Ann Reg Sci (2012) 49:17–34
DOI 10.1007/s00168-010-0432-2

ORIGINAL PAPER

Inter-regional wage differentials with individual heterogeneity: evidence from Brazil

Ricardo S. Freguglia · Naercio A. Menezes-Filho

Received: 25 February 2010 / Accepted: 27 December 2010 / Published online: 18 January 2011

Abstract This paper uses administrative data to follow Brazilian workers over time and examine what happens to the inter-regional wage differentials after controlling for unmeasured workers' characteristics that are fixed over time. Since the data allow us to track the same workers over the years, we are in the unusual position of obtaining the individual wages before and after the migration process. As a significant share of workers changed States in the sample period, it is possible to examine to what extent the wage differentials reflect the concentration of high-skilled individuals in some States. The results show that the overall wage variability across States drops to almost one third of its original value and the ranking of the State effects is significantly altered after we take into account the workers' fixed effects. A great deal of the inter-regional differentials, therefore, reflects differences in the average ability of workers across States.

JEL Classification R23 · J31 · J24

1 Introduction

Evidence of regional differences in average wages has been produced in a variety of countries with distinct institutional and structural arrangements. Persistent wage

R. S. Freguglia (✉)
Department of Economics, Federal University of Juiz de Fora, Campus Universitário, Juiz de Fora, Minas Gerais 36036-330, Brazil
e-mail: ricardo.freguglia@ufjf.edu.br

N. A. Menezes-Filho
Department of Economics, Insper Institute of Education and Research and University of São Paulo, Rua Quatá, 300, São Paulo 04546-042, Brazil
e-mail: naercioamf@isp.edu.br

differentials across regions could exist to compensate for differences in amenities (climate or pollution) or in the cost of living. They could also be the result of the concentration of human capital in some cities or regions, which generate knowledge spillovers and differential productivity (Lucas 1988). Different explanations for the existence of regional differences in wages have different implications in terms of economic policies.

Many studies have decomposed the wage differentials in terms of differences in the market value of individual characteristics, such as, education and work experience, and differences in the rates of return of those characteristics. Even after adjusting for the cost of living and for the unequal spatial distribution of workers' characteristics, most studies find that regional wage differentials tend to persist over time.[1]

This paper aims at examining the role of differences in unmeasured characteristics among workers living in different regions as an additional explanation for the observed regional wage differentials. The central idea is that the observed wage premium received by workers in certain areas may contain a return for unmeasured attributes. Therefore, estimations that disregard these attributes may be upward biased and ascribe to other factors, such as compensating differentials or spillovers, the role of individual-level productivity. This, in turn, could give rise to wrong policy recommendations, such as providing subsidies to the development of local industry, when in fact investments in human capital are necessary.

In the labor economics literature, several industry-level studies have used the unobserved heterogeneity approach to wage determination.[2] These studies typically find that the estimated industry wage premium contains a return to unmeasured workers' attributes, though the magnitude of this effect differs across studies. Murphy and Topel (1987), Keane (1993), and Abowd et al. (1999) find that unobserved heterogeneity explains between 66 and 90%, of the apparent differential in log-wages across industries. Carruth et al. (2004) reach a similar conclusion in a recent study based on industry switchers in the United Kingdom. They explicitly address the role of unobserved heterogeneity as an explanation of observed inter-industry differentials, finding that unmeasured abilities explain 90% of inter-industry wage differences. An exception is Krueger and Summers (1988), which demonstrates the potential importance of efficiency wages theories and find little evidence to support unobserved heterogeneity as an explanation to industry pay.

We can also assess the role of unobserved heterogeneity in wage differentials in the regional context. This kind of investigation is especially interesting when income inequality among regions is high, which is usually the case in developing countries. The sources of inter-regional wage differentials in Brazil, for example, have been the target of several studies, which conclude that these differentials are high, persistent and only partially explained by differences in human capital differences across States.[3]

[1] See, for example, García and Molina (2002), Pereira and Galego (2007) and Motellón et al. (2009).

[2] See Murphy and Topel (1987), Krueger and Summers (1988), Gibbons and Katz (1992), Keane (1993) and Abowd et al. (1999).

[3] Bacha and Taylor (1978), Dabos and Psacharopoulos (1991), Pinheiro and Ramos (1994), Barros and Mendonça (1995), Cowell et al. (1996), Gatica et al. (1995) and Menezes-Filho (2001).

More recently, Arbache and Carneiro (1999) detect inter-regional and inter-industry differences in wages due to union power in Brazil. Azzoni and Servo (2002) examine wage differentials among the largest metropolitan regions in Brazil in the 1990s and show that, although the cost of living does have a role in explaining wage inequality in Brazil, the remaining regional differentials are still important. Ferreira et al. (2006) find that over 55% of the wage differentials between northeast and southeast regions are due to disparities in educational attainment. To the best of our knowledge, no study thus far has examined the role of unobserved heterogeneity at the individual level as an explanation of observed inter-regional wage differentials, which is the main contribution of this paper.[4]

We examine the inter-regional wage differentials of Brazilian workers using longitudinal data. Since the data permit us to track the same workers over the years, we are in the unusual position of observing individual wages before and after the migration process across States. We use a fixed-effects model to control for unobserved workers characteristics, such as ability and motivation, and apply the methodology of Haisken-DeNew and Schmidt (1997) to estimate standard errors and compare the wage variability of different specifications.

The results show that the estimated wages differentials across regions are substantially lower than the simple OLS specifications would suggest. A large amount of the wage variability is a consequence of heterogeneity across individuals living in different States. Although the estimates confirm the existence of persistent earnings differences across regions, their size and statistical significance are considerably reduced when controls for unmeasured abilities are allowed for.

This paper is organized as follows. Section 2 describes the data and presents some preliminary evidence about the inter-regional wage inequality in Brazil. In Sect. 3, we describe the econometric methodology, which combines the fixed-effects model and the two-step approach of Haisken-DeNew and Schmidt (1997) to estimate the wage differentials. Section 4 presents the basic econometric results. Robustness tests are presented in Sect. 5, while Sect. 6 concludes.

2 Data and initial evidence

Our empirical analysis is based on longitudinal micro-data from the Relatório Anual de Informações Sociais—Migração (RAIS-Migra), of the Labor Ministry of Brazil, between 1995 and 2002. This database is a special edition of RAIS, an annual administrative survey used by the government to identify workers eligible to receive social benefits and monitor the labor market. RAIS is a census of all formal sector workers in Brazil (about 24 million) and has information on a detailed set of workers

[4] It is important to clarify the main idea behind the unobserved individual characteristics. When hiring and monitoring workers, employers consider both measured (to the econometrician) abilities, such as schooling and work experience and unmeasured ones, such as ability, motivation, talent and effort. Regions that pay relatively high wages to workers, regardless of their human capital characteristics, may be aggregating individuals with those unmeasured skills. Thus, the regional wage premium will contain a return on these unmeasured attributes and the estimated inter-regional wage differentials will be biased upwards.

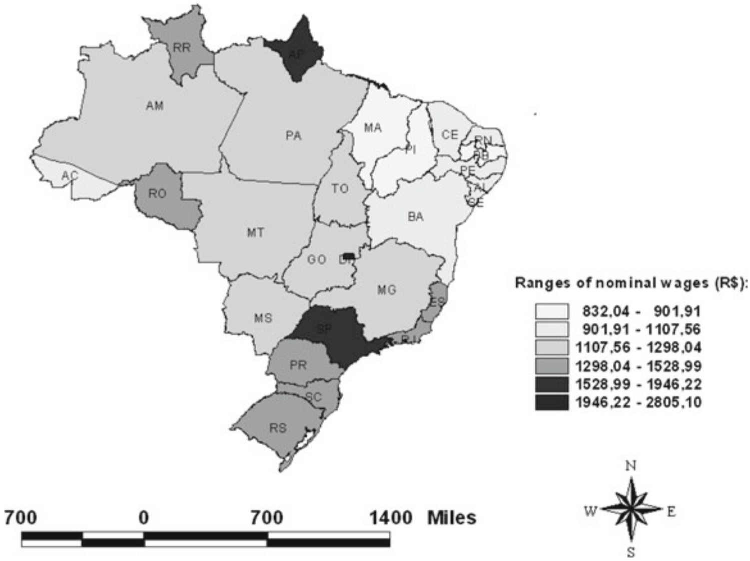

Fig. 1 Distribution of nominal monthly wages. *Source* RAIS-Migra (1996–2002)

characteristics, such as, gender, age, education, occupation and wages. Informal and self-employed workers are not covered by RAIS.

RAIS-Migra follows longitudinally the career of all individuals in the formal labor market, allowing researchers to investigate questions that are relevant to the Brazilian labor market. In this paper, we consider the formal sector workers living in any of the 26 Brazilian states plus the Federal District (Brazilian capital). Due to the large number of observations in the original dataset, which makes estimation unfeasible, a 1% random sample was generated in order to estimate the econometric models. The final sample has 611,632 pooled and balanced observations, with 76,454 individuals by year.

As we can observe the same worker in different years in our sample (1995–2002), migrants are identified in the RAIS-Migra data by the reported place of work.[5] When the State where the individual works in the current year (t) is different from the State where he worked in the previous year ($t-1$), he is identified as a migrant. The percentage of movers among Brazilian states is about 1% (see Table 3 in Appendix). A description of migrant flows by State is presented in Fig. 4 in the Appendix.

The main variable of interest is the monthly wage[6] deflated by the consumer and by the regional cost of living indexes.[7] Figures 1 and 2 describe the distribution of nominal and real monthly wages among the Brazilian states, using the pooled data from RAIS. The figures show that the variability of real monthly wages across states is

[5] The place of birth is not reported in the RAIS-Migra.

[6] Expressed in R$ (Brazilian Reais).

[7] The ICV (a Brazilian cost of living index) used in this paper was computed by Azzoni et al. (2003).

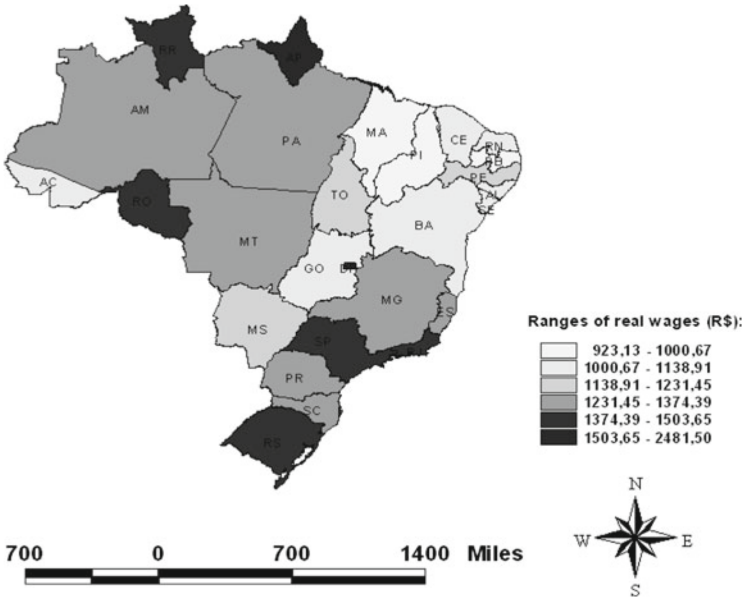

Fig. 2 Distribution of real monthly wages. *Source* RAIS-Migra (1996–2002)

significantly lower than that of nominal wages. Hence, in the remaining of the analysis will focus on real monthly wages, net of the regional differences in living costs.

The independent variables used in the empirical exercises below are age, tenure (monthly), gender, four educational levels, two-digit industry and occupation levels, size of establishment[8], and year dummies. The sample is composed of workers between the ages of 14 and 65, with non-zero monthly income. Data definitions and summary statistics are presented in Table 4 of the Appendix.

3 Methodology

To compute the wage differentials, we carry out a procedure originally developed by Haisken-DeNew and Schmidt (1997) (HDS hereafter), changing the focus from industries to states. The HDS approach, described in Eqs. (1) and (2), provides economically sensible coefficients and correct standard errors in a single regression step:

$$\ln w_{ij} = \alpha + \beta X_i + \delta R_j + \varepsilon_{ij}, \qquad i = 1, \ldots, N;\ j = 1, \ldots, K \tag{1}$$

where $ln\ w_{ij}$ is the natural logarithm of the real monthly wage for worker i in State j, α is the constant, X_i is a vector of individual-level control variables, R_j is the vector of state dummies for worker i, δ is the vector of coefficients associated with the state dummies, β is the coefficient vector of the control variables, and ε_{ij} is the random

[8] The size of the establishment is potentially relevant to the inter-regional wage differentials. For instance, large establishments tend to pay higher wages and be concentrated in some regions.

disturbance term. Since in Eq. (1) the cross-product matrix of regressors is not of full rank, a linear constraint is imposed on the δ as follows:

$$\sum_j n_j \delta_j = 0 \tag{2}$$

where n_j is the percentage of workers—the employment share—in each state j.

The advantage of the restricted least squares (RLS) procedure of HDS is that all k dummy coefficients and standard errors are reported, i.e., the results are independent of the choice of the reference category. This procedure corrects the problems of the traditional methodology of overstating differential standard errors and understating the overall dispersion. The coefficients can be interpreted as percentage-point deviations from the States' weighted average wages.

We use two different measures to describe the overall dispersion in State wages. First, we calculated the weighted average absolute differential:

$$|\delta| = \sum_j \left| n_j \delta_j \right| \tag{3}$$

As our second measure, we calculate the standard deviation of the state wage differentials (SD).[9]

$$\mathrm{SD}(\delta) = \sqrt{\sum_j n_j \delta_j^2 - \sum_j n_j \sigma_j^2} \tag{4}$$

where σ_j^2 is the variance of $\hat{\delta}_j$.

Initially, we estimate wage regressions by pooled ordinary least squares (POLS) including a vector of year dummies. The main problem with this specification is the possible correlation between the error term (ε_{ij}) and the State dummies, if workers with the best unobserved attributes tend to locate in the same regions. If this is the case, such endogeneity would bias the estimated wage differentials. With longitudinal data, this problem may be solved by allowing for individual-specific fixed-effects (c_i):

$$\begin{aligned} \ln w_{ijt} = c_i + \beta X_{it} + \delta R_{jt} + T_t + \varepsilon_{ijt}, \quad & i = 1, \ldots, N;\ j = 1, \ldots, K, \\ t = 1996, \ldots, 2002 & \end{aligned} \tag{5}$$

where $ln\ w_{ijt}$ is the natural logarithm of the real monthly wage of worker i in State j in year t, X_{it} is a vector of individual-level control variables, R_{jt} is the vector of state dummies, c_i are individual-specific effects, T_t are the time dummies, δ is the vector of coefficients associated with the state dummies, β is the coefficient vector of the control variables, and ε_{ijt} is the random disturbance term associated with worker i in State j and year t.

[9] The correction for the least squares sampling error is the second term in Eq. (4). See HDS.

Several econometric methods can be used to estimate model (5), depending on the identifying assumptions one is willing to make. In this paper, we use and compare the results of pooled OLS (POLS), random and fixed-effects models. Both POLS and random effects models require absence of correlation between the specific effects and the explanatory variables, including the State dummies. The validity of this assumption can assessed by means of a Hausman test. Fixed-effect models do not require that assumption, but one cannot estimate the effects of time-invariant variables, such as age and education, on wages, since they will be absorbed into the individual-specific fixed effects. Moreover, fixed-effect models use only the variation within-groups to estimate the coefficients of interest, instead of the between-groups (long run) variation.

In order to identify the effect of the State dummies on wages, conditional on the fixed effects, we have to rely on the individuals that change States over time, i.e., the migrants. Therefore, instead of estimating the impact of migration on wages, we allow the State of residence to affect wages through a set of State-specific dummies, which can be identified in the fixed-effect models because individuals change their place of work over time. In the estimation procedure, the variables will be transformed into deviations from individual-specific means in order to eliminate the fixed effects.[10]

4 Results and discussion

The preliminary analysis is based on levels regressions for each year of the sample, to describe the overall variability of real monthly wage differences across States. The functional form of these regressions is based on the Mincerian equation (Mincer 1974), expanded by a set of explanatory variables (Eq. 1 above).

Figure 3 reports the overall variability[11]—SD(δ)—of the wage differentials for each year between 1996 and 2002. Three different specifications were used, which progressively include more controls (see notes in the bottom of the figure). In general, the overall variability is between 17 and 25% in the seven years considered. As expected, the inclusion of more controls (third specification) reduces the overall variability. Despite a marked decline over time, overall variability of wage differentials was still quite large in 2002, in the range of 17%.

Table 1 presents the main results of the paper.[12] The first column presents a pooled OLS analysis of wage differentials among workers from the 26 Brazilian states plus the Federal District. The coefficients are large and statistically significant, and the overall variability—SD(δ)—is 18.5%. The Breusch-Pagan test, reported in the bottom of the table, rejects the null of no serial correlation, an indication of omitted individual-specific effects. The second column presents the results of the random effects model. One can notice that the estimated coefficients are almost all significant and that the

[10] It would also be interesting to detect any spatial correlation patterns across States in Brazil, but this would be unfeasible at the moment given our sample size (622,632 observations), thus we defer a detailed treatment of this issue to future work.

[11] The overall variability is the employment-weighted adjusted standard deviation of the States log wage differentials. See Eqs. 2–4 above for details.

[12] The control variables included in the regression are listed at the bottom of the Table 1. Their estimated coefficients are available from the authors upon request.

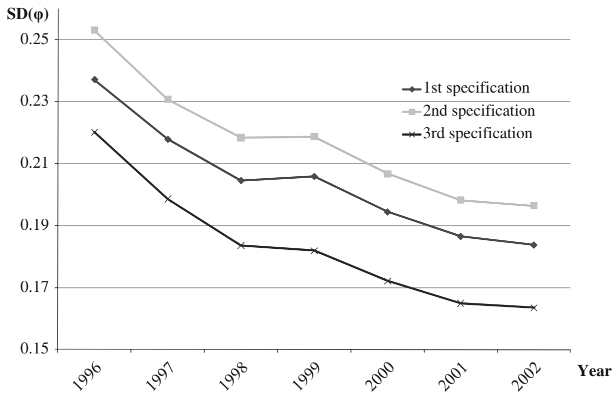

Fig. 3 Overall variability of wage differentials from level regressions *Source* RAIS-Migra (1996-2002) *Notes*: The reported SD(δ)'s are based on HDS procedure. The OLS regressions have three different specifications. The first specification reports the raw inter-regional wage differentials. The second includes personal controls: age, age squared, tenure, tenure squared, 4 education dummies, and gender dummy. The third includes personal controls (see 2nd specification) and workplace controls: 25 industry dummies (2 digit), 82 occupation dummies (2 digit), 6 firm size dummies, and time dummies. The estimated coefficients of the control variables are available from the authors on request

Table 1 Wage differentials

Dependent variable: logarithm of real wages			
States	(1)	(2)	(3)
RO	0.122*** (0.011)	−0.011 (0.008)	0.245*** (0.043)
AC	−0.080*** (0.012)	0.045*** (0.009)	−0.050 (0.066)
AM	0.044*** (0.008)	0.005 (0.005)	0.221*** (0.029)
RR	0.198*** (0.021)	−0.046*** (0.015)	0.338*** (0.060)
PA	−0.124*** (0.006)	0.053*** (0.004)	0.058** (0.023)
AP	0.219*** (0.017)	−0.102*** (0.012)	0.024 (0.054)
TO	−0.144*** (0.013)	0.071*** (0.009)	0.503*** (0.045)
MA	−0.316*** (0.006)	0.118*** (0.005)	−0.018 (0.033)
PI	−0.275*** (0.009)	0.100*** (0.006)	−0.122*** (0.042)
CE	−0.347*** (0.005)	−0.0110 (0.008)	−0.019 (0.019)
RN	−0445*** (0.007)	0.123*** (0.005)	0.018 (0.035)
PB	−0.728*** (0.006)	0.172*** (0.004)	0.076*** (0.028)
PE	−0.194*** (0.004)	0.067*** (0.003)	0.093*** (0.018)
AL	−0.255*** (0.007)	0.089*** (0.005)	0.039 (0.036)
SE	−0.254*** (0.008)	0.084*** (0.006)	−0.027 (0.034)
BA	−0.192*** (0.004)	0.073*** (0.003)	0.061*** (0.015)
MG	−0.016*** (0.002)	0.012*** (0.002)	−0.015 (0.010)
ES	−0.037*** (0.006)	0.014*** (0.004)	−0.021 (0.022)
RJ	−0.051*** (0.002)	0.000 (0.002)	−0.005 (0.009)
SP	0.149*** (0.001)	−0.043*** (0.001)	−0.066*** (0.005)

Table 1 continued

Dependent variable: logarithm of real wages			
States	(1)	(2)	(3)
PR	0.049*** (0.003)	0.000 (0.002)	0.014 (0.011)
SC	0.114*** (0.004)	−0.026*** (0.003)	0.020 (0.016)
RS	0.137*** (0.003)	−0.031*** (0.002)	0.074*** (0.015)
MS	−0.070*** (0.007)	0.015*** (0.005)	0.008 (0.027)
MT	0.046*** (0.008)	0.022*** (0.006)	0.187*** (0.026)
GO	−0.138*** (0.005)	0.056*** (0.004)	0.014 (0.017)
DF	0.360*** (0.005)	−0.123*** (0.004)	−0.005 (0.012)
SD(δ)	0.185	0.149	0.069
$\lvert\delta\rvert$	0.144	0.111	0.050
Observations	611,632	611,632	611,632
Individuals	76,460	76,454	76,454
R^2 within	–	0.0947	0.1046
R^2 between	–	0.4668	0.1843
R^2 overall	0.0491	0.4216	0.1595
Hausman	–	$\chi^2(145) = 80{,}265.30$***	$\chi^2(145) = 80{,}265.30$***
Breusch-Pagan	–	$\chi^2(1) = 1{,}200{,}000.00$***	$\chi^2(1) = 1{,}200{,}000.00$***

Source RAIS-Migra (1996–2002)
Notes The reported SD(δ) is based on HDS procedure. $|\delta|$ is the weighted average absolute differential. Column (1) is a pooled OLS regression and includes a constant plus personal, workplace and time controls: tenure, tenure squared, age, age squared, and dummies for gender, education (4), industry—2 digit (25), occupation—2 digit (82), and year (7). Column (2) is a random effects regression and follows (1). Column (3) is a fixed-effects regression and omits the time-invariant variables: age, age squared, and gender and education dummies. NT is the total number of observations; N is the number of individuals. Standard errors in parentheses
*** Significant at 1% level; ** Significant at 5% level; * Significant at 10% level. See text for details

overall variability is about 15%, just 3.5 percentage points lower than in the POLS regression.

The result of a Hausman test (comparing the random- and fixed-effects models) rejects the null of no correlation between the specific effects and the explanatory variables. We therefore focus the analysis on the last column of Table 1, which presents the results of the fixed-effects models. One can notice from these results that the number of significant coefficients drops by half. There are only 11 significant coefficients at the 1% level, in comparison with 22 in the random effects regression. This result contrasts even more with the POLS regression, where all coefficients were significant at the 1% level. Additionally, one can notice that most of the coefficients from the fixed-effects regression are lower (in absolute values) in comparison with the POLS coefficients.

São Paulo (SP) is the richest State of Brazil. Its coefficient drops from 0.149 in the POLS specification to −0.066 in the fixed-effects specification. This means that, after taking into account unmeasured attributes, such as ability and motivation, workers actually receive a wage drop when they move to São Paulo.

The most important results concern the overall variability of the wage differentials. While the employment-weighted adjusted standard deviation of the States log wage differentials—SD(δ)—is 18.5% in the OLS regression, the fixed-effect regression exhibits an overall variability of only 6.9%. Similarly, the weighted average absolute differential—$|\delta|$—falls from 14.4 to 5%. These results show that there is a large part of inter-regional wage differences—approximately 63%—that can be explained by the individual heterogeneity.[13] Hence, our results are in line with that of Carruth et al. (2004). Although their approach refers to wage differentials across sectors in the United Kingdom, both studies emphasize the importance of unobserved worker traits.

Therefore, after individual fixed effects are controlled for, the inter-regional wage differentials still exist, but lose importance. As a consequence, the regional effect, i.e., the effect on wages caused by the State where the workers are living, has a more limited role.

5 Robustness tests

We now report the results of some selected robustness exercises to test the sensitivity of our results to the maintained assumptions. Panel A of Table 2 contains our results so far, as a baseline. Once again, one can notice the drop in wage differentials after the individual-specific effects are taken into account.

Another suitable comparator to our fixed-effects estimates is the first-differences specification. Panel B presents the results of this exercise. There is little difference from the fixed-effects model in terms of the inter-regional wage dispersion. As the model of unobserved effects is generally defined with serially uncorrelated idiosyncratic errors, the fixed-effects estimator is preferable with respect to first differences one (Wooldridge 2002).

Panels C and D contain the results for men and women separately. It is well known that rates of return to education differ markedly between men and women, and there are important differences in the labor force participation by gender, which can impact wages. Using the fixed-effects approach, the estimated differentials are greater for women than for men, although a higher proportion of the overall variation in wages is explained in the case of men. The main findings are not sensitive to this partition of the data, however, and hence are not driven by differences in labor force participation by gender.

Panels E and F present the results separately for the more educated and less educated workers, respectively. Interestingly enough, the overall wage variability is higher for the more qualified workers, even after controlling for unmeasured abilities, which suggest that the remaining wage differentials may be related to knowledge spillovers.

We also examine the behavior of the wage variability separately for the low tenure and high tenure workers. The overall variability is much higher for the more experi-

[13] The proportion of the inter-regional wage differentials explained by unobserved individual heterogeneity is computed by $[1-(0.069/0.185)]100 = 62.76\%$.

Table 2 Robustness tests

Description	Sample selection	$\|\delta\|$	SD(δ)
A. Basic results	POLS: $N = 611{,}632$	0.144	0.185
(Table 1)	RE: $N = 611{,}632$	0.111	0.149
	FE: $N = 611{,}632$	0.050	0.069
B. First differences	FD: $N = 535{,}178$	0.061	0.072
C. Men only	FE: $N = 337{,}262$	0.056	0.073
D. Women only	FE: $N = 274{,}370$	0.064	0.104
E. Skilled	FE: $N = 103{,}391$	0.082	0.106
F. Unskilled	FE: $N = 508{,}241$	0.059	0.074
G. Low tenure	FE: $N = 305{,}798$	0.046	0.069
H. High tenure	FE: $N = 305{,}834$	0.108	0.135

Source RAIS-Migra 1996–2002
Notes See notes to Table 1

enced workers. This again suggests that the remaining wage differentials are strongly related to human capital.

6 Conclusions

In this paper, we used longitudinal data to estimate the inter-regional wage differentials in Brazil, after controlling for unmeasured workers characteristics, through fixed-effects methods. We find that the spatial wage differentials are much lower than previously thought. A large amount of the wage variability across States is a consequence of unmeasured differences among individuals that cannot be removed through simple OLS estimation. Almost 63% of the wage differentials can be explained by the unobservable individual heterogeneity. Although wage differentials still exist, their size and statistical significance are considerably lower.

These results have important implications in terms of economic policy. Based on the simple OLS regressions, one would have thought that workers living in Maranhão (MA), one of Brazil's poorest States, earned 31% lower wages than the average Brazilian State. Hence, there was a strong case for policy intervention in terms of subsidies, infrastructure, etc. But the fixed-effects models reveal that the Maranhão wage differential is basically zero. This means that policy should actually focus on human capital development, especially at early stages of the life cycle, where children are still developing their cognitive and non-cognitive skills, which will be very important for the wages they will earn over the life cycle.

Appendix

See Tables 3, 4 and Fig. 4.

Table 3 Flow of inter-state migrants

Year	Migrants		Non-migrants		Total	
	Frequence	%	Frequence	%	Frequence	%
1995	–	–	76,454	–	76,454	–
1996	605	0.79	75,849	99.21	76,454	100
1997	608	0.80	75,846	99.20	76,454	100
1998	624	0.82	75,830	99.18	76,454	100
1999	524	0.69	75,930	99.31	76,454	100
2000	617	0.81	75,837	99.19	76,454	100
2001	561	0.73	75,893	99.27	76,454	100
2002	527	0.69	75,927	99.31	76,454	100
Total	4,066	0.67	607,566	99.33	611,632	100

Note The total refers to migrant workers plus non-migrant workers
Source RAIS-Migra 1996–2002

Table 4 Variable definitions and summary statistics

Variable	Definition and description	*N*	Mean	SD
Dependent variable				
Log of real wage	Log of monthly wage deflated by IPCA (price index) and ICV (cost of living index)	611,632	6.74	0.92
Independent variables				
Tenure	Experience in the current job (in months)	611,632	116.18	87.07
Age	Individual age (in years) stated on RAIS	611,632	38.46	9.75
Gender	(1,0) if male	337,262	55.14	–
Education level				
Primary	(1,0) education dummy	206,497	33.76	–
Secondary	(1,0) education dummy	123,037	20.11	–
High school	(1,0) education dummy	178,707	28.22	–
College	(1,0) education dummy	103,391	16.9	–
Region				
RO	(1,0) region dummy – Rondônia	3,289	0.54	–
AC	(1,0) region dummy – Acre	2,534	0.41	–
AM	(1,0) region dummy – Amazonas	6,875	1.12	–
RR	(1,0) region dummy – Roraima	871	0.14	–
PA	(1,0) region dummy – Pará	11,335	1.85	–
AP	(1,0) region dummy – Amapá	1,335	0.22	–
TO	(1,0) region dummy – Tocantins	2,433	0.4	–
MA	(1,0) region dummy – Maranhão	9,964	1.63	–
PI	(1,0) region dummy – Piauí	5,348	0.87	–
CE	(1,0) region dummy – Ceará	18,489	3.02	–

Table 4 continued

Variable	Definition and description	*N*	Mean	SD
RN	(1,0) region dummy – Rio G$^{\text{de}}$ Norte	7,021	1.15	–
PB	(1,0) region dummy – Paraíba	9,975	1.63	–
PE	(1,0) region dummy – Pernambuco	21,223	3.47	–
AL	(1,0) region dummy – Alagoas	7,986	1.31	–
SE	(1,0) region dummy – Sergipe	6,193	1.01	–
BA	(1,0) region dummy – Bahia	28,198	4.61	–
MG	(1,0) region dummy – Minas Gerais	65,591	10.72	–
ES	(1,0) region dummy – Espírito Santo	10,261	1.68	–
RJ	(1,0) region dummy – Rio de Janeiro	65,682	10.74	–
SP	(1,0) region dummy – São Paulo	181,451	29.67	–
PR	(1,0) region dummy – Paraná	35,677	5.83	–
SC	(1,0) region dummy – Santa Catarina	23,398	3.83	–
RS	(1,0) region dummy – Rio G$^{\text{de}}$ do Sul	43,615	7.13	–
MS	(1,0) region dummy – M. Grosso Sul	7,090	1.16	–
MT	(1,0) region dummy – Mato Grosso	6,519	1.07	–
GO	(1,0) region dummy – Goiás	15,487	2.53	–
DF	(1,0) region dummy – Dist. Federal	13,792	2.25	–
Size of establishment				
Until 19	(1,0) size dummy	92,752	15.16	–
From 20 to 99	(1,0) size dummy	89,381	14.61	–
From 100 to 249	(1,0) size dummy	67,145	10.98	–
From 250 to 499	(1,0) size dummy	62,906	10.28	–
From 500 to 999	(1,0) size dummy	64,081	10.48	–
1,000 or more	(1,0) size dummy	235,367	38.48	–
Industry (2 digit)				
Public administration	(1,0) industry dummy	239,690	39.19	–
Real estate, renting and business activities	(1,0) industry dummy	37,910	6.2	–
Agriculture, hunting, forestry, fishing and related service activities	(1,0) industry dummy	17,430	2.85	–
Food products and beverages manufacturing	(1,0) industry dummy	19,645	3.21	–
Communication and lodging	(1,0) industry dummy	34,331	5.61	–
Rubber, tobacco, leather, fur and similar goods manufacturing	(1,0) industry dummy	4,968	0.81	–
Wholesale	(1,0) industry dummy	10,760	1.76	–
Retail sale	(1,0) industry dummy	42,231	6.9	–
Construction	(1,0) industry dummy	11,982	1.96	–
Electrical machinery, radio, tv and communication equipment	(1,0) industry dummy	3,801	0.62	–
Education	(1,0) industry dummy	23,992	3.92	–
Mining and quarrying	(1,0) industry dummy	2,570	0.42	–
Footwear manufacturing	(1,0) industry dummy	3,333	0.54	–

Table 4 continued

Variable	Definition and description	*N*	Mean	SD
Machinery and equipment manufacturing	(1,0) industry dummy	6,906	1.13	–
Metal products manufacturing	(1,0) industry dummy	11,795	1.93	–
Chemicals and chemical products manufacturing	(1,0) industry dummy	11,633	1.9	–
Textiles and wearing apparel manufacturing	(1,0) industry dummy	11,476	1.88	–
Financial intermediation	(1,0) industry dummy	18,834	3.08	–
Wood and of procucts of wood and cork manufacturing	(1,0) industry dummy	5,902	0.96	–
Transport equipment manufacturing	(1,0) industry dummy	9,058	1.48	–
Health and social work	(1,0) industry dummy	26,765	4.38	–
Non-metallic mineral product manufacturing	(1,0) industry dummy	5,064	0.83	–
Paper and paper products manufacturing	(1,0) industry dummy	7,350	1.20	–
Public utility services	(1,0) industry dummy	13,572	2.22	–
Transport, storage and communications	(1,0) industry dummy	30,634	5.01	–
Occupation (2 digit)				
Subgroup 01	(1,0) if Chemists, physicists and related workers	552	0.09	–
Subgroup 02	(1,0) if Engineers, architects and related workers	4,124	0.67	–
Subgroup 03	(1,0) if Technicians, design technicians and related workers	14,322	2.34	–
Subgroup 04	(1,0) if Crew officials, pilots and related workers (inland, coastal and deep-sea)	374	0.06	–
Subgroup 05	(1,0) if Biologists, agricultural engineers and similar workers	672	0.11	–
Subgroup 06	(1,0) if Medical doctors, oral surgeons, veterinary doctors, nursing professionals	12,202	1.99	–
Subgroup 08	(1,0) if Statisticians, computer systems analysts and related workers	3,239	0.53	–
Subgroup 09	(1,0) if Economists, administrators, accountants and related workers	4,663	0.76	–
Subgroup 12	(1,0) if Lawyers	1,033	0.17	–
Subgroup 13	(1,0) if Teachers	75,365	12.32	–
Subgroup 15	(1,0) if Writers, journalists, announcers and related workers	1,193	0.2	–

Table 4 continued

Variable	Definition and description	*N*	Mean	SD
Subgroup 16	(1,0) if Sculptors, artistic painters, photographers and related workers	303	0.05	–
Subgroup 17	(1,0) if Musicians, artists, entertainment industry managers and producers	330	0.05	–
Subgroup 18	(1,0) if Sports referees, professionals athletes and related workers	700	0.11	–
Subgroup 19	(1,0) if Scientific, technical and artistic occupation professionals	2,470	0.4	–
Subgroup 21	(1,0) if Senior members of the legislative, executive and judiciary powers	25,042	4.09	–
Subgroup 22	(1,0) if Diplomats	14	0	–
Subgroup 23	(1,0) if Directors of enterprises	1,819	0.3	–
Subgroup 24	(1,0) if Enterprise managers	9,037	1.48	–
Subgroup 30	(1,0) if Intermediaries, administrative, accountancy and finance managers	8,822	1.44	–
Subgroup 31	(1,0) if Public business administration agents	70,390	11.51	–
Subgroup 32	(1,0) if Secretaries, stenographers and related workers	7,080	1.16	–
Subgroup 33	(1,0) if Accounting services workers, cashiers and related workers	11,867	1.94	–
Subgroup 34	(1,0) if Accounting machine, calculator and data processor operators	3,457	0.57	–
Subgroup 35	(1,0) if Telecommunications and transportation services managers	989	0.16	–
Subgroup 36	(1,0) if Supervisors, collectors, forwarders in public transportation (except train)	4,448	0.73	–
Subgroup 37	(1,0) if Mail classifiers, mailers and messengers	3,109	0.51	–
Subgroup 38	(1,0) if Telephone operators, telegraphers and related workers	2,244	0.37	–
Subgroup 39	(1,0) ifAdministrative services and related workers	56,525	9.24	–
Subgroup 41	(1,0) if Commercial salespersons (wholesale and retail)	233	0.04	–
Subgroup 42	(1,0) if Sales and purchasing supervisors, purchasing agents and related workers	4,449	0.73	–
Subgroup 43	(1,0) if Technical sales agents and sales representatives	1,704	0.28	–
Subgroup 44	(1,0) if Insurance, real estate and securities brokers, sales agents, auctioneers	325	0.05	–
Subgroup 45	(1,0) if Wholesale, retail seller, street vendors and related workers	15,987	2.61	–
Subgroup 49	(1,0) if Trade workers and related workers n.e.c.	5,019	0.82	–
Subgroup 50	(1,0) if Hotels, restaurants, bars and similar establishments and related workers	516	0.08	–
Subgroup 52	(1,0) if Butlers, housekeepers and related workers	154	0.03	–
Subgroup 53	(1,0) if Chefs and Cooks, waiters, barmen and related workers	12,156	1.99	–
Subgroup 54	(1,0) if Attendants (domestic and hotels) and agents (passengers transp.services)	2,446	0.4	–

Source RAIS-Migra 1996–2002

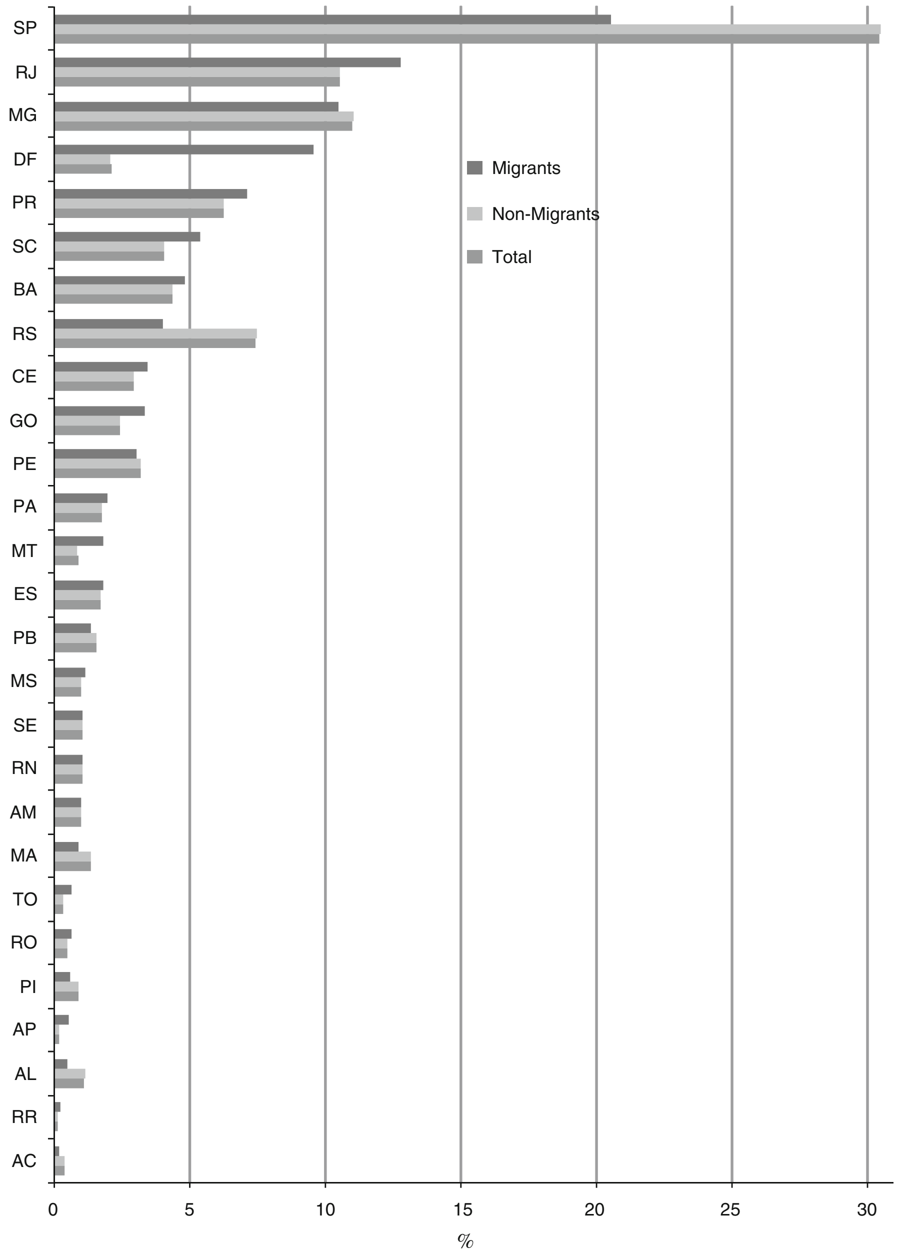

Fig. 4 Percentage of workers by state (1996–2002). *Note* Migrants refer to the percentage of total migrant workers by state; non-migrants refer to the percentage of total non-migrant workers by state; the total refers to the percentage of total workers (migrants plus non-migrants) by state. *Source* RAIS-Migra 1996–2002

Acknowledgments The authors record their thanks, without implication, to Geoffrey Hewings and to two anonymous referees for their valuable comments. Financial support from CAPES Foundation and from CNPq and data support from the Labor Ministry of Brazil are gratefully acknowledged.

References

Abowd JM, Kramarz F, Margolis DN (1999) High wage workers and high wage firms. Econometrica 67:251–333

Arbache JS, Carneiro FG (1999) Unions and interindustry wage differentials. World Dev 27:1875–1883

Azzoni C, Carmo H, Menezes T (2003) Comparações da Paridade do Poder de Compra entre cidades: aspectos metodológicos e aplicação ao caso brasileiro. Pesquisa e Planejamento Econômico 33:91–126

Azzoni C, Servo L (2002) Education, cost of living and regional wage inequality in Brazil. Pap Reg Sci 81:157–175

Bacha E, Taylor L (1978) Brazilian income distribution in the 60's: facts, model results and controversy. J Dev Stud 14:271–297

Barros RP, Mendonça R (1995) A evolução do bem-estar e da desigualdade no Brasil. Rev Brasil Econ 49:329–352

Carruth A, Collier W, Dickerson A (2004) Inter-industry wage differences and individual heterogeneity. Oxford Bull Econ Stat 66:811–846

Cowell FA, Ferreira FHG, Litchfield J (1996) Income distribution in Brazil 1981–1990: parametric and non-parametric approaches. Discussion paper # DARP 21, London School of Economics. http://ideas.repec.org/p/cep/stidar/21.html

Dabos M, Psacharopoulos G (1991) An analysis of the sources of earnings variation among Brazilian males. Econ Educ Rev 10:359–377

Ferreira P, Duarte A, Salvato M (2006) Disparidades Regionais ou Educacionais? Um exercício contrafactual. In: Teixeira EC, Barbosa MJ (eds) Investimento e Crescimento Econômico no Brasil. Editora da UFV, Viçosa

García I, Molina JA (2002) Inter-regional wage differentials in Spain. Appl Econ Lett 9:209–215

Gatica J, Mizala A, Romaguera P (1995) Interindustry wage differentials in Brazil. Econ Dev Cult Change 43:315–331

Gibbons R, Katz L (1992) Does unmeasured ability explain inter-industry wage differentials? Rev Econ Stud 59:515–535

Haisken-DeNew JP, Schmidt CM (1997) Inter-industry and inter-region differentials: mechanics and interpretation. Rev Econ Stat 79:516–521

Keane MP (1993) Individual heterogeneity and interindustry wage differentials. J Hum Resour 28:134–161

Krueger AB, Summers LH (1988) Efficiency wages and the inter-industry wage structure. Econometrica 56:259–293

Langoni CG (1973) Distribuição de renda e crescimento econômico no Brasil. Editora Expressão e Cultura, Rio de Janeiro

Lucas RE (1988) On the mechanics of economic development. J Monet Econ 22(1):3–42

Menezes-Filho N (2001) Educação e Desigualdade. In: Lisboa, Menezes-Filho (eds) Microeconomia e Sociedade. Contracapa, Rio de Janeiro

Mincer J (1974) Schooling, experience and earnings. National Bureau for Economic Research, New York

Motellón E, López-Bazo E, El-Attar M (2009) Regional heterogeneity in wage distributions: evidence from Spain. Research Institute of Applied Economics Working Paper 2009/03. http://www.ub.edu/irea/working_papers/2009/200903.pdf

Murphy KM, Topel RH (1987) Unemployment, risk, and earnings: testing for equalizing wage differences in the labor market. In: Lang K, Leonard J (eds) Unemployment and the structure of labor markets. Basil Blackwell, Oxford

Pereira J, Galego A (2007) Regional wage differentials: static and dynamic approaches. Cefage UE—Working paper 2007/07

Pinheiro AC, Ramos L (1994) Inter-industry wage differentials and earning inequality. Estud Econ 21: 79–111

Relatório Anual de Informações Sociais—Migração (RAIS-Migra)—1995–2002, Brasília, Labor Ministry of Brazil

Wooldridge JM (2002) Econometric analysis of cross-section and panel data. MIT, Massachusetts