)
Dummies de tempo	Sim	Sim	Sim	Sim	Sim	Sim
<i>Sorting</i>						
Efeito de indivíduo	-	Sim	Sim	Sim	-	-
Efeito de firma	-	Sim	-	-	Sim	-
Efeito indivíduo-firma	-	-	-	-	-	Sim
Observações	4.665.388	4.665.388	4.665.388	4.665.388	4.665.388	4.665.388
R ²	0,411	0,915	0,225	0,893	0,758	0,241
R ² ajustado	0,411	0,906	0,165	0,885	0,751	0,131

Fonte: Rais-Migra/MTE.

Elaboração da autora.

Notas: ¹ O método condicional com efeitos de firma estimados primeiro gerou estimativas similares ao modelo com efeitos de indivíduos estimados primeiro, e ao modelo estimado com uma rotina alternativa (*azreg6*).² O método intrafirma gerou estimativas similares ao método de efeitos fixos de firmas.

Obs.: 1. O erro padrão encontra-se abaixo de cada estimativa, entre parênteses.

2. *, ** e *** = significativo a 5%, a 10% e a 1%, respectivamente.

Todas as estimações incluíram os efeitos de localização e de tempo, variando somente na inclusão das características observadas dos trabalhadores e heterogeneidade não observada de trabalhador e firma. As características observadas dos trabalhadores formais foram estatisticamente significantes e apresentaram os sinais esperados de acordo com a literatura que lida com equações de salários. As estimativas da idade e experiência apresentaram efeitos positivos sobre os salários reais, porém de maneira não linear, sugerindo a tendência de os salários crescerem a taxas decrescentes com a idade e a experiência. As mulheres receberam salário real em média 52% abaixo dos salários reais dos homens. O retorno à educação, mensurado pelas variáveis *dummies* para os ciclos escolares, é reduzido significativamente após o controle dos efeitos fixos de firma, de trabalhador ou de ambos, em relação à categoria de referência (analfabetos). O retorno à educação variou entre 27% e 3,4% para o ensino fundamental, entre 76% e 3,9% para o ensino médio e entre 143% e 15% para o ensino superior.

Os resultados encontrados mostram que o efeito fixo de indivíduo foi mais relevante para explicar a variação dos salários reais, em logaritmo, do que o efeito de firma. Comparando o R^2 ajustado da coluna (IV) com o R^2 da coluna (V) é possível perceber que, considerando os mesmos controles em ambas as regressões, a inclusão da heterogeneidade individual não observada torna a regressão capaz de explicar cerca de 88% da variação de salários reais do trabalho formal no Brasil, enquanto 75% dessa variação seria explicada na regressão que inclui somente os efeitos de firma. A inclusão de ambas as heterogeneidades não observadas, de firma e trabalhador, dados os demais controles, responde por cerca de 91% da variação de salários reais, conforme R^2 da coluna (II). A especificação da coluna (I), que desconsidera os efeitos fixos de firma e trabalhador, explica somente 41% da variação salarial real, o que sugere a importância das heterogeneidades não observadas sobre os diferenciais salariais.

As especificações que controlam apenas os efeitos individuais explicaram entre 16% e 88% da variação salarial individual, enquanto a especificação que inclui apenas os efeitos de firma explicou 75% da mesma variação. No estudo de Abowd, Kramarz e Margolis (1999) os componentes individuais também mostraram maior contribuição para explicar os diferenciais de salários individuais da França do que os efeitos de firma. Por sua vez, de acordo com o trabalho, os resultados do método condicional com efeito individual estimado primeiro e do método consistente explicaram 77% e 84%, respectivamente, da variação de salários individuais da França. Por sua vez, a evidência deste artigo para o Brasil explica uma parcela maior da variação de salários individuais do que o estudo para a França, com o método condicional, correspondendo a 91% da variação salarial. Isso pode sugerir que os efeitos não observados individuais são mais importantes para a determinação

de salários do Brasil e podem obter maior remuneração no mercado de trabalho desse país em comparação com a França.

O método condicional que inclui os efeitos fixos de indivíduo e firma, descrito na coluna (II), apresentou estimativas muito similares ao método de efeitos fixos intratrabalhador-firma, da coluna (VI). A variação do R^2 entre as colunas é explicada pela forma com a qual cada método estima os efeitos fixos. O resultado pode sugerir que ambos os métodos seriam capazes de identificar estimativas similares dos parâmetros subjacentes aos modelos. Por sua vez, o método de efeitos fixos de indivíduo e o método condicional intraindivíduo, apresentados nas regressões das colunas (III) e (IV), respectivamente, encontraram estimativas idênticas. Portanto, os métodos condicionais devem encontrar estimativas iguais às do método de efeitos fixos tradicional, mas podem ser adotados para incluir simultaneamente um grupo maior de efeitos fixos.

A estimação dos efeitos fixos de localização assumiu os municípios que não pertenciam a alguma das 24 RMs como a categoria de referência. Analisando os resultados encontrados para os efeitos de localização nas RMs do Brasil, é possível perceber que oito entre as 24 RMs apresentaram estimativas significantes e sem variações na direção (sinal) do efeito, em todos os métodos que controlam a heterogeneidade não observada de firma ou de trabalhador. São essas as RMs de Maringá (-7,7% a -11,7%), de Goiânia (-2,5% a -6,3%), de Belém (-1,6% a -6,6%), de Fortaleza (-1,4% a -4,7%), de São Paulo (1% a 4,2%), do Recife (2,3% a 3,7%), de Porto Alegre (4% a 9,4%) e de Salvador (12,2% a 24,3%).

A estimação por POLS encontrou um prêmio salarial urbano em quinze RMs: Macapá (58,3%), Campinas (52,1%), Norte/Nordeste Catarinense (42,5%), São Paulo (42,4%), Vale do Aço (38,4%), Baixada Santista (36,7%), Belo Horizonte (28,3%), Vale do Itajaí (26,8%), Porto Alegre (26,1%), Curitiba (22,9%), Rio de Janeiro (19,1%), Florianópolis (17,5%), Vitória (12,9%), Londrina (9,7%) e Goiânia (3,3%). No modelo que obteve o melhor desempenho na explicação da variação salarial, o qual controla os efeitos de firma e trabalhador – coluna (II) –, apenas sete RMs apresentaram prêmio salarial urbano: Salvador (23,8%), Vale do Aço (7,9%), Porto Alegre (7,4%), Vitória (3,7%), Maceió (3,1%), Recife (2,3%) e São Paulo (1%). Portanto, após considerar a heterogeneidade não observada de firmas e trabalhadores, o aparente prêmio salarial urbano desapareceu em nove RMs do Sul e do Sudeste e na RM de Macapá, e sua magnitude foi significativamente reduzida em três RMs do Sudeste e uma do Sul. Os resultados sugerem que grande parte dos efeitos positivos de localização sobre os salários reais, em áreas urbanas densas, resulta da heterogeneidade não observada de firmas e trabalhadores formais, particularmente no Sul e no Sudeste do Brasil.

4.2 Diferenciais salariais entre RMs: decomposição dos efeitos fixos de localização

Na tabela 4 são apresentados os resultados que permitem decompor os efeitos fixos de localização em componentes associados às heterogeneidades não observadas de trabalhadores e firmas, tal como proposto por Abowd, Kramarz e Margolis (1999) para os diferenciais salariais entre setores na França. Para decompor os efeitos de localização entre efeitos de firma e de trabalhador foi necessário estimar cada efeito a partir da equação de salários individuais, utilizando as mesmas variáveis observadas em cada regressão. Nesse sentido, os efeitos fixos de localização foram estimados pelo método POLS a partir da equação de salários individuais sem a inclusão dos efeitos de firma e indivíduo, controlando apenas as *dummies* de educação, a experiência e os efeitos de tempo. Do mesmo modo, os efeitos fixos de indivíduos e firmas foram estimados simultaneamente pelo método condicional a partir da equação salarial, considerando o mesmo grupo de controles utilizados na estimação dos efeitos de localização e os incluindo.

TABELA 4
Resultados da decomposição dos efeitos de localização

Efeito fixo de RM	(I) $\hat{\alpha}$	(II) $\hat{\alpha}$	(III) $\hat{\alpha}$	(IV) $\hat{\alpha}$	(V) $\hat{\alpha}$	(VI) $\hat{\alpha}$
Média ($\hat{\theta}$)	0,0789 (0,1262)	-	0,6865*** (0,0533)	0,7429*** (0,1198)	-	0,7808*** (0,0566)
Média ($\hat{\gamma}$)	-	1,1450*** (0,0370)	1,2594*** (0,0313)	-	1,1723*** (0,0450)	1,1822*** (0,0356)
Norte	-0,1747*** (0,0491)	-0,1674*** (0,0228)	-0,0674*** (0,0201)	0,0458 (0,0451)	-0,1743*** (0,0237)	-0,0303 (0,0214)
Nordeste	-0,5646*** (0,0388)	-0,1833*** (0,0199)	-0,0125 (0,0209)	-0,3364*** (0,0379)	-0,1797*** (0,0202)	-0,0011 (0,0205)
Sul	-0,1477*** (0,0304)	-0,0207 (0,0157)	-0,0016 (0,0129)	-0,1700*** (0,0255)	-0,0160 (0,0163)	-0,0150 (0,0129)
Centro-Oeste	-0,3585*** (0,0616)	-0,0706** (0,0318)	0,0108 (0,0266)	-0,1561*** (0,0543)	-0,0725** (0,0319)	0,0283 (0,0262)
Setor público	-	-	-	-0,1935*** (0,0165)	0,0113 (0,0106)	-0,0383*** (0,0091)
<i>Dummies</i> de tempo	Sim	Sim	Sim	Sim	Sim	Sim
Constante	0,3474*** (0,0496)	0,2283*** (0,0239)	0,1303*** (0,0208)	0,9030*** (0,0630)	0,1882*** (0,0446)	0,2537*** (0,0356)
Observações	336	336	336	336	336	336
R^2	0,557	0,890	0,928	0,691	0,890	0,931
R^2 ajustado	0,532	0,883	0,923	0,673	0,883	0,927

Fonte: Rais-Migra/MTE.

Elaboração da autora.

Obs.: 1. O erro-padrão encontra-se abaixo de cada estimativa, entre parênteses.

2. *, ** e *** = significativo a 5%, a 10% e a 1%, respectivamente.

O painel utilizado para a decomposição estatística é construído a partir das 24 RMs do Brasil, ao longo do período de análise. Todas as estimações da tabela 4 foram realizadas por meio do método POLS de painel. Para a decomposição, foram obtidos os efeitos fixos médios locais dos indivíduos e das firmas. O modelo de decomposição incluiu, ainda, variáveis *dummies* para as macrorregiões do Brasil, para lidar com as significativas diversidades produtivas regionais existentes no país. A categoria de referência foi a região Sudeste, considerada a mais dinâmica do país.

Todas as especificações da tabela 4 controlam os efeitos temporais, sendo o ano de 1995 o período de referência. A coluna (I) da tabela apresenta a regressão dos efeitos fixos de localização em RMs ($\hat{\alpha}$) sobre os efeitos individuais médios por RMs ($\hat{\theta}$), além dos controles indicados. A coluna (II) inclui apenas os efeitos médios de firma ($\hat{\gamma}$), além dos controles considerados. Por fim, a coluna (III) decompõe os efeitos de localização em efeitos médios de trabalhador e firma, controlando as diferenças regionais e temporais existentes entre as RMs no período considerado.

Os resultados mostram que as heterogeneidades não observadas das firmas e dos trabalhadores formais respondem estatisticamente por quase toda a variação dos efeitos de localização em RMs, conforme R^2 ajustado de 92% na coluna (III) da tabela 4. Estes resultados sugerem que praticamente todos os ganhos salariais atribuídos à localização em áreas urbanas densas podem decorrer de características específicas e não observadas das firmas e dos trabalhadores concentrados nessas áreas. Os efeitos residuais que ainda restaram, em torno de 8%, poderiam resultar dos atributos locais associados às aglomerações urbanas.

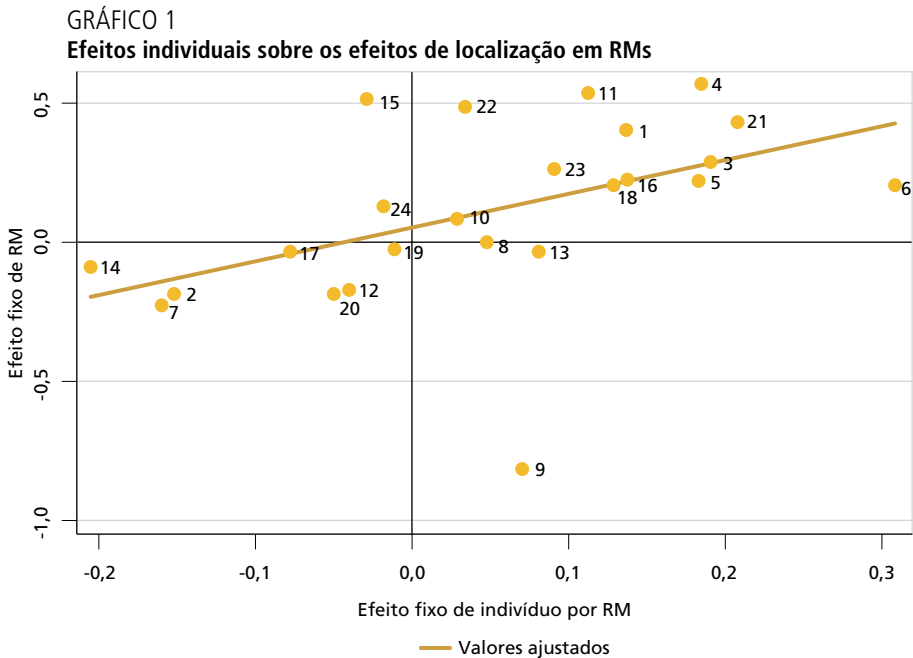
A avaliação da importância relativa dos efeitos fixos sugere que as heterogeneidades de firmas mostraram-se mais relevantes para explicar os diferenciais salariais no mercado formal entre as RMs do que as heterogeneidades individuais. Os efeitos de firma explicaram cerca de 88% da variação dos efeitos de localização sobre os salários, conforme R^2 ajustado da coluna (II) da tabela 4. Por sua vez, os efeitos fixos dos trabalhadores explicaram 53% da variação dos efeitos de localização em RMs, conforme R^2 ajustado da coluna (I). No entanto, vale ressaltar que ambos respondem separadamente por proporções bastante altas da variação salarial entre RMs, medida pelos seus efeitos de localização. Neste aspecto, a importância relativa dos efeitos fixos dos trabalhadores não pode ser reduzida.

No estudo de Abowd, Kramarz e Margolis (1999), sobre a decomposição dos diferenciais salariais entre setores da França, os componentes individuais apresentaram maior poder explicativo (90%) do que os componentes de firma (7%-25%). Ambos os componentes explicaram entre 85% e 96% do diferencial setorial de salários. Por sua vez, considerando a aplicação da decomposição para os diferenciais de salários entre localizações do Brasil, este artigo encontrou evidências

de maior contribuição dos efeitos de firma, embora o estudo de Abowd, Kramarz e Margolis (*op. cit.*) tenha focado os diferenciais entre setores.

Por outro lado, é possível que a presença do setor público no banco de dados explique o maior poder explicativo dos efeitos de firma em relação aos efeitos individuais. Abowd, Kramarz e Margolis (1999) excluíram o setor público em sua análise para a França. Em geral, os salários reais médios do setor público do Brasil estão acima dos salários privados (Braga, Firpo e Gonzaga, 2009). Isso pode ocorrer já que a determinação salarial no setor público não segue os critérios de mercado. Nesse sentido, as áreas com maior parcela de empregos públicos podem apresentar um diferencial de salários maior associado às firmas. Desse modo, os resultados nas colunas (IV), (V) e (VI) incluem uma *dummy* para controlar a proporção de empregos no setor público em cada RM na amostra. O controle da proporção de empregos públicos locais reduziu a diferença entre o poder explicativo dos efeitos de firma (88%) e de indivíduo (67%), mas o componente não observado das firmas ainda exerceu maior contribuição para explicar os diferenciais salariais das RMs.

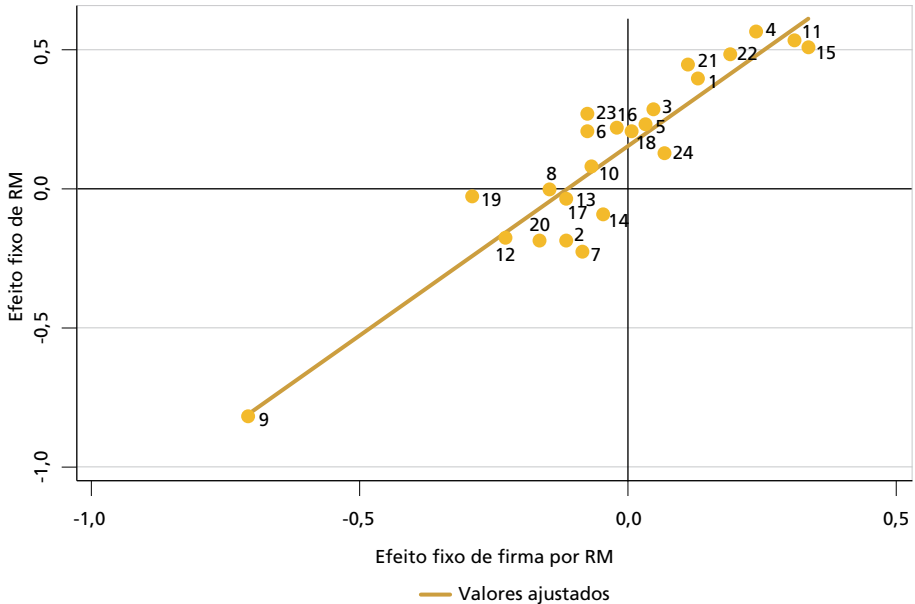
Os gráficos 1 e 2 mostram a força da relação entre os componentes não observados dos indivíduos e das firmas e os efeitos de localização nas RMs.



Elaboração da autora.

GRÁFICO 2

Efeitos de firmas sobre os efeitos de localização em RMs



Elaboração da autora.

Os gráficos apresentam a relação entre os efeitos fixos médios de trabalhador e firma e os efeitos de localização das RMs, sem controlar as variáveis regionais e temporais incluídas nas regressões. Um ajuste linear entre as variáveis é incluído nos gráficos. Com base neles, é possível perceber que os efeitos de firma possuem uma relação mais forte com os efeitos de localização do que os efeitos de trabalhadores. Essas evidências são compatíveis com os resultados encontrados na decomposição dos efeitos de localização.

Os resultados gerais encontrados neste artigo mostraram que, embora os efeitos fixos individuais sejam os componentes mais importantes para explicar a variação dos salários reais do trabalhador formal do Brasil, entre 1995 e 2008, tais efeitos apresentaram menor contribuição relativa em comparação com os efeitos de firma para explicar a variação salarial associada à localização em RMs. No entanto, ambos os componentes relacionados à heterogeneidade não observada de firma e indivíduos possuem um peso substancial para explicar a variação de salários e dos efeitos de localização. A decomposição estatística mostrou que apenas 8% da variação salarial entre as áreas metropolitanas pode ser explicada por fatores específicos locais, enquanto os componentes específicos de firma e indivíduo explicam quase todo o diferencial de salários locais no período de análise.

5 CONSIDERAÇÕES FINAIS

O objetivo deste artigo foi avaliar a contribuição da heterogeneidade não observada de firma e indivíduo para os efeitos de localização sobre os salários e para a variação salarial do trabalhador formal do Brasil. O referencial teórico mostrou que os efeitos positivos da aglomeração urbana sobre os salários podem ser confundidos com a concentração de trabalhadores e firmas mais produtivos nos grandes centros urbanos, caracterizando um processo de *sorting* entre localizações.

Para encontrar a contribuição dos efeitos de firma e trabalhador sobre os salários reais e sobre os efeitos de localização foram estimadas equações de salários com as características observadas dos trabalhadores e os efeitos fixos de localização, trabalhador e firma. A estratégia empírica recorreu aos métodos e avanços da literatura recente em economia do trabalho, para lidar com múltiplos efeitos fixos em grandes bancos de microdados pareados de trabalhador-firma, baseados no estudo de Abowd, Kramarz e Margolis (1999).

Os resultados mostraram que as características observadas dos trabalhadores formais, os efeitos temporais e os efeitos de localização em RMs explicavam 41% da variação de salários reais dos indivíduos. A inclusão da heterogeneidade não observada de trabalhadores e firmas ampliou a variação explicada de salários reais para 91%. Os resultados das estimações considerando somente os efeitos de indivíduos (*absorbed*) conseguiram explicar cerca de 88% da variação de salários, enquanto a estimação considerando somente os efeitos não observados de firma foi capaz de explicar cerca de 75% de tal variação, dados os mesmos controles observados em ambas as regressões.

A estimação pelo modelo POLS encontrou um prêmio salarial urbano em quinze RMs. Após controlar os efeitos de firma e trabalhador, apenas sete RMs apresentaram prêmio salarial urbano. Ao considerar a heterogeneidade não observada de firmas e trabalhadores, o aparente prêmio salarial urbano desapareceu em dez RMs, e sua magnitude foi significativamente reduzida em quatro RMs. As evidências indicaram que grande parte dos efeitos positivos de localização, em áreas urbanas densas, sobre os salários reais do trabalhador, resultava da heterogeneidade não observada das firmas e dos trabalhadores dessas áreas, particularmente no Sul e no Sudeste do Brasil.

As estimativas dos efeitos fixos de localização das RMs foram explicadas quase inteiramente (92%) pelos efeitos fixos de trabalhadores e firmas, controlando para as diferenças entre as macrorregiões e os efeitos de tempo. Os resultados sugerem que uma grande parcela dos diferenciais salariais atribuídos à localização em aglomerações urbanas decorre das heterogeneidades não observadas das suas firmas e dos seus trabalhadores. Por sua vez, os efeitos residuais, em torno de 8%, podem resultar dos atributos locais associados às aglomerações urbanas. No entanto, as heterogeneidades

de firmas foram relativamente mais relevantes para os diferenciais salariais entre as RMs no mercado de trabalho formal do Brasil. Os efeitos de firma explicaram cerca de 88% da variação dos efeitos de localização, enquanto os efeitos fixos dos trabalhadores formais explicaram cerca de 53% dessa variação.

O estudo de Abowd, Kramarz e Margolis (1999) para a França encontrou maior poder explicativo dos efeitos individuais para a variação salarial entre os setores, excluindo o setor público. Embora o presente estudo não tenha buscado decompor os diferenciais de salários setoriais, mas os diferenciais de localização, é possível que a presença do setor público nos dados do Brasil explique a maior importância relativa dos efeitos de firma. No entanto, a inclusão de uma variável de controle para a proporção de empregos públicos reduziu a diferença entre as contribuições dos componentes de indivíduos e firmas, mas os efeitos de firma permaneceram com maior poder explicativo.

Portanto, as evidências encontradas neste artigo, para o mercado de trabalho formal do Brasil, entre 1995 e 2008, mostraram que as heterogeneidades não observadas de firmas e trabalhadores são componentes fundamentais na determinação dos salários reais dos indivíduos e respondem por grande parcela dos efeitos de localização sobre os salários. Os efeitos fixos individuais foram mais importantes para explicar a variação dos salários reais individuais, mas relativamente menos relevantes do que os efeitos de firma para explicar os diferenciais salariais de localização em RMs. No entanto, ambos os efeitos de firma e trabalhador responderam por parcelas estatisticamente substanciais da variação de salários individuais e por localização em áreas de aglomeração urbana.

REFERÊNCIAS

- ABOWD, J. M. *et al.* The relative importance of employer and employee effects on compensation: a comparison of France and the United States. **Journal of the Japanese and International Economies**, v. 15, n. 4, p. 419-436, 2001.
- ABOWD, J. M.; CREECY, R. H.; KRAMARZ, F. **Computing person and firm effects using linked longitudinal employer-employee data**. Maryland: US Census Bureau, 2002. (Longitudinal Employer-Household Dynamics Technical Papers, n. TP-2002-06). Disponível em: <<https://goo.gl/G5jGgy>>. 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.
- 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.

BRAGA, B. G.; FIRPO, S.; GONZAGA, G. Escolaridade e diferencial de rendimentos entre o setor privado e o setor público no Brasil. **Pesquisa e Planejamento Econômico**, v. 39, n. 3, p. 431-464, dez. 2009.

BRASIL. Ministério do Trabalho e Emprego. **Microdados da Relação Anual de Informações Sociais (Rais)**. Brasília: MTE, 1990.

_____. _____. **Microdados da Relação Anual de Informações Sociais (Rais)**. Brasília: MTE, 1997.

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., Foz do Iguaçu, 2009. **Anais...** Foz do Iguaçu: Anpec, 2009.

COMBES, P-P. *et al.* Sorting and local wage and skill distributions in France. **Journal of Regional Science and Urban Economics**, n. 42, p. 913-930, 2012.

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

CORREIA, S. **Explanation of the HDFE iteration with 3 FEs**. Boston: Boston College, 2014. Disponível em: <<https://goo.gl/3vJ7KE>>. Acesso em: 3 jan. 2015.

ENTORE, H.; GOLLAC, M.; KRAMARZ, F. New technologies, wages, and worker selection. **Journal of Labor Economics**, v. 17, n. 3, p. 464-491, 1999.

FREGUGLIA, R. S. **Efeitos da migração sobre os salários no Brasil**. 2007. Tese (Doutorado em Economia) – Universidade de São Paulo, São Paulo, 2007.

FREGUGLIA, R. S.; MENEZES-FILHO, N. A. **Inter-regional wage differentials with individual heterogeneity**: evidence from Brazil. São Paulo: Insper, 2011. (Insper Working Paper, n. 231). Disponível em: <<https://goo.gl/fScYc2>>. Acesso em: 14 dez. 2011.

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**. Massachusetts: NBER, 1994. (NBER Working Papers, n. 4728).

_____; _____. Cities and skills. **Journal of Labor Economics**, v. 19, n. 2, p. 316-342, 2001.

GOUX, D.; MAURIN, E. Persistence of interindustry wage differentials: a reexamination using matched worker-firm panel data. **Journal of Labor Economics**, v. 17, n. 3, p. 492-533, 1999.

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.

HALFDANARSON, B.; HEUERMANN, D. F.; SÜDEKUM, J. **Human capital externalities and the urban wage premium: two literatures and their interrelations**. Bonn: IZA, 2008. (IZA Discussion Paper, n. 3493).

LEONARD, J. S.; VAN AUDENRODE, M. **Persistence of firm and individual wage components**. Berkeley: University of California, 1996. Disponível em: <<https://goo.gl/vvKvDM>>. Acesso em: 14 jan. 2015.

MCCAFFREY, D. F. *et al.* 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-FILHO, N. A.; MUENDLER, M. A.; RAMEY, G. The structure of worker compensation in Brazil, with a comparison to France and the United States. **Review of Economics and Statistics**, v. 90, n. 2, p. 324-346, 2008.

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

MITTAG, N. **New methods to estimate models with large sets of fixed effects with an application to matched employer-employee data from Germany**. Berlin: FDZ, 2012. (FDZ-Methodenreport, n. 1) Disponível em: <<https://goo.gl/aPWHPq>>. Acesso em: 3 jan. 2015.

OUAZAD, A. **A2REG: Stata module to estimate models with two fixed effects**. Boston: Boston College, 2008. Disponível em: <<https://goo.gl/QN8B69>>. Acesso em: 13 jul. 2018.

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

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, 16., Fortaleza, 2011. **Anais...** Fortaleza: Anpec, 2011.

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

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

SILVEIRA NETO, R.; AZZONI, C. Disparidades regionais de renda no Brasil: qual o papel das amenidades? *In*: ENCONTRO REGIONAL DE ECONOMIA, 9., 2004, Fortaleza. **Anais...** Fortaleza: Anpec Nordeste, 2004.

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data.** Cambridge: MIT, 2002.

APÊNDICE

ALGORITMO DE ITERAÇÃO DE PONTO FIXO

Correia (2014) desenvolveu um algoritmo implementável no *software* Stata para estimar a solução proposta por Guimarães e Portugal (2010) para os modelos com múltiplos efeitos fixos. O estudo assume a seguinte especificação geral:

$$y = X\beta + D_1\alpha_1 + D_2\alpha_2 + D_3\alpha_3 + \epsilon, \quad (1A)$$

onde D_i são as variáveis indicadoras e α_i os efeitos fixos correspondentes. A estratégia de iteração de ponto fixo é aplicada alternando-se entre a estimação de β e α_i nas equações de solução:

$$\begin{aligned} \beta &= (X'X)^{-1}X'(y - D_1\alpha_1 - D_2\alpha_2 - D_3\alpha_3), \\ \alpha_1 &= (D_1'D_1)^{-1}D_1'(y - X\beta - D_2\alpha_2 - D_3\alpha_3), \\ \alpha_2 &= (D_2'D_2)^{-1}D_2'(y - X\beta - D_1\alpha_1 - D_3\alpha_3), \\ \alpha_3 &= (D_3'D_3)^{-1}D_3'(y - X\beta - D_1\alpha_1 - D_2\alpha_2). \end{aligned} \quad (2A)$$

onde $(D_i'D_i)^{-1}D_i'$ é obtido como uma média de grupos dos resíduos da regressão de y sobre X . A estimação de β seria dada pela regressão de y (transformado) sobre X . No entanto, a implementação mantém y como variável dependente e inclui $D_i\alpha$ como covariada. Quando o procedimento de estimação converge, o vetor $D_i\alpha$ deve conter as estimativas dos coeficientes das variáveis indicadoras. Definindo $Z_2 := D_2 \hat{\alpha}_2$ e $Z_3 := D_3 \hat{\alpha}_3$ e supondo valores para a iteração, o procedimento desenvolvido por Correia (2014) é dado como:

- a) calcula $P_1 y$ e $\tilde{y} := M_1 y$;
- b) inicia com $Z_2^{(0)} = Z_3^{(0)} = 0$;
- c) até a convergência de Z_2 e Z_3 , são definidos:
 - $Z_2^{(n)} = P_2 \left[\tilde{y} + P_1 \left(Z_2^{(n-1)} + Z_3^{(n-1)} \right) - Z_3^{(n-1)} \right]$;
 - $Z_3^{(n)} = P_3 \left[\tilde{y} + P_1 \left(Z_2^{(n)} + Z_3^{(n-1)} \right) - Z_2^{(n)} \right]$.
- d) calcula $Z_1 = P_1 (y - Z_2 - Z_3)$ e, com ele, computa $y^* = y - Z_1 - Z_2 - Z_3$;
- e) após repetir os estágios de (a) a (d) para cada variável, efetua a regressão das variáveis transformadas para obter $\hat{\beta}: y^* = X^* \hat{\beta} + e$;

- f) para obter os efeitos fixos, utiliza $y = X\hat{\beta} + Z_1 + Z_2 + Z_3 + e$, com $\hat{\beta}$ e e do estágio (e). Computa $\tilde{e} := y - X\hat{\beta} = Z_1 + Z_2 + Z_3 + e$ e aplica o estágio anterior para obter Z_1 , Z_2 e Z_3 .

