Location and wages: the contribution of firm and worker effects in Brazil[1]

Diana Lúcia Gonzaga da Silva
Universidade de São Paulo, IPE-USP

Carlos R. Azzoni
Universidade de São Paulo, Departamento de Economia, IPE-USP

Resumo: O objetivo deste artigo é avaliar a contribuição da heterogeneidade não observada de indivíduos e firmas para os efeitos de localização sobre os salários e para a variação de salários no Brasil. No primeiro estágio, os efeitos de localização são estimados através de uma equação salarial, controlando as características observadas dos trabalhadores e a heterogeneidade não observada de trabalhadores e firmas. No segundo estágio, os efeitos de localização estimados são regredidos sobre os efeitos fixos de firma e trabalhador. Foi utilizado um painel de microdados para o período de 1995-2008 (RAIS-Migra). O modelo proposto por Abowd et al. (1999) para a decomposição salarial foi estimado, para lidar com múltiplos efeitos fixos em grandes bancos de dados pareados de trabalhador e firma. Uma contribuição deste artigo é lidar com mais controles do que o usual neste tipo de análise. Para a literatura nacional, uma importante contribuição é o controle simultâneo de efeitos de firma e trabalhador. Os resultados mostram que os efeitos de firma e trabalhador são responsáveis por substancial variação de salários entre indivíduos (93%) e para a variação nos efeitos de localização entre áreas metropolitanas (95%). No primeiro e segundo estágios, as características individuais são mais importantes do que os efeitos de firma para explicar os diferenciais salariais (indivíduos 91%, firmas 80%) e os efeitos de localização (92%, 41%). Controlando para todos esses efeitos, o "puro" efeito de aglomeração pode ser somente de 5%. Portanto, ambos os efeitos respondem por parcela substancial da variação de salários reais e efeitos de localização sobre os salários no Brasil.

Palavras-chave: determinação salarial, *sorting*, efeitos de firma, efeitos de localização, efeitos de indivíduos.

Abstract: The objective of this paper is to assess the contribution of unobservable firm and individual heterogeneity for the location effects on wages and for the variation of wages in Brazil. In the first stage we estimate the effects of location through a wage equation, controlling for observable worker characteristics and unobserved heterogeneity of workers and firms. In a second stage, the estimated location effects are regressed on the fixed effects of firms and workers. We use micro data panel for the period 1995-2008 (RAIS-Migra). We estimate the model proposed by Abowd et al. (1999) for the wage decomposition, to deal with multiple fixed effects in large databases matching workers and firms. One contribution of this paper is to deal with more controls than usual in this type of analysis. As for the literature on the Brazilian case, the simultaneous control for firm and worker effects is also an important contribution. The findings show that firm and worker effects account for a substantial variation of wages across individuals (93%) and for the variation in location effects across metropolitan areas (95%). In the first and second stages individual characteristics are more important than firm effect to explain wage differentials (individuals 91%, firms 80%) and location effects (92%, 41%). Controlling for all these effects, the "pure" agglomeration effects would amount to only 5%. Therefore, both effects account for substantial shares of the variation of real wages and location effects on wages in Brazil.

Keywords: wage determination, sorting, firm effects, location effects, individual effects.

Área 10 - Economia Regional e Urbana
JEL: J24, J31, R23, C23

[1] Grant #2014/03077-7, São Paulo Research Foundation (FAPESP).

1. Introduction

Several studies in Labor Economics sought to understand wage disparities. Using mincerian equations, research in labor economics advanced in the control of observable and unobservable characteristics. However, a wage differential in favor of large urban centers remained, which motivated the emergence of integrated studies in regional, urban and labor economics to include agglomeration economies in the explanation of the wage differential. Thus, location became a wage determinant. In the Brazilian case, for example, the real wages of João Pessoa, a metropolitan area (MR) in the poor Northeast region, corresponded to 65% of the real wages of São Paulo MR in 2008. Wages of metropolitan areas in every state are higher than the respective state means, suggesting the existence of agglomeration gains in these areas. Comparing across metropolitan areas, even within the most dynamic regions of the Centre-South of the country, a wage differential can be detected. Several factors, such as differentials in cost of living, mobility, productivity, amenities, regional labor market segmentation (formal *vs.* informal, small *vs.* large firms, public *vs.* private sector) may explain the observed differences.

Controlling for the observed characteristics of workers is not sufficient to dispel the wage inequalities. Recent integrated studies in regional and labor economics include worker's unobserved skills, allowing for the correction of the selection bias associated with the attraction of skilled workers by major urban centers. This reduces much of the observed wage advantage associated with dense urban areas. Combes et al. (2008) have shown that the differences in the composition of skills were responsible for 40-50% of the spatial wage disparities of French workers, suggesting evidence of sorting by skills. Combes et al. (2012) showed a negative selection by skills in the migration of workers to less dense areas, and positive, for denser areas of France. Freguglia and Menezes-Filho (2011) showed that 63% of the wage differential in Brazil was explained by unobserved characteristics of workers. However, a wage differential persisted, even after considering cost of living, observed and unobserved skills of workers and occupational characteristics, indicating the existence of location-specific effects. The gains of agglomeration in the labor market may emerge from increased productivity of workers in the presence of positive externalities, such as access to greater diversity of occupations, the accumulation of human capital and knowledge spillovers, social interactions, etc. A broader market allows sharing of suppliers and workers with similar skills, in addition to enabling better matching between workers and firms. The denser areas also facilitate learning and the development of new technologies (PUGA, 2009).

The literature indicates wage premium ranging between 5% and 10% (HALFDANARSON et al., 2008). Glaeser and Maré (1994; 2001) indicate a positive relationship between wages and the size of the cities, with 32-33% higher wages for workers in metropolitan areas of the United States. A branch of research in the labor literature has emerged, incorporating urban agglomeration economies as a relevant factor to the determination of wages. The few existing studies in the Brazilian case show evidence of agglomeration gains: Rocha et al. (2011) found a 0.12 differential minimum wage for workers in metropolitan regions, after controlling for observable and unobservable characteristics; Campos and Silveira Neto (2009) found wage gains of 16% for workers in metropolitan areas in 2000.

The gains associated with the characteristics of workers and firms located in certain areas can be confused with gains of agglomeration. Combes et al. (2012) suggested the sorting by workers' skills and the differentiated returns of agglomeration on skills, as explanations for wage differentials. If the choice of location is not exogenous, unobserved attributes of firms and workers in dense areas can bias the estimates of the urban wage premium. Sorting by skills arises because workers with better skills tend to cluster in dense labor markets. Combes et al. (2008) have highlighted that unbiased estimates require the inclusion of fixed effects of firms, since these can be correlated with the effects of sectors. The selection of firms can arise if only the most productive survive in dense areas. Combes et al. (2008) and Combes et al. (2012) developed a unified approach of the determinants of earnings, including the effects of location, sectors and unobserved skills. However, data limitations precluded the adoption of controls associated with the observed characteristics of workers, such as education. Existing studies in Brazil control just part of the spatial sorting associated with unobserved workers heterogeneity.

The empirical limitation to estimate models with unobserved heterogeneity of workers and firms by least squares (full least square) arises from the computational complexity involved, since a large dimension of firm and worker identifiers in regressions of longitudinal data is required. Since Abowd et al (1999), several studies have sought computationally feasible alternatives to estimate the parameters of interest (ABOWD et al., 1999; 2002; ANDREWS et al., 2006; GUIMARÃES; PORTUGAL, 2010; McCaffrey, 2012; MITTAG, 2012; GAURE, 2013). Abowd et al (1999) developed alternative approaches to estimating the parameters in wage equations, given the computational limitations of the least squares solution. Their results indicate that unobserved individual heterogeneity is a very important source of wage variation in France, more relevant than the effects of firm, and both effects were weakly correlated. The fixed effects of workers explained about 90% of the sectoral wage differential in France.

This study seeks to investigate effect of location on wages in the Brazilian case, considering the presence of sorting, controlling for firm and worker effects. We use a matched panel of micro data of workers and firms to estimate wage equations, including the observed characteristics of workers and firms, the unobserved heterogeneity of workers and firms, and the effects of location in metropolitan regions. Besides this introduction, four additional sections compose this paper. The following section presents a wage decomposition model. The econometric model is presented in the third section, including the data base and the empirical strategy. The fourth section presents and discusses the results, and the last section presents the conclusions of the study.

2. A wage decomposition model

The statistical decomposition of wages is based in Abowd et al. (1999). Consider the wage equation:

$$(1) \quad y_{it} = \mu_y + (x_{it} - \mu_x)\beta + \theta_i + \psi_{J(i,t)} + \varepsilon_{it},$$

where y_{it} is the logarithm of wages of individual $i = 1, ..., N$, in period $t = 1, ..., T$; x_{it} is a time-varying vector of exogenous characteristics of individual; θ_i is the pure individual effect; $\psi_{J(i,t)}$ is the pure firm effect, for the firm employing individual i, at time t (indicated by $J(i,t)$); μ_y and μ_x are the general averages of x_{it} and y_{it}; and ε_{it} is the statistical residue. Assuming a random sample of N individuals observed in T years, ε_{it} will present the following properties:

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

$$(1.2) \quad cov[\varepsilon_{it},\varepsilon_{ns}|i,t,n,s,J(i,t),J(n,s),x_{it},x_{ns}\,] = \begin{cases} \sigma_\varepsilon^2 \;for\; i = n \;and\; t = s \\ 0 \;for\; i \neq n \;and\; t \neq s \end{cases}$$

In matrix notation, equation (1) becomes:

$$(2) \quad y = X\beta + D\theta + F\psi + \varepsilon,$$

where X is a (N*xP) matrix of time-varying observable characteristics (in deviations from the mean); D is a (N*xN) matrix of individual indicators; F is a (N*x mJ) matrix of the firm employing individual i in time t (J is the total number of firms); y is the (N*x1) vector of wages (in deviations from the mean); ε is the vector of residues; and $N^*=NT$. The parameters are the vectors β (Px1), θ (Nx1), ψ (mJx1) and the variance of the error σ^2_ε.

Equations (1) and (2) indicate the conditional wage expectation, given the observed characteristics and indicators of individuals and firms. Although some studies have developed least squares solutions for equation (2) with small samples (LEONARD et al., 1996; ENTORF et al., 1999; GOUX; MAURIN, 1999), the effort of recent literature has been finding feasible solutions to large databases matching firms and workers. Some studies estimate partial versions of equation (2), which leads to ambiguous interpretations of their parameters. Abowd et al. (1999) shows that, if the pure firm effect (ψ) is ignored, the individual effect will be the sum of the pure individual effect (θ) and an omitted variable bias term:

$$(3) \quad \theta^* = \theta + (D'M_X D)^{-1} D'M_X F\psi,$$

Where $M_A \equiv I - A(A'A)^{-1}A'$, for an arbitrary A matrix. The time-varying individual characteristics parameter (β*) will be the sum of the parameter β of the complete model with an omitted variable bias term:

$$(4) \quad \beta^* = \beta + (X'M_D X)^{-1} X'M_D F\psi.$$

If the pure individual effect (θ) is ignored, the firm effect will be the sum of the pure firm effect (ψ) and an omitted variable bias term:

$$(5) \quad \psi^{**} = \psi + (F'M_X F)^{-1} F'M_X D\theta.$$

Vector β** will be the sum of the β from the complete model to an omitted variable bias term:

$$(6) \quad \beta^{**} = \beta + (X'M_F X)^{-1} X'M_F D\theta.$$

Therefore, partial estimations of Equation (2) lead to biased effects of firms and individuals.

The literature on inter-industry wage differentials has pointed out that these cannot be explained just by the observed characteristics of workers and firms. However, the role of unobserved heterogeneity is still controversial. In Abowd et al. (1999), the pure industry effect is defined as the aggregation of the pure effects of firms in the respective industry, corresponding to the industry identifiers in equation (2). The firm residual effect is defined as a deviation from the industry effect. Therefore, the equation (2) becomes

$$(7) \quad y = X\beta + D\theta + FA\kappa + \ (F\psi - FA\kappa) \ + \varepsilon, \qquad \kappa \equiv (A'F'FA)^{-1} A'F'F\psi$$

Matrix A (JxK) allocates each of the J firms to each of the K industries and K(j) denotes the sectoral classification of firm j. The vector of parameters κ (Kx1) is interpreted as a weighted average of the pure firm effect. The effect (Fψ - FAκ) can be represented as $M_{FA}F\psi$. The terms FAκ and $M_{FA}F\psi$ (7) result from the decomposition of Fψ in two orthogonal components. In this context, two incomplete versions of (7) can produce inconsistent estimates.

Equation (8) indicates a situation in which the residual firm effect is omitted. In this case, the industry effect will be the sum of the pure effect of industry (κ) and an omitted variable bias term

$$(8) \quad \kappa^* = \kappa + (A'F'M_{[D\ X]}FA)^{-1} A'F'M_{[D\ X]} M_{FA} F\psi,$$

where $M_{[D\ X]}$ is a matrix M_Z with $Z \equiv [\ D \mid X\ ]$. The vector of observable β* will be the sum of the β from the complete model to an omitted variable bias term:

$$(9)\quad \beta^* = \beta + (X'M_{[D\ FA]}X)^{-1}X'M_{[D\ FA]}M_{FA}F\psi.$$

Equation (10) shows an incomplete version of equation (7), without the firm effect ($M_{FA}F\psi$) and the individual fixed effect (θ), implyingκ^{**}:

$$(10)\quad \kappa^{**} = \kappa + (A'F'M_XFA)^{-1}A'F'M_X(M_{FA}F\psi + D\theta)$$

$$\equiv (A'F'M_XFA)^{-1}A'F'M_XF\psi + (A'F'M_XFA)^{-1}A'F'M_XD\theta.$$

Therefore, incomplete versions of Equation (7) will provide inconsistent industrial effects. Abowd et al. (1999) uses the equation (10) to determine what proportion of the estimated inter-industry wage differentials is explained by worker effects and firm effects.

2.1. Statistical model

Based on the general model of Equation (2), Abowd et al. (1999) have developed alternative statistical approaches, including firm and worker fixed effects. For identification, a cross-product matrix containing sub-matrices of variables in the general model is pre-multiplied by the vector of parameters. This matrix is given by:

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

Its dimension depends on the number of workers (N) and firms (J). The usual computational methods for the estimation of the parameters β, θ, and ψ by least squares are not generally adequate. As the estimation of the full and unrestricted model computational is difficult, Abowd et al. (1999) proposed alternative methods to try to preserve the general structure of the complete model. Workers' mobility between firms is a necessary condition for the statistical identification of the model, i.e., to find the fixed effects separately, regardless of the computational approach adopted.

The so called consistent method computes first differences to data within firm-worker, holding the suppositions in equation (1), and using the definition:

$$(11)\quad \psi_j = \phi_j + \gamma_j s_{it},$$
$$F\psi_j = F_0\phi + F_1\gamma$$

Where s_{it} denotes the seniority of worker *i* in firm $j = J(i,t)$; ϕ_j is the firm-specific intercept; γ_j is the coefficient of seniority in the specific firm, F_0 and F_1 are N^*xJ matrices and γ is a $Jx1$ vector. Thus, applying first differences to all information for which $J(i,n_{it}) = J(i,n_{it-1})$, gives:

$$(12)\quad y_{in_{it}} - y_{in_{it-1}} = (x_{in_{it}} - x_{in_{it-1}})\beta + \gamma_{J(i,n_{it})}\left(s_{in_{it}} - s_{in_{it-1}}\right) + \varepsilon_{in_{it}} - \varepsilon_{in_{it-1}}$$

$$\Delta y = \Delta X\beta + \tilde{F}\gamma + \Delta\varepsilon,$$

Where n_{it} refers to worker i from the first to the last year he is present in the sample, Δy is $\tilde{N}^* \times 1$, ΔX is $\tilde{N}^* \times P$, $\tilde{F}$ is $\tilde{N}^* \times J$, $\Delta\varepsilon$ is $\tilde{N}^* \times 1$ and $\tilde{N}^*$ is the number of combinations (i, t) satisfying the condition J(i, n_{it}) = J(i, n_{it-1}) in the sample. The estimates of this method are:

$$(13) \quad \tilde{\beta} = (\Delta X' \mathrm{M}_{\tilde{\mathrm{F}}} \Delta X)^{-1} \Delta X' \mathrm{M}_{\tilde{\mathrm{F}}} \Delta y$$

$$(14) \quad \tilde{\gamma} = (\tilde{F}'\tilde{F})^{-1} \tilde{F}'(\Delta y - \Delta X \tilde{\beta}).$$

However, the consistent method is inefficient to estimate the complete model (2) because the first difference will eliminate workers whose firm in t differs from the firm in t-1. In addition, given the constraint of J(i,n_{it}) = J(i,n_{it-1}), the consistent method cannot be used to identify the fixed effects of firm and worker separately.

Conditional methods do not restrict the sample and can identify those fixed effects separately. The denomination conditional results from its relationship with the linear models standard techniques and by its origin in the panel data literature, associated to models with workers' fixed effects. Additional suppositions of orthogonality are needed: the interactions between X, D and F will be proxies for the correlations between these variables and the estimation assumes conditional orthogonality, given the interactions.

In this context, a Z (N*xQ) matrix is defined from Q information functions on X, D and F. For its construction, Abowd et al. (1999) included the firm size and it square, industry, worker's experience and his age at the end of the schooling period. The least squares solution for (2) can be found by considering some orthogonality assumptions conditional on Z. Under the supposition that X and D are orthogonal to the projection of F on the null space of Z, equation (2) becomes:

$$(15) \quad y = X\beta + D\theta + Z\gamma + M_Z F\psi + \varepsilon$$

Where $\gamma \equiv (Z'Z)^{-1}Z'F$. The supposed conditional orthogonality between X and F and between D and F, given Z, implies that

$$(16) \quad X' M_Z F = 0$$

$$(17) \quad D' M_Z F = 0$$

Given these suppositions, the least squares solution for the parameters in (15) is

$$(18) \quad \begin{bmatrix} \hat{\beta} \\ \hat{\theta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X'X & X'D & X'Z \\ D'X & D'D & D'Z \\ Z'X & Z'D & Z'Z \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ D'y \\ Z'y \end{bmatrix}$$

$$(19) \quad \hat{\psi} = (F' M_Z F)^{-} F' M_Z y$$

Where [ ]$^{-1}$ is a generalized inverse, required because $rank(F' M_Z F) = mJ - 1 - Q$. The order-independent estimation is a conditional method to compute solutions (18) and (19), whose stages are independent. It is implemented in two stages. The first follows a within-D (within worker) longitudinal estimation, in which X and Z are projected over D, obtaining

$$(20) \begin{bmatrix} \hat{\beta} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} X' M_D X & X' M_D Z \\ Z' M_D X & Z' M_D Z \end{bmatrix}^{-1} \begin{bmatrix} X' M_D y \\ Z' M_D y \end{bmatrix},$$

$$(21)\quad \hat{\theta} = (D'D)^{-1}D'(y - X\hat{\beta} - Z\hat{\lambda}).$$

The second stage, within-F (within firm), computes the least square solution for the parameters of F and Z from

$$(22)\quad y = F\psi + Z\pi + \upsilon,$$

Where π $(Qx1)$ is a vector of auxiliary parameters and $\upsilon \sim \mathrm{N}(0, \sigma_{\upsilon}^2 I)$, given the conditional orthogonality suppositions. The second-stage solution is:

$$(23)\quad \hat{\pi} = (Z'M_F Z)^{-1}Z'M_F y,$$

$$(24)\quad \hat{\psi} = (F'F)^{-1}F'(y - Z\hat{\pi}).$$

In the order-dependent estimation, the stages are not independent, so that the estimates of the parameters may be different, depending on which effects are estimated first. If the effects of workers are estimated in the first stage, the parameters β, θ and λ are recovered, as in the order-independent method, according to (20) and (21). In the second stage, the effects of firm are estimated using the equations in (11). To this purpose, all the observations about workers employed in the same firm are grouped in the set {j} ≡ {(i, t) | J (i, t) = j}, with Nj elements, such that:

$$(25)\quad \hat{y}_{\{j\}} \equiv y_{\{j\}} - x_{\{j\}}\hat{\beta} - \hat{\theta}_{\{j\}}$$

$$(26)\quad y_{\{j\}} \equiv \begin{bmatrix} \cdots \\ y_{ns} \\ \cdots \end{bmatrix}, \forall (n,s) \in \{j\}$$

With $x_{\{j\}}$ and $\theta_{\{j\}}$ similar to $y_{\{j\}}$, and using the first stage estimates for $x\hat{\beta}$ and $\hat{\theta}$. Thus, the firm-level equation is:

$$(27)\quad \hat{y}_{\{j\}} = F_{\{j\}}\begin{bmatrix} \phi_j \\ \gamma_j \end{bmatrix} + \zeta_{\{j\}}$$

$$F_{\{j\}} \equiv \begin{bmatrix} & \cdots \\ 1 & s_{ns} \\ & \cdots \end{bmatrix}, \forall (n,s) \in \{j\}$$

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

The least square estimator of (27) is

$$(28)\quad \begin{bmatrix} \hat{\phi}_j \\ \hat{\gamma}_j \end{bmatrix} = (F'_{\{j\}}F_{\{j\}})^{-1}F'_{\{j\}}\hat{y}_{\{j\}}, \text{ for } j = 1,\ldots,J.$$

In the case the firm effects are estimated first, the first stage uses the estimator of $\hat{\psi}$ given by (19) or (24). The second stage finds β and θ from:

$$(29)\quad y - F\hat{\psi} = X\beta + D\theta + \xi,$$

$$\xi = \varepsilon + F(\psi - \hat{\psi}).$$

Therefore, the estimators of β and θ are:

$$(30)\quad \hat{\beta} = (X'\mathrm{M_D}X)^{-1}X'\mathrm{M_D}(y - F\hat{\psi}\,),$$

$$(31)\quad \hat{\theta} = (D'D)^{-1}D'(y - X\hat{\beta} - F\hat{\psi}).$$

The estimators of β and θ obtained by the order-independent method are identical to those obtained by the order-dependent conditional method, in which the effects of workers are estimated first. The order-independent estimator of ψ is identical to its order-dependent estimator with the effect of firm estimated first.

3. Empirics

The database of this study consists of a 5% random sample of workers taken from a wide longitudinal matched database of workers and firms in Brazil, extracted from administrative records by the Ministry of Labor and Employment (RAIS-Migra, MTE). Our sample covers the period 1995-2008. The sample is composed of 2,328,018 observations, corresponding to a balanced panel with 166,287 workers, employed every year, with positive income and age between 18 and 65 years; there are 126,704 firms and 324,419 combinations of worker and firm. The sample does not include workers and firms in the public sector (government), since wage formation in this case does not follow the same market influences as in the private sector. Wages were deflated by a national price index (IPCA). Following Freguglia (2007), whenever the metropolitan region did not have a specific price index, the index of a neighboring similar region was used[2]. Workers from 294 labor market areas (LMA) defined by the Brazilian statistics agency (IBGE) were included ("arranjos populacionais", encompassing 56% of population[3]). Over the period of this study, there has been a change in the number and border of municipalities in Brazil, due to extinction, emancipation and foundation. Therefore, this study defined 4.253 Minimum Comparable Areas (MCA) for the period 1995-2008, keeping constant the areas of each unity of investigation related to the municipalities.

The identification of the fixed effects of firm and location requires mobility of workers between firms and metropolitan regions. Table 1 presents the within variation (within worker); the between variation is the variation between workers. As can be seen, the mobility condition is met for the identification of the estimators.

Table 1. Between and within variations in the sample

	Variation	(%)
Region	Between	0.9325
	Within	0.0675
Firm	Between	0.8516
	Within	0.1463

Source: Author's calculation from RAIS-Migra (MTE).

The identification strategy of this study consists in the adoption of panel data modeling, to control for the worker and firm unobserved heterogeneity and the respective sorting on wages, and location effects. The control for the worker's fixed effects allows for eliminating the bias in the estimation of the parameters in the wage equation. It also allows for correcting the self-selection problem resulting from the attraction of skilled

[2] Anyway, although metropolitan inflation rates may differ in the short run, they tend to be quite similar over time.
[3] Available in: <http://www.ibge.gov.br/home/geociencias/geografia/geografia_urbana/arranjos_populacionais/default.shtm?c=9>. The size distribution of the 294 LMA is: 189 small ($Pop \leq 100{,}000$); 81 medium ($100{,}000 < Pop \leq 750{,}000$); 24 large ($Pop > 750{,}000$). The large LMAs (24 LMAs and 2 large municipalities) are labeled Metro Regions in the remainder of the paper.

workers to the major urban centers. The inclusion of the effects of location in labor market areas and of firm allows controlling for the sorting or self-selection by location, or by the attraction of more productive firms to more developed areas.

Table 2 presents the basic descriptive statistics of the variables in the sample. The majority of workers are male (70%) with average age 39 years, without a high school degree (49.7%); this pattern is replicated in the metropolitan areas: (69%), 39 years old and (47%), respectively. The average wage in the metropolitan areas (R$ 2,767) was larger than the average wage in overall sample (R$ 2,563), a sign of a possible wage premium. About 2% of the sample has moved between LMA (migration LMA) and 7% between firms (migration firm) in the period.

Table 2 – Descriptive statistics of the variables in the sample (1995-2008)

Variable	**Total**				**Medium and large LMA**			
	Average	**St Dev**	**Min**	**Max**	**Average**	**St Dev**	**Min**	**Max**
Education < 11	0.4968		0	1	0.4704		0	1
Education = 11	0.2719		0	1	0.2768		0	1
11<Education <15	0.0484		0	1	0.0526		0	1
Education ≥ 15	0.1829		0	1	0.2001		0	1
Experience	123.56	86.77	0	678.80	124.04	87.5	0	678.80
Age	38.63	8.59	18	65	38.73	8.58	18	65
Real wage	2,563.28	2,921.43	66.37	153,731.30	2,767.62	3,056.94	66.37	153,731.30
ln(wreal)	7.42	0.91	4.20	11.94	7.51	0.90	4.20	11.94
Dsex	0.30		0	1	0.31		0	1
Small firm	0.40		0	1	0.37		0	1
Medium firm	0.27		0	1	0.27		0	1
Large firm	0.33		0	1	0.35		0	1
Population of MCA	2,315,843	356,5942	782	11,000,000	2,787,518	3,749,819	1,791	11,000,000
Density of MCA	2,518.71	2,861.04	0.11	13,304.23	3,026.59	2,895.36	1.35	13,304.23
Large LMA	0.5929		0	1	0.7159		0	1
Medium LMA	0.2352		0	1	0.2841		0	1
Small LMA	0.0347		0	1	0		0	0
Education change	0.0384		0	1	0.0384		0	1
Migration btw LMA	0.0242		0	1	0.0232		0	1
Migration firm-LMA	0.0033		0	1	0.0029		0	1
Migration firm	0.0738		0	1	0.0768		0	1
Total employment MCA	751,390	1,208,928	1	4,489,076	905,861	1,275,181	80	4,489,076
Employment density MCA	798.17	1,010.12	0	7,141.43	960.90	1,038.28	0.07	7,141.43

Source: Author's calculation from RAIS-Migra (MTE).

We have estimated the following econometric model

$$(32)\quad \ln(w_{it}) = \beta_0 + \beta_1 educ_{it} + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + \beta_4 age_{it} + \beta_5 age_{it}^2 + \beta_6 Gender_i + \beta_7 Sector_{it} + \beta_8 Firm\ size_{it} + \beta_9 T_t + \alpha LMA\ Size_{it} + \gamma F_{ijt} + \theta_i + \varepsilon_{it}$$

In which $\ln(w_{it})$ is the natural logarithm of the actual wage of worker *i* in year *t* = 1995,..., 2008; $educ_{it}$ indicates schooling, in educational degrees; exp_{it} is the experience in the job (months employed in the same firm); age is the worker's age; dummy variables for gender, sector, firm size (small: <99 employees; medium: 100-499; ≥500) and LMA size were included. The location dummies vector, $LMA\ Size_{it}$, captures the fixed effects of location in each labor market area, by size of its population in time *t*: small LMAs; Medium LMAs; Large LMAs (metro region-MR). The non-LMA cities are taken as the reference for the LMAs dummies. If the location fixed effect is positive and significant, vector α provides the magnitude of the wage urban premium. The unobserved workers' skills are represented by θ_i. The vector F_{ijt} denotes the firm *j* fixed effect in time *t*. The time fixed effect is denoted by T_t and the error term is ε_{it}.

The fixed-point iteration algorithm proposed by Portugal and Guimarães (2010) is adopted for the estimation of the model, in its *Stata*[4] version developed by Correia (2014), to estimate models with multiple fixed effects. Take the following model used to obtain OLS estimates of β:

$$(33) \quad y = X\beta + D_1\alpha_1 + D_2\alpha_2 + D_3\alpha_3 + \epsilon,$$

where D_i are the indicator variables and α_i are the corresponding fixed effects. Following the Frisch-Waugh-Lovell theorem, the algorithm regresses y and X on each D_i, generating the residuals u_y and u_x; then u_y is regressed against u_x. Actually, the algorithm makes a linear regression demeaning the fixed effects. The fixed-point iteration strategy is presented in Appendix A.

This is the method used to estimate the complete model (32) and find the OLS conditional solution, with the inclusion of fixed effects. In addition to the conditional method, alternative versions of the general model are estimated, through the intra-worker-firm consistent method and using the traditional panel data methods, pooled ordinary least squares (POLS) and the fixed effects method.

4. Results

As a first exercise, we estimate equation (32) with a limited number of variables: observed worker and job characteristics, time dummies and a dummy indicating if the worker is in one of the LMAs (the non-LMA cities being the reference category). The POLS estimation, including controls for firm and worker, indicates an average LMA premium of 28.9%. However, when controls for the heterogeneity of firms and workers are included, in different estimation methods, this premium disappears, and even becomes negative (and significant). Table 3 presents the estimated location effects for the LMAs considered by IBGE as metropolitan regions, estimated by different methods, arranged in increasing order of population size. Since these are the LMAs considered as "large", there is a dummy for small and medium LMAs (non-LMA cities are the reference category).

Colum I shows the POLS results including the observed characteristics of workers and jobs, the location effects and time dummies, but without controlling for the fixed effects of firm and worker. Columns II and IV show the results using the traditional Fixed Effects method for panel data. Column VI shows the estimates obtained with the transformation intra worker-firm, as in the consistent method of Abowd et al. (1999), but using deviations from the mean instead of first differences. The units of analysis in this case are the spells firm-worker (i,j), from which the within-group transformation is performed, with the Fixed Effects estimator. Inside each spell, the fixed effects of firm and individual do not vary, and the within transformation eliminates them. Any variable constant in time inside each spell will not have its influence assessed. Column III shows the results of the conditional estimation, in which only the fixed effects of individuals (i) are absorbed. Column V shows the results when only the firm fixed effects (j) are absorbed. Both models include the fixed effects of location and the observed characteristics of workers. Finally, column VII shows the conditional estimation results, in which both the firm and worker fixed effects are absorbed, while the location effects are includes as variables in the regression. These results can be called conditional, with the fixed effects of worker estimated first. The iteration method used in these regressions works as the conditional approach proposed by Abowd et al (1999). The intra-firm (j) and intra individual-firm (ij) fixed effects exclude the sectoral effects (fixed). In the estimation of the firm or worker effects (conditional and FE), age was replaced by age range dummies.

[4] The Stata version algorithm is *reghdfe* (*Linear and instrumental-variable regression absorbing any number of high-dimensional fixed effects*) developed by Correia (2014).

Table 3 – Location effects, 1st stage (dependent variable: ln w)

LMAs	**POLS**	**FEi**	**CONDi**	**FEj**	**CONDj**	**FEij**	**CONDij**
	I	**II**	**III**	**IV**	**V**	**VI**	**VII**
Sorocaba/SP	0.4208***	0.0113	0.0113	0.0270	0.0221	0.0256**	0.0166
	(0.0051)	(0.0079)	(0.0079)	(0.0182)	(0.0182)	(0.0117)	(0.0118)
Campo Grande/MS	0.1686***	0.0004	0.0004	0.0433	0.0434	0.0374	0.0420
	(0.0070)	(0.0126)	(0.0126)	(0.0413)	(0.0413)	(0.0259)	(0.0263)
Cuiabá/MT	0.1985***	0.0313***	0.0313***	0.1386***	0.1375***	0.0443***	0.0534***
	(0.0077)	(0.0115)	(0.0115)	(0.0265)	(0.0265)	(0.0167)	(0.0169)
Florianópolis/SC	0.2801***	-0.0097	-0.0097	0.0621**	0.0575*	-0.0019	0.0002
	(0.0047)	(0.0090)	(0.0090)	(0.0300)	(0.0301)	(0.0189)	(0.0193)
Aracaju/SE	-0.076***	0.0080	0.0080	0.0249	0.0241	0.0048	0.0056
	(0.0062)	(0.0141)	(0.0141)	(0.0473)	(0.0473)	(0.0295)	(0.0298)
S. José dos Campos/SP	0.5731***	0.0347***	0.0347***	-0.0363	-0.0424	-0.079***	-0.075***
	(0.0042)	(0.0081)	(0.0081)	(0.0261)	(0.0261)	(0.0167)	(0.0166)
Teresina/PI	-0.176***	-0.076***	-0.076***	-0.101***	-0.101***	-0.108***	-0.106***
	(0.0058)	(0.0140)	(0.0140)	(0.0348)	(0.0348)	(0.0209)	(0.0214)
João Pessoa/PB	-0.1913***	-0.0555***	-0.0555***	0.0012	-0.0001	-0.0285	-0.0286
	(0.0059)	(0.0126)	(0.0126)	(0.0406)	(0.0406)	(0.0249)	(0.0254)
Maceió/AL	-0.0440***	0.0350**	0.0350**	0.1174**	0.1202***	0.1107***	0.1152***
	(0.0058)	(0.0143)	(0.0143)	(0.0465)	(0.0465)	(0.0282)	(0.0289)
Natal/RN	-0.1631***	0.0097	0.0097	0.0770**	0.0777**	0.0192	0.0277
	(0.0056)	(0.0115)	(0.0115)	(0.0309)	(0.0309)	(0.0189)	(0.0193)
São Luís/MA	-0.0411***	0.0066	0.0066	-0.1194**	-0.1174**	-0.1493***	-0.1484***
	(0.0060)	(0.0144)	(0.0144)	(0.0584)	(0.0584)	(0.0386)	(0.0382)
Baixada Santista/SP	0.4309***	0.0062	0.0062	0.0290	0.0288	0.0046	0.0100
	(0.0038)	(0.0085)	(0.0085)	(0.0217)	(0.0217)	(0.0135)	(0.0137)
Vitória/ES	0.1621***	0.0369***	0.0369***	0.1008***	0.1041***	0.0459**	0.0581***
	(0.0042)	(0.0088)	(0.0088)	(0.0308)	(0.0309)	(0.0192)	(0.0195)
Manaus/AM	0.4130***	0.0943***	0.0943***	-0.0688*	-0.0705**	-0.0491*	-0.0468*
	(0.0053)	(0.0128)	(0.0128)	(0.0359)	(0.0359)	(0.0251)	(0.0249)
Campinas/SP	0.5497***	0.0447***	0.0447***	-0.0083	-0.0111	-0.0174*	-0.0152
	(0.0029)	(0.0048)	(0.0048)	(0.0141)	(0.0141)	(0.0093)	(0.0093)
Belém/PA	0.1109***	-0.0248**	-0.0248**	-0.0674*	-0.0689*	-0.0027	-0.0127
	(0.0050)	(0.0115)	(0.0115)	(0.0379)	(0.0379)	(0.0241)	(0.0245)
Goiânia/GO	0.1831***	-0.0141*	-0.0141*	-0.0445**	-0.064***	-0.0029	-0.0266*
	(0.0045)	(0.0082)	(0.0082)	(0.0217)	(0.0218)	(0.0135)	(0.0137)
Porto Alegre/RS	0.4363***	0.0310***	0.0310***	0.0109	0.0064	0.0073	0.0019
	(0.0025)	(0.0048)	(0.0048)	(0.0121)	(0.0121)	(0.0079)	(0.0080)
Curitiba/PR	0.3004***	0.0156***	0.0156***	-0.0100	-0.0094	-0.036***	-0.035***
	(0.0027)	(0.0050)	(0.0050)	(0.0151)	(0.0151)	(0.0099)	(0.0099)
Fortaleza/CE	-0.136***	-0.049***	-0.049***	-0.052***	-0.053***	-0.072***	-0.074***
	(0.0034)	(0.0069)	(0.0069)	(0.0149)	(0.0149)	(0.0092)	(0.0094)
Brasília/DF	0.4161***	0.1088***	0.1088***	0.0788***	0.0818***	0.0454***	0.0420***
	(0.0031)	(0.0060)	(0.0060)	(0.0159)	(0.0161)	(0.0101)	(0.0103)
Salvador/BA	0.1358***	0.0006	0.0006	-0.0051	-0.0065	-0.066***	-0.063***
	(0.0030)	(0.0063)	(0.0063)	(0.0165)	(0.0165)	(0.0109)	(0.0110)
Recife/PE	-0.032***	0.0130*	0.0130*	0.0232	0.0198	0.0039	0.0052
	(0.0030)	(0.0067)	(0.0067)	(0.0179)	(0.0179)	(0.0111)	(0.0113)
Belo Horizonte/MG	0.2838***	0.0215***	0.0215***	0.0502***	0.0500***	0.0163*	0.0192**
	(0.0022)	(0.0043)	(0.0043)	(0.0129)	(0.0129)	(0.0087)	(0.0087)
Rio de Janeiro/RJ	0.2712***	0.0547***	0.0547***	0.0566***	0.0576***	0.0032	0.0047
	(0.0018)	(0.0042)	(0.0042)	(0.0113)	(0.0113)	(0.0075)	(0.0074)
São Paulo/SP	0.5576***	0.0434***	0.0434***	0.0462***	0.0460***	-0.0121**	-0.0123**
	(0.0015)	(0.0029)	(0.0029)	(0.0075)	(0.0075)	(0.0050)	(0.0050)
Medium LMA	0.2105***	0.0214***	0.0214***	0.0161**	0.0154**	0.0043	0.0053
	(0.0014)	(0.0023)	(0.0023)	(0.0063)	(0.0063)	(0.0040)	(0.0041)

LMAs	POLS	FEi	CONDi	FEj	CONDj	FEij	CONDij
	I	II	III	IV	V	VI	VII
Small LMA	0.0747***	-0.0066*	-0.0066*	0.0648***	0.0635***	-0.0030	-0.0029
	(0.0024)	(0.0039)	(0.0039)	(0.0109)	(0.0109)	(0.0069)	(0.0070)
R^2	0.560	0.234	0.908	0.383	0.801	0.089	0.929
Adjusted R^2	0.560	0.142	0.901	0.242	0.790	0.048	0.921
R^2 Within	-	0.203	0.203	0.283	0.283	0.181	0.177

Source: Author's calculation from RAIS-Migra (MTE).
Notes: Complete table in Appendix B; Significant at *5%, **10%, ***1%.

As for the observed characteristics of workers, the results came out as suggested in the labor market literature (Appendix B). Wage growths with age and experience at decreasing rates. Women receive 34% less than men do. The returns to education are reduced after controlling for the fixed effects of firm, worker, or both. Workers with a university degree received wages 121% (POLS) higher then workers with less than eleven years of schooling (10% CONDij); completing the basic education (11 years) increased the wage by 49% (POLS) and 0,7% (CONDij). Large and medium size firms paid 5.8% and 3.1% more than small firms. The sectoral dummies indicate that only five sectors paid more than manufacturing.

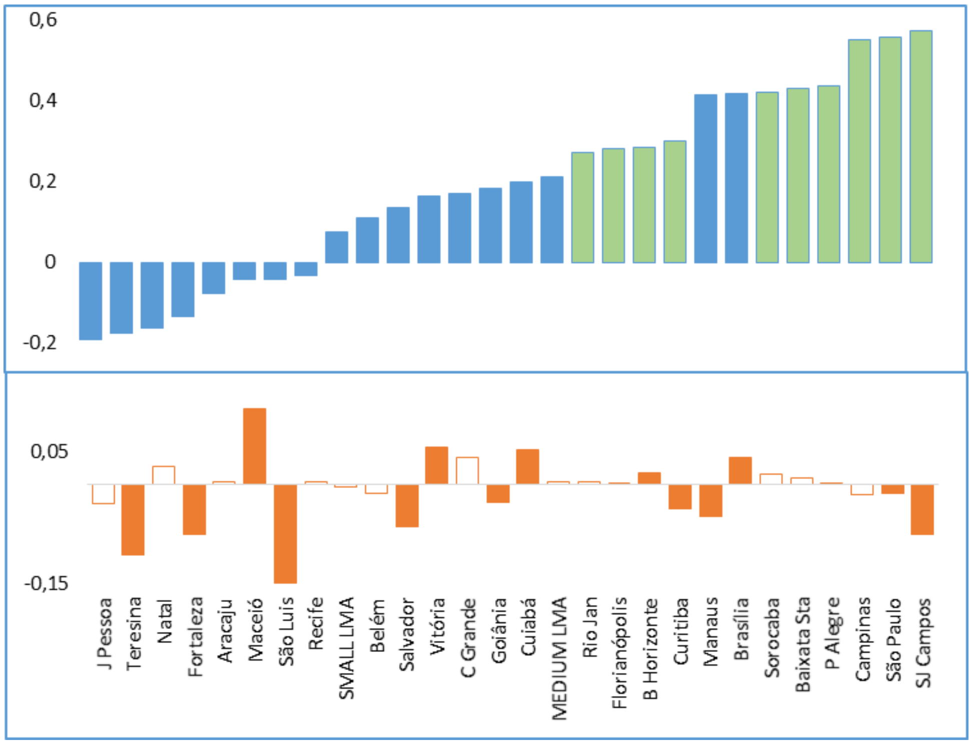
The individual fixed effects are more important to explain the wage variation than the firm effects. A comparison columns III and V indicates that the unobserved individual heterogeneity is more important for explaining wage variation (R^2=91%) than the unobserved firm heterogeneity is (R^2=80%). Including both effects raise the R^2 to 93% (Column VII). Comparing these R^2 to column I, where the effects of firm and worker are neglected (R^2=56%), it is clear that controlling for them is very important. The conditional method including fixed effects for individual and firm (Colum VII) shows similar results to Column VI (fixed effects intra worker-firm), but only the former is capable of including firm and worker effects individually. This indicates that both methods would be able to identify similar estimates of the parameters. The individual fixed effects and the conditional method intra-individual (column II and III) also presented identical estimates.

Figure 1 presents the results for the estimated urban premium for the 26 LMAs considered by IBGE as metropolitan regions. They are displayed in increasing order of the premium estimated by POLS. The POLS urban premium for the LMAs belonging to the economic core of the country, the southeast and south regions, is displayed in a different color. It is clear that these LMAs are well above the others in terms of urban premium. The exceptions are Brasília and Manaus. The first is the Federal District, including the nation's capital city, with well-known high wage levels due to high-level public workers[5]. Manaus, in the middle of the Amazon region, hosts a free import zone of electronics and transportation equipment, and tends to be an outlier in most economic studies of any sort. Besides these 26 metropolitan LMAs, the chart also shows the small and medium LMAs as groups, indicating that the POLS wage premium is higher for the second group (21%), although the small LMAs also show a wage premium in relation to the set of non-LMA cities (7.5%). The small columns in brown refer to the wage premium obtained by the conditional estimation, in which the fixed effects of firm and worker are absorbed, while the effects of location are include as variables in the regression. Clearly, the size of the effects are a fraction of the ones estimated by POLS, indicating that controlling for unobserved heterogeneity is an important step in the estimation. Interestingly, the resulting estimated wage premium are either negative or non-significant for the LMAs of the economic core of the country. In the case of São Paulo LMA and the neighboring SJ Campos LMA, the estimated premium is negative. This suggests that, even if these LMA exhibit high wage levels, these levels should be even higher, given the quality of firms and workers present in the areas. This is also the case of Manaus and Curitiba.

[5] Even if we did not include public workers in the sample, the effect of the high wages paid by government on the areas' private labor market is evident.

Therefore, after considering the unobserved heterogeneity of firms and workers, the urban wage premium disappeared in 14 out of those 26 metropolitan regions. Besides these 14 metropolitan areas, the urban wage premium disappeared in the small and medium LMAs. The results indicate that much of the positive impact of location on wages results from the unobserved heterogeneity of firms and workers.

Figure 1 – Wage premium for LMAs (reference: non-LMA cities)



Column colors: blue – POLS estimated urban premium (green – economic core); brown – urban premium conditionally estimated; shaded columns – significant coefficients.

Table 4 presents the decomposition of the location fixed effects of the 26 metro regions (large LMAs) estimated in the first stage into the observed characteristics of the workers and the unobserved heterogeneity of workers and firms. The location effects were taken from column I in Table 3, and they only include the observed characteristics of individuals and jobs, and the time effects. In order to decompose these effects into the effects of firm and worker, we regressed the three fixed effects on the LMA-averages of workers, sectors and firms characteristics, and employment and area density. Columns (1) to (3) show the results of the FE regressions, with correction for (robust) standard errors. All second stage estimations used the panel POLS method. Columns (4)-(6) show the decomposition, adopting large LMA clusters for the correction of the standard errors. The results indicate that the unobserved heterogeneity of firms and workers accounts for almost all variation of the location effects (R^2=95%, column 6). This suggests that almost all wage premium of denser metro areas come from these unobserved characteristics. The residual 5% could represent the local attributes related to the LMAs. The effect of employment density on wage variation in large LMAs was 0.9%, while the effect of area (scale) was 3.1%.

Table 4 – Decomposition of the location effects (2nd stage)

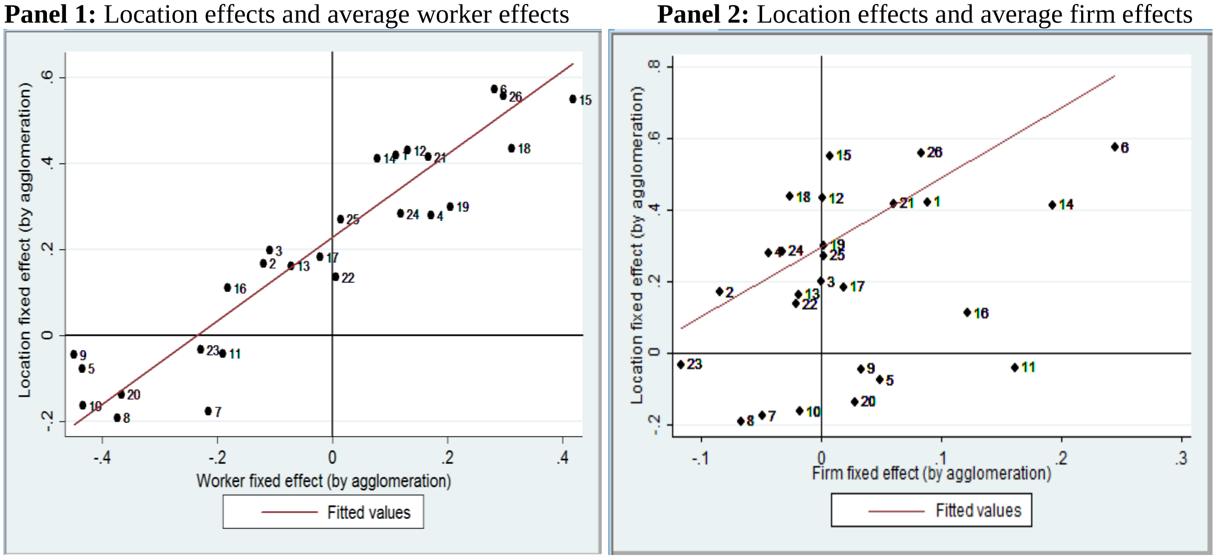
	(1) FE_LMA	(2) Mean(FEi)	(3) Mean(FEj)	(4) FE_LMA|X	(5) FE_LMA|X	(6) FE_LMA|X
ln(emply_density)	0.0094***	0.0421***	-0.0237***			
	(0.0002)	(0.0002)	(0.0001)			
ln(LMA_area)	0.0310***	0.1179***	-0.0830***			
	(0.0004)	(0.0004)	(0.0002)			
FEi|X				0.9769***		0.8709***
				(0.0662)		(0.0462)
FEj|X					2.2000***	0.7057***
					(0.5173)	(0.1448)
Time effects	Sim	Sim	Sim			
Constant	7.3936***	4.8011***	3.8766***	0.2276***	0.2894***	0.2197***
	(0.0224)	(0.0228)	(0.0120)	(0.0123)	(0.0316)	(0.0120)
# Observ.	1380168	1380168	1380168	1380168	1380168	1380168
R^2	0.941	0.942	0.840	0.919	0.408	0.950
R^2 Adjusted	0.941	0.942	0.840	0.919	0.408	0.950

Source: Author's calculation from RAIS-Migra (MTE).
Note: Significant at *5%, **10%, ***1%.

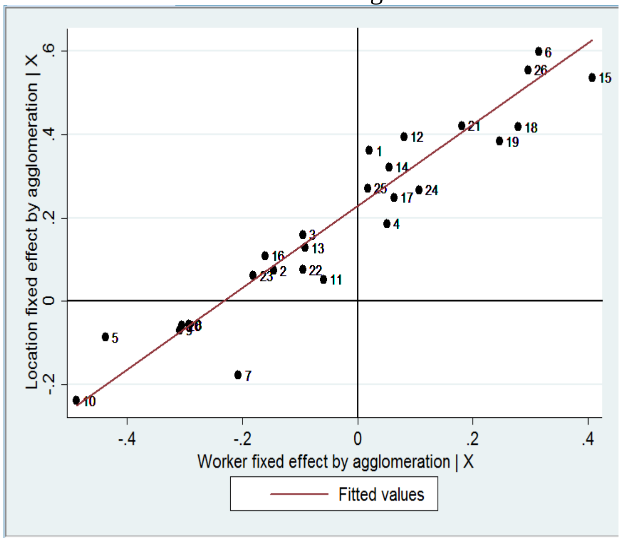
The worker heterogeneity appears to be more important in the explanation of wage differentials than the firm heterogeneity. The latter explained only 41% of wage variation (column 5), while the former explained 92% (column 4). However, it should be emphasized that both effects account individually for an important share of wage variation. Therefore, the relative importance of the firm fixed effects should not be disregarded. Figure 2 shows the relative strength of each component.

Figure 2 – Relative importance of firm and worker effects

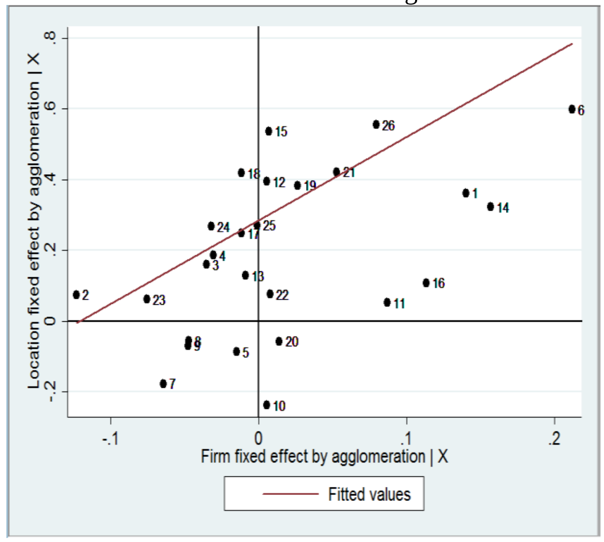
Panel 1: Location effects and average worker effects **Panel 2:** Location effects and average firm effects



Panel 3: Location effects and worker average effects, conditioned on the variables averages



Panel 4: Location effects and firm average effects, conditioned on the variables averages



Panels 1 and 2 present the relationship between the fixed effects of worker and firm and the location effects (each dot refers to one of the 26 large LMAs), without controlling for the observed characteristics of workers. Panels 3 and 4 present the same relations, after controlling for the observed characteristics of the formal workers. All figures include a fitted linear regression line. The effects of worker show a slightly stronger correlation to the effects of location than the effects of firms.

Finally, since the LMA-averages of workers, sectors and firms refer to variables already used in the first stage, we regressed the location effects without include these variables and using regional macro dummies. In this case, the contribution of the firm effects increased to 80%, but the contribution of the worker effects was still more relevant (95%). The residual effect, attributable to agglomeration, previously estimated as 5%, dropped to 3%. The effect of density varied between 1.9% and 3.6%, and the area effect (scale), varied between -0.3% and 5%.

5. Conclusions

This paper aimed at evaluating the contribution of the unobserved heterogeneity of firm and worker to the localization effects on wages in the Brazilian case. Theoretically, the positive effects of urban agglomeration on wages could be confused with the concentration of more productive workers and firms in major urban centers, featuring a sorting process between locations. Wage equations were estimated in order to identify the contribution of firm and worker effects on wages, including the observed characteristics of the workers and jobs and fixed effects for location, worker and firm. The empirical strategy resorted to methods presented in the labor economics literature, to deal with multiple fixed effects in large data bases with matched worker and firm observations.

The results showed that the observed characteristics of workers and jobs and the effects of location in labor market areas explained 56% of the wage variance. The inclusion of the unobserved heterogeneity of workers and firms raised the explanation to about 93%. For the same control variables, the inclusion of the worker effects (absorbed) alone increased accounted for the explanation to about 91% of the wage variance, while the inclusion of the unobserved firm effects alone produced 80% of explanation.

The POLS results indicate a wage premium in 18 out of 26 MRs, as compared to non-metro areas. After controlling for the effects of firm and worker, only 5 large LMAs maintained their wage premium. The inclusion of the unobserved heterogeneity of firms and workers eliminated the wage premium in 14 MRs, and its magnitude was reduced in the other regions. The location fixed effects coefficients were almost entirely explained by the fixed effects of worker and firm (95%), controlling for the average observed characteristics. The results suggest that a large portion of the wage gains attributed to the location in dense urban areas come from the unobserved characteristics of firms and workers, leaving only the remaining 5% to be associated with the local attributes of urban agglomerations. The heterogeneity of worker is more important than the heterogeneity of firm, since the former explained about 92% of the variation of the localization effects on wages, while the later explained about 41%.

Therefore, the evidence presented in this article showed that the unobserved heterogeneity of firms and workers are key components in the determination of wages and of the localization effects on wages. The worker fixed effects were more important than the firm effects to explain the change of wages and the localization effects on real wages in metropolitan areas. However, both the effects accounted for a substantial variation in real wages and localization effects on wages of Brazilian workers.

References

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, U.S. Census Bureau. 2002. Available in: < ftp://ftp2.census.gov/ces/tp/tp-2002-06.pdf >. Access in Jan 6 2015.

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

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

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:*** XXXVII Encontro Nacional de Economia - ANPEC, Foz do Iguaçu, PR. 2009. Available in: <http://www. anpec.org.br/encontro2009/inscricao.on/ arquivos/000-c6f925639a711d35a1187 c1b9c5483f3. pdf>. Access in 06/16/2011.

COMBES, P-P.; DURANTON, G.; GOBILLON, L. Spatial wage disparities: Sorting matters. **Journal of Urban Economics**, n. 63, pp. 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, pp. 913–930, 2012.

CORREIA, Sergio. Explanation of the HDFE iteration with 3 Fes. 2014. Available in: <http://fmwww.bc.edu/repec/bocode/r/reghdfe_explanation.pdf>. Access in Jan 3 2015.

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

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

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

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

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

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

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

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

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

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. **Índice Nacional de Preços ao Consumidor Amplo (IPCA).** Available in: <http://www.sidra. ibge.gov.br/snipc/ tabelaIPCA.asp?o=3&i=P>. Access in May 5 2012.

LEONARD, J. S.; VAN AUDENRODE, M. Persistence of Firm and Individual Wage Components. **University of California at Berkeley, Working Paper**. 1996. Available in: <http://www.ecn.ulaval.ca/w3/recherche/cahiers/1996/9607.pdf>. Access in Jan 14 2015.

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**, Volume 12 Number 3: pp. 406-432. 2012.

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

MINISTÉRIO DO TRABALHO E EMPREGO (MTE). **RAIS-Migra**. Available in: <http://www.mte.gov.br/pdet/o_pdet/ produtos/BD_estatisticas.asp# raismigra>. Access in Aug 18 2011.

MITTAG, N. New methods to estimate models with large sets of fixed effects with an application to matched employer-employee data from Germany. **FDZ-Methodenreport**, n. 1. 2012. Available in: <http://doku.iab.de/fdz/reporte/2012/MR_01-12_EN.pdf>. Access in Jan 3 2015.

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

ROCHA, R. M.; SILVEIRA NETO, R. M.; GOMES, S. M. F. P. O. Maiores cidades, maiores habilidades produtivas: ganhos de aglomeração ou atração de habilidosos? Uma análise para as cidades brasileiras. ***In***: XVI Encontro Regional de Economia, Fortaleza, julho, 2011. Available in: http://www.bnb.gov.br/content/aplicacao/ eventos/ forumbnb2011/docs/2011_maiores_cidades.pdf>. Access in Aug 2011.

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

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data**. Cambridge: Massachusetts Institute of Technology - MIT, 2002. 740p.

Appendix

Appendix A - The fixed-point iteration for models with high-dimensional fixed effects

The fixed-point iteration strategy developed by Correia (2014) is applied alternating between the estimation of β and α in the solution equations:

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

where, $(D_i'D_i)^{-1}D_i'$ is obtained as a group-average of residuals from the regression of y on X. The β estimation is given by the regression of the transformed y on X. However, the implementation keeps y as the dependent variable and includes $D_i\alpha$ as a covariate. When the procedure converges, the vector $D_i\alpha$ contains the estimates of the coefficients of the indicator variables. Defining $Z_2 := D_2\hat{\alpha}_2$ and $Z_3 := D_3\hat{\alpha}_3$, and assuming values for the iteration, the procedure developed by Correia (2014) is:

(i) Compute $P_1 y$ and $\tilde{y} := M_1 y$;
(ii) Start with $Z_2^{(0)} = Z_3^{(0)} = 0$;
(iii) Until the convergence of Z_2 and Z_3:

(a) $Z_2^{(n)} = P_2\left[\tilde{y} + P_1\left(Z_2^{(n-1)} + Z_3^{(n-1)}\right) - Z_3^{(n-1)}\right]$,
(b) $Z_3^{(n)} = P_3\left[\tilde{y} + P_1\left(Z_2^{(n)} + Z_3^{(n-1)}\right) - Z_2^{(n)}\right]$;

(iv) Compute $Z_1 = P_1(y - Z_2 - Z_3)$. From it, compute $y^* = y - Z_1 - Z_2 - Z_3$;
(v) After repeating steps (i) through (iv) for each variable, regress the transformed variables to obtain $\hat{\beta}$: $y^* = X^*\hat{\beta} + e$;
(vi) To obtain the fixed effects, use $y = X\hat{\beta} + Z_1 + Z_2 + Z_3 + e$; with $\hat{\beta}$ and e from step (v), compute $\tilde{e} := y - X\hat{\beta} = Z_1 + Z_2 + Z_3 + e$ and apply the previous step to obtain Z_1, Z_2 and Z_3.

Appendix B – Wage equations 1^{st} stage (dependent variable: ln wage)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	POLS	FEi	FEj	FEij	CONDij	POLS	FEi	CONDi	FEj	CONDj	FEij	CONDij
educ = 11	0.4811***	0.0161***	0.2590***	0.0123***	0.0078***	0.4888***	0.0162***	0.0162***	0.2590***	0.2590***	0.0123***	0.0077***
	(0.0010)	(0.0011)	(0.0009)	(0.0012)	(0.0011)	(0.0010)	(0.0011)	(0.0011)	(0.0009)	(0.0009)	(0.0012)	(0.0011)
11<educ < 15	0.8905***	0.0509***	0.5152***	0.0383***	0.0342***	0.8511***	0.0512***	0.0512***	0.5152***	0.5152***	0.0382***	0.0342***
	(0.0020)	(0.0019)	(0.0017)	(0.0021)	(0.0020)	(0.0019)	(0.0019)	(0.0019)	(0.0017)	(0.0017)	(0.0021)	(0.0020)
educ ≥ 15	1.2522***	0.1423***	0.8655***	0.0890***	0.0998***	1.2127***	0.1421***	0.1421***	0.8655***	0.8655***	0.0890***	0.0997***
	(0.0013)	(0.0018)	(0.0012)	(0.0020)	(0.0019)	(0.0012)	(0.0018)	(0.0018)	(0.0012)	(0.0012)	(0.0020)	(0.0019)
age	0.0528***					0.0569***						
	(0.0004)					(0.0004)						
age2	-0.001***					-0.001***						
	(0.0000)					(0.0000)						
25-29 age		0.1466***	0.1411***	0.0979***	0.1079***		0.1466***	0.1466***	0.1411***	0.1411***	0.0979***	0.1079***
		(0.0012)	(0.0018)	(0.0012)	(0.0012)		(0.0012)	(0.0012)	(0.0018)	(0.0018)	(0.0012)	(0.0012)
30-39 age		0.2062***	0.2534***	0.1384***	0.1560***		0.2062***	0.2062***	0.2535***	0.2535***	0.1384***	0.1560***
		(0.0015)	(0.0018)	(0.0015)	(0.0014)		(0.0015)	(0.0015)	(0.0018)	(0.0018)	(0.0015)	(0.0014)
40-49 age		0.1882***	0.3265***	0.1286***	0.1439***		0.1882***	0.1882***	0.3265***	0.3266***	0.1286***	0.1439***
		(0.0019)	(0.0019)	(0.0019)	(0.0018)		(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0018)
50-64 age		0.1508***	0.3758***	0.1047***	0.1156***		0.1509***	0.1509***	0.3758***	0.3758***	0.1047***	0.1156***
		(0.0025)	(0.0022)	(0.0023)	(0.0023)		(0.0025)	(0.0025)	(0.0022)	(0.0022)	(0.0023)	(0.0023)
65 age		0.0889***	0.4063***	0.0498***	0.0596***		0.0889***	0.0889***	0.4062***	0.4064***	0.0498***	0.0597***
		(0.0125)	(0.0177)	(0.0111)	(0.0114)		(0.0125)	(0.0125)	(0.0177)	(0.0177)	(0.0111)	(0.0114)
experience	0.0006***	0.0010***	0.0012***	0.0008***	0.0009***	0.0007***	0.0010***	0.0010***	0.0012***	0.0012***	0.0008***	0.0009***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
experience2	0.0000***	-0.000***	0.000***	-0.000***	-0.000***	0.0000***	-0.000***	-0.000***	0.0000***	0.0000***	-0.000***	-0.000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
gender	-0.343***					-0.343***						
	(0.0010)					(0.0010)						
Agriculture	-0.471***	-0.049***			0.012***	-0.412***	-0.049***	-0.049***		-0.025***		0.0117***
	(0.0023)	(0.0027)			(0.0037)	(0.0023)	(0.0027)	(0.0027)		(0.0059)		(0.0037)
Fishery	-0.478***	-0.0194			-0.0370*	-0.384***	-0.0191	-0.0191		-0.0672*		-0.0373*
	(0.0262)	(0.0207)			(0.0212)	(0.0251)	(0.0207)	(0.0207)		(0.0344)		(0.0212)
Extractive indust.	0.0654***	0.0011			-0.041***	0.1287***	0.0022	0.0022		-0.045***		-0.041***
	(0.0049)	(0.0051)			(0.0063)	(0.0047)	(0.0051)	(0.0051)		(0.0100)		(0.0063)
Elect. Gas, Water	0.2164***	0.0921***			0.1167***	0.2691***	0.0929***	0.0929***		0.1043***		0.1175***
	(0.0021)	(0.0050)			(0.0063)	(0.0020)	(0.0050)	(0.0050)		(0.0102)		(0.0064)
Construction	-0.146***	-0.120***			-0.028***	-0.119***	-0.120***	-0.120***		-0.029***		-0.029***
	(0.0032)	(0.0032)			(0.0041)	(0.0031)	(0.0032)	(0.0032)		(0.0066)		(0.0041)

	(1) POLS	(2) FEi	(3) FEj	(4) FEij	(5) CONDij	(6) POLS	(7) FEi	(8) CONDi	(9) FEj	(10) CONDj	(11) FEij	(12) CONDij
Commerce	-0.227***	-0.075***			0.0048*	-0.207***	-0.075***	-0.075***		0.0005		0.0050**
	(0.0015)	(0.0018)			(0.0025)	(0.0015)	(0.0018)	(0.0018)		(0.0040)		(0.0025)
Food; Accommodation	-0.430***	-0.144***			0.0003	-0.440***	-0.144***	-0.144***		-0.0028		0.0010
	(0.0033)	(0.0051)			(0.0074)	(0.0032)	(0.0051)	(0.0051)		(0.0119)		(0.0074)
Transport; Communication	-0.152***	-0.027***			-0.014***	-0.134***	-0.027***	-0.027***		-0.017**		-0.013***
	(0.0016)	(0.0027)			(0.0043)	(0.0015)	(0.0027)	(0.0027)		(0.0069)		(0.0044)
Finance	0.2688***	0.0497***			0.0061	0.2708***	0.0474***	0.0474***		-0.027***		0.0042
	(0.0019)	(0.0035)			(0.0054)	(0.0018)	(0.0035)	(0.0035)		(0.0086)		(0.0054)
Housing	-0.184***	-0.088***			-0.009***	-0.197***	-0.089***	-0.089***		-0.015***		-0.009***
	(0.0015)	(0.0018)			(0.0024)	(0.0015)	(0.0018)	(0.0018)		(0.0038)		(0.0024)
Education	-0.337***	-0.168***			0.0139***	-0.324***	-0.168***	-0.168***		0.0140**		0.0143***
	(0.0019)	(0.0032)			(0.0043)	(0.0018)	(0.0032)	(0.0032)		(0.0070)		(0.0044)
Health; Social	-0.368***	-0.074***			0.0056	-0.338***	-0.075***	-0.075***		-0.0043		0.0058
	(0.0018)	(0.0032)			(0.0043)	(0.0018)	(0.0032)	(0.0032)		(0.0069)		(0.0043)
Other services	-0.284***	-0.096***			0.0057*	-0.269***	-0.097***	-0.097***		-0.0014		0.0063*
	(0.0020)	(0.0026)			(0.0034)	(0.0019)	(0.0026)	(0.0026)		(0.0055)		(0.0035)
Household Serv.	-0.368***	-0.0290			0.0295	-0.377***	-0.0261	-0.0261		-0.0145		0.0300
	(0.0557)	(0.0279)			(0.0255)	(0.0535)	(0.0279)	(0.0279)		(0.0413)		(0.0255)
International Org.	0.1374***	-0.0007			-0.0009	0.1937***	0.0035	0.0035		0.0279		-0.0039
	(0.0364)	(0.0206)			(0.0198)	(0.0350)	(0.0206)	(0.0206)		(0.0320)		(0.0198)
Medium firm	0.2571***	0.0695***	0.0296***	0.0318***	0.0314***	0.2514***	0.0690***	0.0690***	0.0297***	0.0298***	0.0318***	0.0314***
	(0.0011)	(0.0010)	(0.0020)	(0.0013)	(0.0013)	(0.0010)	(0.0010)	(0.0010)	(0.0020)	(0.0020)	(0.0013)	(0.0013)
Large firm	0.3439***	0.1048***	0.0493***	0.0596***	0.0583***	0.3115***	0.1036***	0.1036***	0.0497***	0.0496***	0.0595***	0.0583***
	(0.0011)	(0.0012)	(0.0027)	(0.0017)	(0.0017)	(0.0010)	(0.0012)	(0.0012)	(0.0027)	(0.0027)	(0.0017)	(0.0017)
LMA Dummy	0.2896***	0.0179***	0.0190***	-0.006**	-0.007**							
	(0.0013)	(0.0020)	(0.0051)	(0.0032)	(0.0032)							
Time effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	5.4860***	6.9149***	6.6693***	7.0078***		5.3945***	6.9070***		6.6609***		7.0062***	
	(0.0072)	(0.0025)	(0.0049)	(0.0032)		(0.0070)	(0.0026)		(0.0052)		(0.0034)	
Observations	2328018	2328018	2328018	2328018	2328018	2328018	2328018	2328018	2328018	2328018	2328018	2328018
R^2	0.522	0.203	0.283	0.181	0.929	0.560	0.203	0.908	0.283	0.801	0.181	0.929
Adjusted R^2	0.522	0.142	0.242	0.048	0.921	0.560	0.142	0.901	0.242	0.790	0.048	0.921

Source: Author's calculation from RAIS-Migra (MTE).

Notes: Significant at *5%, **10%, ***1%.