NBER WORKING PAPER SERIES

CITIES AS ENGINES OF OPPORTUNITIES:
EVIDENCE FROM BRAZIL

Radu Barza
Edward L. Glaeser
César A. Hidalgo
Martina Viarengo

Working Paper 32426
http://www.nber.org/papers/w32426

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2024

We acknowledge the support of Cristian Jara-Figueroa in the initial conceptualization of the empirical strategy. Barza and Viarengo gratefully acknowledges the financial support received from the Swiss National Science Foundation (Principal Investigator: Prof. Dr. Martina Viarengo; Research Grant n. 100018-176454). Hidalgo acknowledges the support of the Agence Nationale de la Recherche grant number ANR-19-P3IA-0004, the 101086712-LearnData-HORIZON-WIDERA-2022-TALENTS-01 financed by European Research Executive Agency (REA) (https://cordis.europa.eu/project/id/101086712), IAST funding from the French National Research Agency (ANR) under grant ANR-17-EURE-0010 (Investissements d'Avenir program), and the European Lighthouse of AI for Sustainability [grant number 101120237-HOR-IZON-CL4-2022-HUMAN-02]. The usual caveats apply. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w32426

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Radu Barza, Edward L. Glaeser, César A. Hidalgo, and Martina Viarengo. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Cities as Engines of Opportunities: Evidence from Brazil
Radu Barza, Edward L. Glaeser, César A. Hidalgo, and Martina Viarengo
NBER Working Paper No. 32426
May 2024
JEL No. D63,I24,N90,N96,O10,O11,O18,O43,R10,R23

ABSTRACT

Are developing-world cities engines of opportunities for low-wage earners? In this study, we track a cohort of young low-income workers in Brazil for thirteen years to explore the contribution of factors such as industrial structure and skill segregation on upward income mobility. We find that cities in the south of Brazil are more effective engines of upward mobility than cities in the north and that these differences appear to be primarily related to the exposure of unskilled workers to skilled co-workers, which in turn reflects industry composition and complexity. Our results suggest that the positive effects of urbanization depend on the skilled and unskilled working together, a form of integration that is more prevalent in the cities of southern Brazil than in northern cities. This segregation, which can decline with specialization and the division of labor, may hinder the ability of Brazil's northern cities to offer more opportunities for escaping poverty.

Radu Barza
The Graduate Institute, Geneva
radu.barza@graduateinstitute.ch

Edward L. Glaeser
Department of Economics
315A Littauer Center
Harvard University
Cambridge, MA 02138
and NBER
eglaeser@harvard.edu

César A. Hidalgo
Center for Collective Learning,
Toulouse School of Economics &
Corvinus University of Budapest
cesar.hidalgo@tse-fr.eu

Martina Viarengo
Geneva Graduate Institute
Department of Economics
Chem. Eugène-Rigot
1202 Geneva
Switzerland
and CEPR, Center for International
Development at Harvard University
martina.viarengo@graduateinstitute.ch

Introduction

Do cities in the developing world enable upward mobility for the poor or do they become poverty traps? Wages are typically higher in larger agglomerations (Glaeser and Gottlieb, 2008, Duranton, Henderson and Strange, 2015, Combes and Gobillon, 2015) and also appear to rise more steeply in cities, which is compatible with Alfred Marshall's hypothesis that clustering enables learning (Glaeser and Mare, 2001, De La Roca and Puga, 2016).[1] Yet other studies indicate that cities can leave poor people behind, especially in low- and middle-income countries (Marx, et al., 2013; Kramarz and Viarengo, 2015; Bryan, et al., 2020; Manea, Piraino and Viarengo, 2023). In this paper, we study the link between city size and wage growth in Brazil and test the Marshall-inspired hypothesis that the wage growth of poor urbanites is limited when the skilled and unskilled largely work in different establishments.

We focus on Brazil, a middle-income country with enormous disparities between and within cities (Musacchio, Martínez-Fritscher and Viarengo, 2014; World Bank 2023). We use administrative individual income data from the Annual Social Information Report (RAIS), which covers Brazil's entire formal sector workforce. RAIS provides information on more than 900 million employer-employee observations, enabling us to monitor the income growth of individuals who entered the labor market in 2006 earning less than 1.5 times the minimum wage. We classify these individuals as poor. We have a panel data set spanning 13 years and consisting of 675,632 workers.

Methodologically, we follow the two-stage procedure of De La Roca and Puga (2016), who estimate the elasticity between population size and wages in Spain. We differ from their approach both by examining other agglomeration measures beyond city size and by analyzing a sample of individuals initially identified as poor. We start by calculating the initial premia of working in specific agglomerations, which we refer to as the city-level effect. We then calculate the benefits accumulated over a seven-year period in a specific location (seven years is the average time between relocations in our sample). This term combines the initial agglomeration premium, and the learning incurred while working at a specific location. We refer to this term as the medium-term effect.

[1] According to Chetty et al. (2016), cities in the United States are linked to higher wage growth for adults but lower upward mobility for children.

We estimate an elasticity of the initial premium to city size of .011 in the north and .016 in the south. Seven years after relocation, the elasticity of the medium-term effect relative to city population is .005 in the north and .037 in the south, which is a strikingly large difference. In the north, 52 percent of urban workers who earn less than 1.5 times the minimum wage in their first year in the labor market continue to earn less than 1.5 times the minimum wage after seven years. The corresponding percentage in the south is 32 percent. Brazil's southern cities are functioning as engines of opportunity, while the northern cities of Brazil seem to be comparatively stagnant, as described by Marx, Stoker, and Suri (2013).

To understand what drives these regional differences, we examine the role of industrial composition. The north has a significantly higher proportion of public sector employment than the south, which contributes to the greater segregation of skilled workers in the north. There is a real wage premium in the north for individuals employed in the public sector (as in Panizza, 2001), which may lead to an oversupply of workers in northern cities, as in the Harris and Todaro (1970) model. While cities in the south specialize in manufacturing and higher-skilled services, cities in the north focus on agricultural industries and government jobs. Wages increase with agglomeration size far more for private sector workers in the south and more for public sector workers in the north. We also find that average education levels, population size, and industrial diversity at the city level, do not significantly explain the disparity between north and south, but the level of economic complexity, as defined by Hidalgo and Hausmann (2009), does explain and predict the divide between north and south (see also Freeman and Viarengo, 2014).

Following the logic of Alfred Marshall, Glaeser (1999) predicts that segregation of skilled and unskilled workers will lead to less learning and wage growth by the unskilled. We test that hypothesis by measuring the likelihood with which a low-wage workers share establishments with high-wage earners. The north-south wage premium gap falls by 60 percent when we control the average isolation of the initially poor population at the occupation, industry, or establishment level.

Individuals with varying levels of human capital work together more often in the south than in the north. As such, more opportunities for on-the-job learning exist in the south, while in the north, highly educated individuals are concentrated in the public sector. These results suggest that workplace segregation may be as important as residential segregation which has been extensively studied (Kain, 1969; Ellwood, 1984; Cutler and Glaeser, 1997; Chetty, Hendren, et al., 2015; Chetty et al., 2020).

But if southern cities are dynamic engines of opportunity, then why do so many people go to northern cities? A natural explanation is that these cities first grew to exploit nearby natural resources and have been sustained by public employment. Their processing of these resources, however, is simple and does not involve interactions between more and less skilled workers. Moreover, economies of scale in some natural resource industries can be considerable (i.e., US mines) which implies that employers are large and tend to interface with large suppliers and service firms (Glaeser, Kerr, and Kerr, 2015). Consequently, there is limited opportunity for ground up entrepreneurship in any area beyond local services.

Another explanation is that the government may anchor the local economy with a large public sector, limiting opportunities for private sector employees. This phenomenon is particularly prevalent in Brazil where workers face barriers to move freely among the private and public sectors.

Our work contributes to the growing literature on urban mixing, which explores a city's ability to catalyze links among individuals with different demographic or economic characteristics (Toth et al. 2021; Nilforoshan et al. 2023; Dong et al. 2020; Lee, Peri and Viarengo, 2022). This literature has shown, for instance, that cities that are fragmented by physical barriers (e.g. highways, rivers) exhibit lower levels of urban mixing and higher levels of income inequality (Toth et al. 2021), and that larger cities exhibit higher levels of segregation (Nilforoshan et al. 2023). Our study contributes to this literature by showing that cities that are better at mixing low- and high-skill workers work better as engines of upward mobility.

We test these hypotheses by examining the relationship between city wages and wage growth on the one hand, and industrial structure, overall skills, skill segregation, and the public sector's share of the labor force on the other hand.

The remainder of this article is organized as follows: Section 2.2 describes the three datasets used, Section 2.3 the methodology, Section 2.4 presents and discusses the results, and Section 2.5 concludes.

Data Description

Our primary data set is the Annual Social Information Report – Relação Annual de Informações Sociais (RAIS).

RAIS is a panel dataset at individual level. It links employer and employee databases and includes the entire universe of formal workers in Brazil between 2003 and 2018. It is collected by the Ministry of Labor which surveys all public and private employers about their workers' wages and characteristics on yearly basis. Firms are legally required to answer the survey. RAIS contains a unique time-invariant identifying code for each company and individual, which enables tracking workers across time and space. RAIS contains the monthly wage of an individual, the wage in December, number of hours worked per week, the type and duration of job interruptions, gender, age, education, ethnicity, geographical location of the job, the size of the firm and of the establishment, nature of the work contract, the occupation of the individual and the industry of the firm. We perform the analysis at the individual-occupation-firm-year observation, level adjusting wages by inflation with 2010 as reference year.

The complete RAIS includes data on 5,560 municipalities, 2,500 occupations, and 585 industries for more than 30 million workers each year. We use two subsamples of the data. In the first sample we include all men between 25 and 44, as to ensure a sample with a high labor force participation. The second sample is restricted to the "initially poor," i.e., workers who earn less than 1.5 times the minimum hourly wage when they first enter the labor market in 2006. We then follow these initially poor individuals over time and locations. This results in a panel of 675,632 workers observed between 2006 and 2018. We do not include observations prior to 2006 in our sample. We analyzed data from 2003 to 2006 to confirm that individuals who entered the labor market in 2006 had not participated in the labor market between 2003 and 2006.

The geographical analysis focuses on the 558 microregions, which are defined by the Brazilian Institute of Geography and Statistics by grouping several nearby municipalities (Figure 1). Microregions range in size from less than 3,000 inhabitants in Fernando de Noronha to over 14 million people in Sao Paolo. Thirty microregions contain more than one million individuals.

For the replications of De La Roca and Puga (2017) we sampled five percent of the male population aged 24-44 in 2006. The sampling was conducted across five strata: age, education, location, industry, and occupation. We then track these individuals until 2018. This ensures a representative sample of non-poor males. We obtained panel data for 926,863 individuals over a 13-year period.

We further divide these two samples into the private and public sectors. The core analysis focuses on private sector workers since wages in the public sector are determined at the national level and do not reflect the competitive wages determined by local labor market conditions. We only use data from the public sector to compare wage structures between the private and public sectors in different regions.

Methodology

Combes, Duranton, and Gobillon (2008), Combes et al. (2010), Combes and Gobillon (2015), De la Roca and Puga (2017) suggest a two-step procedure to estimate the benefits of economic agglomerations. In the first stage we estimate the Mincerian wage equation with location fixed effects, as presented in Equation 1

$$\omega_{jrit} = \xi_r + \alpha_t + \alpha_i + \mu_j + \sum_{k=1}^{R} \gamma_k e_{jkt} + \boldsymbol{x}'_{jt}\beta + \varepsilon_{jrit} \quad (1)$$

Where ω_{jrit} represents the log of the monthly wage for individual *j*, in urban region *r*, industry *i*, at time *t*. $\boldsymbol{x}_{jt}$ is a vector of individual characteristics including age, age-squared, company tenure, hours and days worked per week, and education. ξ_r represents the region fixed-effect, α_t the time fixed-effect,.μ_j the individual fixed effect, α_i the industry fixed effects, and ε_{jrit} the error term. The term $\sum_{k=1}^{R} \gamma_k e_{jkt}$ captures the dynamic learning, where e_{jkt} reflects the experience acquired by working in location *r* ($k=\overline{1,R}$) until time *t*.

In the second stage we regress the location fixed effects from the first stage on population measures.[2] Formally, we estimate:

[2] Combes et al. (2008) show that the effect of the introduced heteroscedasticity through sampling error while using the estimated coefficients as a dependent variable are negligible, i.e., using feasible generalized least-square estimate in the second stage to correct for the heteroscedasticity introduced through sampling errors due to using as dependent variable an estimate from a first stage does not significantly change the estimated coefficients of the second stage. De la Roca and Puga (2017) argue that the variance of the estimated city effect is better estimated by a two-stage procedure than by a one stage estimate. It is also

$$\hat{\xi}_r = \gamma \log(population) + \upsilon_r \qquad (B2)$$

The coefficient γ is the agglomeration effect (Combes and Gobillon. 2015).

We report in the analysis two city premia: initial premium and the medium-term premium. Both the initial and the medium-term premia are the estimated γ coefficients in Equation 2, where the dependent variable is different. The initial premium is estimated by regressing the first stage estimated microregion fixed effects on population. This is also known as the static effect, or the level effect of agglomeration on wages. The medium-term premium is the γ estimated in a regression where the dependent variable equals the location fixed effects estimated in the first stage plus the dynamic benefits of acquiring experience in each location ($\sum_{k=1}^{R} \gamma_k e_{jkt}$) over an average period of 7.7 years. This dependent variable includes both the initial premiums received from working in a large urban area plus the annual learning benefits that accumulate over time. By construction, the difference between the medium and the initial premium reflects the wage returns to the learning while working. According to our data, individuals change locations every 7.1 years on average. To compare our results to De La Roca and Puga (2017), we calculate the medium term at 7.7 years. We estimate these effects for the two samples.

When estimating Equation 2 we are weighing by the local employment. There is an econometric case for choosing weighted regressions as the wage premium estimates for larger cities are more accurate. Additionally, there is a conceptual case because weighted regressions better capture the experience of the average Brazilian.

Results

The structure of the results section is as follows. We first replicate the analysis conducted by De La Roca and Puga (2017) using a representative sample of males aged 24-44 in the private sector. We then repeat the analysis using the initially poor workers in the private sector. Additionally, we analyze the geographical heterogeneity and provide separate findings for the northern and southern regions of Brazil. The results show that initially

known that the agglomeration effect is different for different industries. This would be the equivalent of adding a fixed effect for a two-way dummy at industry – region level. When regressing these fixed effects on population measures, the estimated γ coefficient is different for each industry. Balland et al. (2020) show that this γ is a function of the knowledge intensity of the industry, meaning that the agglomeration effect is stronger for more knowledge intense industries.

poor males in the private sector face limited opportunities for upward mobility in the northern region. To analyze the factors contributing to the formation of urban agglomerations in different territories, we will compare the private and public sectors in the northern and southern regions. We then examine urban characteristics such as education levels, economic complexity of the microregion, and the interaction between low-skilled and high-skilled workers.

City-Specific Learning Effects

We now turn to our core empirical exercise. We assess the average wage gain when working in large economic agglomerations both in the short run and in the medium term. To estimate these gains, we use the methodology proposed by De La Roca and Puga (2017). We first report estimates for an equivalent sample as the one used by De La Roca and Puga (2017), next we focus on the sample of initially poor male workers and then allow for heterogeneity across Brazil.

Table 2 presents the results. The first stage regressions (Equation 1) are shown in columns (1) and (4) for the broader male population and the initially poor sample. Columns (2) and (5) report coefficients of the initial premia, while columns (3) and (6) report the medium-term premia. These values are obtained from the second-stage regressions (Equation 2).

The estimated coefficient for the initial premium in the larger male sample is .020, with a medium-term premium of .030. In the initially poor male sample, the coefficients are .020 and .039, respectively. These results indicate that the wage elasticity to population size is 2.4 percent for the overall male population when the population doubles (Chang, 2016). The wage elasticity for poor males is two percent. The medium-term estimates show that agglomeration effects have a greater impact on the learning outcomes of initially poor males. This claim is supported by comparing the coefficients in column (6) and column (3), as well as by analyzing the difference between the initial premium and medium-term premium for each sample. The wage gain from learning in large urban agglomerations for individuals who were initially poor is 1.9 percent, as indicated by a difference of .019. In the larger male sample, this difference is 0.006.

When comparing this coefficient to the .022 coefficient estimated in Table 2 of De La Roca and Puga (2017), one can state that the initial agglomeration effects are similar in

Spain and in Brazil. The coefficient for Brazil's medium-term has a magnitude of .022, which is 1.7 times smaller than the equivalent estimate in De La Roca and Puga (2017), which has a point estimate of .051. The findings indicate that learning while working in Brazilian cities is lower compared to Spanish cities.[3]

The difference between the returns to the initially poor and the larger male sample may reflect a differential effect within the same cities. In the US, Autor (2020) reports little or no returns for less educated people living in bigger cities, but this is primarily due to the less educated living in less educated cities (Glaeser, 2020). Our results show that although the poor initially experience a smaller benefit from working in large urban agglomerations compared to the overall male workers, the learning is stronger for the initially poor individuals. To further investigate the factors behind the differences, we now explore regional heterogeneity within Brazil.

Geographical Heterogeneity

Brazil is formally divided into five large regions. Based on the geographical position and the similarity of GDP per capita, we combine the North and Northeast regions and define it as north. We label the rest of Brazil as south. These regions are at different levels of development and urbanization. Indeed, the south is more urbanized and much richer than the north (Figure 2). We estimate the overall effect of being in the north versus being in the south.

Table 3 shows the results for private sector workers. Panel A shows the results for the full sample of workers in the private sector, and Panel B for the initially poor sample of workers in the private sector. Column (1) to (3) show the initial wage premium, and column (4) to (6) the intermediate term wage premium. Column (1) and (4) reproduce exactly the results of Table 2.

Columns (2) and (5) show that being in the north is associated with earnings losses ranging from -.133 to -.208 log points. The initial earnings losses are of -.142 log points for non-poor and -.133 for the initially poor. The medium-term losses when working in

[3] By construction, the medium term incorporates the initial premia plus the learning generated by working in a specific location. Thus, comparing the difference between medium-term premium and initial premium reveals the returns to learning.

the north are of -.135 for non-poor and these losses rise to -.208 log points for the initially poor workers. These negative effects are large, but not unexpected, considering the acknowledged poverty in northern Brazil.

We aim to determine if urban agglomerations have similar positive effects in both regions. To achieve this, we interact the logarithm of the population with the north dummy. The results are reported in columns (3) and (6). The negative coefficients of the interaction terms show that the positive effects of urban agglomerations are lower in the northern region compared to the southern region. These effects are statistically significant, but their economic significance is negligible as it amounts to less than 1 REAL, except for the medium terms for initially poor individuals. Cities in the north have a lower wage elasticity in the medium term, with a difference of .032 log points. This result shows that urban agglomerations in the north generate lower levels of learning. Both the direct effects and the interactions indicate that northern cities are doing far less to generate wage growth for poorer Brazilians. If workers in northern cities experience lower long-term learning and wage growth, it raises the question of why urban agglomerations still form in the north. We analyze the factors that attract workers to local labor markets in these cities. We draw inspiration from the Harris and Todaro (1970) two-sector model and investigate whether the public sector in the northern region is an attractive option for individuals seeking employment in these cities. One notable distinction between cities in northern and southern Brazil is the higher percentage of public sector employment. In many northern cities, the public sector employs more than 25 percent of the workforce. Our previous focus was limited to private-sector employees. In the next subsection, we will analyze public-sector wages and compare them to the analysis of private-sector wages. To gain a comprehensive understanding of regional wage dynamics, we merge data from both sectors and report results on the pooled sample as well.

The Public Sector

Analyzing the role of the public sector in the heterogeneity of city premia represents a novel contribution to the literature, as the public sector is rarely considered in empirical studies in urban economics. Table 4 divides the sample into the north and the south and allows us to look at public and private sector workers. There are four panels (A-D) reflecting the first two divisions (north all, north initially poor, south all, south initially poor), and six regressions within each panel reflecting the third division and the initial vs. intermediate-term results.

Table 4 presents the results. The purpose of this analysis is to compare wage premia in the private and public sectors to understand whether individuals are attracted to northern cities by higher returns in the private or the public sectors. The results show an initial wage premium of 0.026 log points for the pooled two-sector sample for non-poor males. The premium is divided into .013 log points for private-sector workers and .070 for public-sector workers. The estimated coefficients for initially poor workers in the northern region are .020, .011, and .030, respectively. The findings show that employees in the public sector in large northern cities receive wage premia that are three to five times higher than those in the private sector. The patterns are consistent in the medium term. Specifically, for the non-poor sample, the medium-term wage premium in the pooled sample is .047 log points. This translates to a .026 log point increase for the private sector and a .124 log point increase for the public sector. This demonstrates that working in the public sector in the northern region results in nearly six times higher long-term returns. The average coefficient for initially poor workers is .020 in the pooled sample, .005 in the private sector, and .087 in the public sector. The results clearly show that the public sector primarily drives wage growth in the northern region. The findings support the hypothesis that individuals might be attracted to northern cities because of the higher returns offered by working in the public sector. This is particularly relevant for the initially poor individuals who receive insignificant long-term benefits in the private sector. However, individuals in the public sector can experience premia that are nearly 17 times higher. The results show that there is a higher return to city size in northern cities, but this is primarily driven by opportunities in the public sector. These findings support the Harris and Todaro model, which suggests that cities have a dual labor market, which can be either the formal or the informal market or the public sector and the private sector. Jobs are limited in the privileged sector (the public one in our case). If workers secure employment in the public sector in northern Brazil, they receive a higher wage compared to other sectors. Public sector results are primarily influenced by the consistent wage levels across different locations, with increases being based on seniority. While the cost of living is not the same across locations, wages are not adjusted in the private sector dependent on this.

Panels C and D of Table 4 show results for the southern cities. In this case, the premia for non-poor workers are .023 in the pooled sample, .021 in the private sector, and .006 in the public sector. The medium-term premia are .029, .026, and .011, respectively. The initial premium coefficient is smaller for the initially poor sample (.018, .016, and -.004, respectively), but the intermediate-term premium is larger (.042, .037, and .067, respectively). The coefficients for the public sector are smaller in the non-poor sample for both the initial and medium-term premia. For the initially poor individuals, the initial premium for the public sector is negative but very small in magnitude. The medium-term in the public sector is large, almost twice as large as the medium-term in the private sector. The results show that the private sector effectively promotes upward mobility in the short and long term. The public also provides long-term learning opportunities and offers greater medium-term benefits for individuals who are initially poor.

The overall returns to city size indicate that the returns to city size are slightly higher in the north for the entire sample. This is primarily due to the high returns to city size within the public sector of northern Brazil. For the initially poor sample, the initial wage premium coefficients are larger in the south than in the north for the private sector. However, there is a significant difference in the public sector premia, with negative initial premia in the south and large, positive premia in the north.

One possible interpretation for the higher coefficient in the south is that southern cities are more successful in facilitating upward mobility for individuals who start with low wages. If we subtract the initial wage premium from the intermediate-term wage premium, we can determine the coefficient of wage growth based on city size. The coefficient for the initially poor samples is .021 in the south and -.006 in the north. There is a consistent association of wage growth among initially poor individuals in southern cities, which does not appear in northern cities. We now investigate why northern Brazil is different?

The North-South Wage Premia Gap

We focus on two core hypotheses. The first hypothesis emphasizes skills and learning. Wage levels, and especially wage growth, should be a function of the level of skilled individuals and the frequency of their interactions with less skilled individuals. A significant body of research, starting with Rauch (1993) and including Moretti (2003), has extensively documented the strong correlation between the skill composition of a city and its economic vitality. Chauvin et al. (2016) documented that this is true in Brazil and other middle-income countries. They estimated that a 10 percent increase in the share of college graduates in an area results in a 10 percent increase in individual earnings.

The second hypothesis emphasizes industrial diversification and economic complexity. Yet formal education may matter little for the actual skills that determine earnings in Brazil's labor market. Consequently, we follow the literature on complexity and economic growth (Hidalgo and Hausmann, 2009; Hidalgo, 2021) and focus on the presence of rarer and more advanced industries as a measure of local human capital. Formally, we define the complexity of the economic activity in a given region (ECI_r) as:

$$ECI_r = \frac{K_r - mean(\vec{K})}{std(\vec{K})} \qquad (B3)$$

Where, *r* stands for region, and $(\vec{K})$ is the eigenvector associated with the eigenvalue of the matrix:

$$M_{r,r'} = \frac{1}{k_r} \sum_i \frac{M_{r,i}\, M_{r',i}}{k_i} \qquad (B4)$$

$k_r = \sum_i M_{r,i}$ which is the number of the industries in which the region has a Revealed Comparative Advantage $(RCA_{r,i} > 1)$, and $k_i = \sum_r M_{r,i}$ is the number of regions with a comparative advantage in a specific industry *i*. A region is said to have a Revealed Comparative Advantage (Balassa and Noland, 1989) when $RCA_{r,i} > 1$, where:

$$RCA_{r,i} = \frac{\frac{x_{r,i}}{\sum_{i'} x_{r,i'}}}{\frac{\sum_{r'} x_{r',i}}{\sum_{i'} \sum_{r'} x_{r',i'}}} \qquad (B5)$$

For region *r*, and industry *i*, i' refers to all industries except industry *i (*and the same is true about *r'*). $x_{r,i}$ is the number of firms in region *r* that operate in industry *i*.

The learning-in-cities model predicts that the skill level in an area should predict earnings growth over the life cycle, because unskilled individuals learn from skilled individuals. Yet for learning to occur, the unskilled must *de facto* interact with the skilled.

Consequently, we also test the hypothesis that economic segregation between skilled and unskilled predicts wages, wage growth, and explains part of the north-south diverge.

We measure segregation with an exposure measure defined as:

$$Exposure_c = \sum_i \frac{NonPoor_{ic}}{Poor_{ic}+NonPoor_{ic}} \frac{Poor_{ic}}{Poor_c} - \frac{NonPoor_c}{Poor_c+NonPoor_c} \quad (6)$$

The values of $Poor_{ic}$ and $NonPoor_{ic}$ refer to the number of poor and non-poor in city c and in group *i*, where group *i* can refer to an occupation, company, or industry. The values of $Poor_c$ and $NonPoor_c$ refer to the number of poor and non-poor in the city as a whole.

This measure represents the likelihood that a low-income worker will encounter a high-income worker within the same occupation in each microregion. This measure captures the level of poverty segregation, which is similar to the isolation measures commonly used in studies on residential segregation. Our exposure measures are based on data from 2003-2005, i.e., prior to estimating wage effects. The measures are higher in the north compared to the south.

The North-South Initial Wage Premium Gap

We test both hypotheses about the factors that contribute to learning and prosperity in cities. To do so, we alter Equation 2 and introduce a dummy variable that equals 1 if the agglomeration is located in the north. We also include microregion-specific measures such as the share of college graduates, complexity, or exposure. We thus estimate:

$$CFE_c = \alpha * North_c + \gamma * Log(Pop_c) + \beta * C_c + \varepsilon_c \quad (7)$$

where CFE_c is the estimated intermediate or initial wage impact of microregion c, $North_c$ is an indicator that takes on a value of 1 if the microregion is in the north, and $Log(Pop_c)$ is the natural logarithm of the population in microregion c. CX_c is one of the microregion-specific measure: the share of college-graduates, the complexity measure, or the isolation index. The terms α, β, and γ are parameters to be estimated and ε_c refers to the error term. Our focus is to determine if the negative effects of being located in the north are reduced when we control for the microregion-specific characteristics.[4] Table 5 report

[4] We do not include an added interaction between the north indicator and log of population because that would make it harder to interpret the coefficient on the other interaction with city population. We are interested in changes of the point estimates of the north dummy.

results on the initial wage premium. Table 6 repeats the analysis for the intermediate term premium.

Table 5 shows results for the initial premium in the private sector. Panel A reports estimates for the entire sample of male workers, and Panel B for the initially poor males.

We compare these regressions with results reported in column (2) of Table 3. In Table 3, the second regression in Panel A shows a coefficient on location in the north of -.142. This static estimate of the north dummy also tells us the average wage gain of an individual that relocates from north to south. Panel B shows a coefficient of -.133, which suggests a comparable effect for initially poor workers.

In Table 5, we add in each column one of our three microregion-specific measures– (1) share of college graduates, (2) economic complexity, and (3) the exposure. In all specifications we control for the log of city population. The estimated coefficient for the share of college graduates is negative in both samples. This finding is robust across specifications and different from the findings of Chauvin et al. (2017). In specifications where we interact the share of college graduates with the microregion population, we observe strong positive interactions. Consequently, there seems to be a correlation between a high percentage of college graduates in less populated micro-regions and lower wages. This could be attributed to the prevalence of highly educated individuals employed in the public sector within these microregions. Results in column (1) does little to reduce the coefficient on the north variable.

In column (2) we control for the complexity measure. Results show that complexity is positively associated with the initial wage premium. In this case, the variable is significant both economically and statistically, and causes the coefficient on city population to flip signs. The estimated coefficient of .051 shows that a one standard deviation increase in the complexity measure is associated with a .051 log point increase in the initial wage premium.

The effect is almost identical in magnitude for the initially poor sample of workers. More complex cities appear to deliver significantly higher earnings for their workers and complexity seems to explain, in a purely statistical sense, why city size is associated with larger initial wage premia. The flip in the sign of the estimated coefficient for population indicates that larger cities are not necessarily more economically complex. This also shows that economic complexity explains wages, and not the size of the population alone.

The greater level of complexity in the south also explains a significant part of the north-south wage premium gap.[5] The coefficient on location in the north shrinks in magnitude from -.142 in column (2) Table 3 to -.106 in column (2) Table 5, or a 24 percent reduction in the north parameter. In panel B, the coefficient on being in the north declines by 26 percent. These results indicate that complexity has greater impacts on the initially poor individuals.

In column (3) we introduce the measure of exposure of the less skilled to the highly skilled in companies across the microregion. This measure is larger and more statistically significant than the complexity measure. The associated coefficient is .069 for all males in the private sector and .072 for the initially poor males. Controlling for exposure reduces the "north effect" to -.048 for non-poor male workers and -.047 for initially poor males. This represents a 66 percent reduction in the negative effect of the north on the initial premiums for non-poor males and a 65 percent reduction for initially poor individuals. The significant wage gap between the north and south can be partially explained by the lack of integration in working environments in northern microregions, as integration is strongly associated with higher earnings. We continue the analysis with the intermediate wage premium.

The North-South Intermediate Wage Premium Gap

Table 6 shows results for the intermediate wage premium. These results are compared to column (5) of Table 3. As before, Panel A reports results for the non-poor sample, and Panel B for the initially poor males. The estimated coefficient in column (5) of Table 3 is -.135 for the non-poor sample and -.208 for the initially poor sample.

In column (1) of Table 6, we show results when for the share of collage-graduates. As for the initial-premium, the coefficient remains negative, but its magnitude is three times higher for non-poor males. For the initially poor sample, the intermediate term coefficient on education is -.064. This coefficient is almost twice smaller than the equivalent estimate for the initial premium. This shows that the coefficient on college share would be positive

[5] Results survive to controlling for income per capita at microregion level. See results reported in Table 7. We also have strong correlations between complexity index and income per capita, and GDP per capita.

if we regress the change between the medium-term and the initial-term premia on the share of college graduates. This positive effect on the growth rate of wages for the initially poor sample is compatible with the learning in cities model.

The estimated effects of complexity also remain large. The estimated coefficient has a magnitude of .057 for the non-poor sample, and .055 for the initially poor. Controlling for complexity reduces the estimated coefficient on location in the north by 33 percent for the non-poor sample and by 17 percent for the initially poor sample.

The results for exposure are reported in the column (3) of Table 6. For the non-poor sample, the coefficient on exposure is .069. Controlling for this variable reduces the estimated coefficient on the northern variable by 72 percent. For the initially poor sample, the exposure coefficient has a magnitude of .076, and the north variable's magnitude declines by 46 percent. The intermediate term impact of the exposure variable is slightly larger than the immediate impact for the initially poor sample, which provides support for the view that this variable captures opportunities for learning in cities for the initially poor.

The ability of economic integration to explain differences in upward mobility across cities is shown by Figure 3. In the first panel of the figure, the initial wage premium is regressed on city population. The slope is positive, but the wage premia for the southern cities are clearly greater than the wage premia in the northern cities holding population constant. In the second panel, we control for the exposure of the less skilled to the more skilled. As the figure shows, the gap between north and south is almost completely "explained" (in a purely statistical sense) by the greater workplace segregation of the north.

The results show that the ability of low skilled people to partner with more skilled people creates most of the opportunity for higher earnings, and that the ability of less skilled people to learn from more skilled people generates wage gains as well over time.

Taken together, these results suggest that northern Brazil is different because its firms involve less integration of skilled and unskilled workers. Both factors - exposure and complexity - correlated strongly with earnings.

Why are Brazil's northern cities simpler and more economically segregated?

In this section, we present our hypothesis on why complexity is more prevalent in the south and segregation is more prevalent in the north, and why these two factors are highly correlated at the city level. This hypothesis consists of two elements. Complex products drive complex organizations, leading to more exposure in larger and more complicated organizations. The second element is that simple products only lead to the development of large cities when those cities also serve significant political functions. According to this hypothesis, northern cities initially developed around basic goods and services, and their growth today is attributed to their public sectors. The cities in the south were founded on intricate goods and still export them today. Industrial and political history explains why southern cities seem to provide more opportunities than northern cities.

We present five facts about complexity, integration, and the public sector in Brazil: (1) average firm size correlates with complexity and manufacturing employment share, (2) average firm size is related to increased interaction between high and low skilled workers, (3) microregions with a larger public sector have smaller firms, less integration, and lower complexity, (4) skilled workers are more likely to work in the public sector in both north and south, but the segregation of skilled workers in the public sector is more pronounced in the north, and (5) the public sector employs a larger share of workers in the north than in the south.

Figure 4 shows the correlation between average establishment size and the share of manufacturing. This fact can be interpreted as confirming the presence of scale economies in manufacturing. Manufacturing firms may create more complex products, which could explain the difference. Figure 5 confirms a link between establishment size and complexity across micro-regions.

This fact that larger firms tend to produce more complex products is expected. More complex products typically involve a wider variety of tasks. Keeping some tasks in-house may be more cost-effective than outsourcing them due to high transaction costs. In some cases, simple products with significant economies of scale are produced within large firms, but there is a general correlation between complexity and the size of establishments across Brazilian microregions.

Figure 6 shows the correlation between average firm size and exposure to low and high skilled workers across metropolitan areas. When we weigh by employment the correlation is strong. Larger companies in urban areas often blend the expertise of both highly skilled and less skilled employees. The interactions within these companies can help less skilled workers improve their skills over time.

Figures 7, 8 and 9 show the connections between public sector size, industrial complexity, establishment size, and the integration of skilled and unskilled workers. Figure 7 shows the strong link between the size of the public sector and industrial complexity. This fact suggests that a larger public sector corresponds to less private economic activity, where simpler tasks are being performed by firms. Figure 8 confirms that simpler tasks are associated with smaller firms. Figure 9 indicates that smaller, simpler firms have lower skill integration.

We do not intend to suggest that these correlations indicate a causal impact of the public sector. A more plausible interpretation is that these figures reflect the impact of a weaker private sector. Historically, Brazilian cities that experience economic decline often maintain political influence, distributing benefits and drawing in new residents. Unfortunately, cities resulting from this politically driven process often lack the ability to promote upward mobility or economic productivity.

Our last findings are simply data points, not statistics. In the northern region, 34% of public sector workers (men aged 25-54) have a university education, while only 6% of private sector workers are similarly educated. In the northern region, the sector known as "public administration, defense, and social security" employs 15.9% of all formal sector workers, with university-educated individuals equally divided between the public and private sectors.

In the southern region, 8.6% of workers are employed in "public administration, defense, and social security." In the northern region, 43% of public sector workers hold a university degree compared to 12% in the private sector. Skilled workers in the south are over twice as likely to work in the private sector than in the public sector. Highly skilled workers are more concentrated in the public sector in the south, but more dispersed among private firms in the same region.

Table 7 investigates why northern cities may be different. Columns (1) and (2) show that companies are typically bigger in big cities, but smaller in the north. Columns (3) and (4) show that workers are less likely to work in single person establishments in larger cities but are 5.7 percentage points more likely to be in such an isolated establishment in the south. Columns (5) and (6) look at the share of companies that have fewer than five employees. This share also declines in big cities but expands dramatically in the north. Together these columns support the view that big cities typically involve big firms that enable mixing between skilled and unskilled workers, but such firms are dramatically absent from northern Brazil.

Conclusion

We began this paper by asking whether Brazilian cities provided economic opportunities for initially poor workers, similar to those found for cities in Spain or the US. We found that the overall effects of city size of wages and wage growth was smaller than in the studies on Spain or the US, and that there is a sharp divergence between Brazil's northern and southern cities. The cities of the north do not have high wage gains in the private sector. The cities in the north have high elasticities of returns with respect to population size.

We then tested whether the differences between north and south could be related to the level of skill and the integration of skilled and unskilled workers in firms. We found that both economic complexity, which we interpreted as a measure of informal skills or knowledge, and exposure of unskilled to skilled co-workers, strongly predict wages and more weakly predicted wage growth. Our integration measure is particularly effective in explaining the differences between north and south, yet complexity and mixing are not completely independent, as on-the-job mixing and learning opportunities should vary among economic sectors. Interestingly, formal measures of human capital, such as share of college graduates, could not explain these wage effects, showing that measures of complexity and urban mixing capture information about on the job learning that transcends formal measures of education.

This north-south dynamic reflects a mix of economic structure and institutions that can be explained using the story of three cities: Salvador da Bahia, Rio de Janeiro, and Sao Paulo.

Salvador was the first capital of colonial Brazil. It was founded in 1549 with an economy centered on sugar exports. Historically, the comparative advantage of Northern Brazil, like that of Southern U.S. and the Caribbean, lay in its tropical climate, which enabled the growth of crops, like sugar, which were hard to produce in Europe. Slaves worked the sugar plantations, whose owners were more like medieval European nobles than the industrial entrepreneurs of the late 19th century.

Neither the core export of the city nor its public role as the colonial capital were likely to turn Salvador da Bahia into a model of upward mobility. Moreover, as sugar's role in the Brazilian economy declined in the 18th and 19th centuries, Salvador remained a regional (although not a national) capital with particularly strong ties to the Portuguese crown. Slavery continued in Brazil until 1888, with whites representing a minority in Salvador. The city's port continues to be important, but Salvador's missed the industrial revolution that came to the more entrepreneurial cities of the South.

Rio de Janeiro replaced Salvador as the capital of the Brazilian colony in 1763. The shift to the south partially reflected the mineral wealth found in nearby Minas Gerais. Rio became a far more important capital than Salvador had ever been. Portugal moved its imperial capital to the city in 1808 and Rio then became Brazil's capital following its independence in 1822. Rio was largely a political, even imperial, city in the 19th century and its growth partially reflected the direction of rents to the capital by its leaders as in Ades and Glaeser (1995). For example, the imperial government repeatedly offered public subsidies to build railroads in Rio.

Although it was a political city just like Salvador, 19th century Rio was so dominant in Brazil that it attracted many of the talented and entrepreneurial Brazilians who industrialized the city. For example, the entrepreneurs who led the Companhia Progresso do Brazil built a textile factory that would be the heart of the Bangu neighborhood for decades. As the country transitioned from monarchy to democracy, its leaders increased their focus on the well-being of the poor, especially those who lived nearby. Oswaldo Cruz, with public support, led numerous campaigns to improve the health of Rio's urban poor, one of which led to the famous Vaccine Revolt.

While Rio had more complex products than Salvador, Sao Paulo is a far more extreme example of a city built overwhelmingly on commerce instead of politics. In the 19th century, Sao Paulo emerged as the center of Brazil's coffee industry. That industry attracted a particularly entrepreneurial group of immigrants who turned Sao Paulo into the industrial hub of the country. Perhaps, the absence of the imperial court made Sao Paulo more attractive for people with few natural ties to the old-world aristocracy. In the 20^{th} century, Sao Paulo's business leaders campaigned successfully to move the country's capital away from Rio, which would level the playing field between the two cities.

By 1960, Sao Paulo's population surpassed that of Rio and by 1980, its population had surpassed that of New York City. Agricultural products, such as soybeans, coffee and raw sugar are among Sao Paulo's core exports, but these are sold by complex agribusinesses that are on the technological frontier, not simple plantations. Moreover, the city also exports complex products such as pharmaceuticals, turbines, and air pumps. Its rubber footwear exports (Havaianas flipflops) are something of a global icon. This economy is large and complex.

These three cities have suggested a hierarchy within Brazil, from the northernmost Salvador, which was built around politics and a simple export good, to Rio de Janeiro, a political juggernaut but also an industrial capital, to Sao Paulo, a city built almost entirely around its economic functions that still exports globally. The Sao Paulo model naturally leads to large firms which require highly skilled people at the top, a logistic network in the middle and less skilled workers at the bottom. This economic implies a certain level of exposure between more and less skilled people at the firm level, which is just not implied by either the public sector or by simpler service businesses.

We have discussed three cities, but we see Rio de Janeiro as essentially a convex combination of two different models of urban form: the city based on manufactured exports and the city based on political power and natural resources. According to this interpretation, cities that reflect political power and natural resources will tend to have far simpler companies and a great segregation of the skilled into the public sector. Cities with commercial sectors will tend to have larger and more integrated firms.

The northern cities were built around a simple colonial economy, and they have remained large partially because of their political functions. The southern cities have more dynamic economies and larger firms. The northern model of cities, which fits many cities in the low-income world, leads to much less exposure of poor to rich and much less upward mobility. The southern model, which resembles cities in high-income countries, seems more conducive to both economic productivity and upward mobility.

Our results are based on data on one country. The results show that the dynamic benefits of urbanization hinge upon the skilled and unskilled working together. In the cities of southern Brazil, workplace integration prevails. In the poor cities of northern Brazil, the poor and rich interact less. This segregation, which may decline with industrial deepening, seems to prevent Brazil's northern cities from providing more pathways out of poverty.

Further work is needed to bring evidence on the role of urban agglomerations in other low- and middle- income countries. This would further the understanding of policies that can transform urbanization in engines of economic growth. Future work can also establish whether urban agglomerations in low- and middle countries can provide upward mobility for their workers.

Acknowledgements

We acknowledge the support of Cristian Jara-Figueroa in the initial conceptualization of the empirical strategy. Barza and Viarengo gratefully acknowledges the financial support received from the Swiss National Science Foundation (Principal Investigator: Prof. Dr. Martina Viarengo; Research Grant n. 100018-176454). Hidalgo acknowledges the support of the Agence Nationale de la Recherche grant number ANR-19-P3IA-0004, the 101086712-LearnData-HORIZON-WIDERA-2022-TALENTS-01 financed by European Research Executive Agency (REA) (https://cordis.europa.eu/project/id/101086712), IAST funding from the French National Research Agency (ANR) under grant ANR-17-EURE-0010 (Investissements d'Avenir program), and the European Lighthouse of AI for Sustainability [grant number 101120237-HOR-IZON-CL4-2022-HUMAN-02]. The usual caveats apply.

Bibliography

Ades, Alberto F., and Edward Glaeser. 1995. "Trade and circuses: explaining urban giants." *The Quarterly Journal of Economics* 110 (1): 195-227.

Athey, Susan, Billy Ferguson, Matthew Gentzkow, and Tobias Schmidt. 2021. "Estimating Experienced Racial Segregation in US Cities Using Large-Scale GPS Data." *Proceedings of the National Academy of Sciences* 118 (46): e2026160118.

Autor, David. 2020. "The Faltering Escalator of Urban Opportunity." *Securing Our Economic Future*, 108–36.

Balassa, Bela, and Marcus Noland. 1989. "'Revealed' Comparative Advantage in Japan and the United States." *Journal of International Economic Integration*, 8–22.

Balland, Pierre Alexandre, Cristian Jara-Figueroa, Sergio G Petralia, Mathieu P A Steijn, David L Rigby, and César A Hidalgo. 2020. "Complex Economic Activities Concentrate in Large Cities." *Nature Human Behaviour*, 1–7.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh." *Econometrica* 82 (5): 1671–1748.

Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. 2017. "What Is Different about Urbanization in Rich and Poor Countries? Cities in Brazil, China, India and the United States." *Journal of Urban Economics* 98: 17–49.

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2020. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." *The Quarterly Journal of Economics* 135 (2): 711–83.

Chetty, Raj, Nathaniel Hendren, and others. 2015. "The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimates." *Harvard University and NBER* 133 (3): 1–145.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics* 63 (2): 723–42.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. "Estimating Agglomeration Economies with History, Geology, and Worker Effects."

In *Agglomeration Economics*, 15–66. University of Chicago Press.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. “The Empirics of Agglomeration Economies.” In *Handbook of Regional and Urban Economics*, 5:247–348. Elsevier.

Correia, Sergio. 2016. "A feasible estimator for linear models with multi-way fixed effects". Preprint at http://scorreia. com/research/hdfe.pdf.

Cutler, David M, and Edward L Glaeser. 1997. “Are Ghettos Good or Bad?” *The Quarterly Journal of Economics* 112 (3): 827–72.

De La Roca, Jorge, and Diego Puga. 2017. “Learning by Working in Big Cities.” *Review of Economic Studies* 84 (1): 106–42.

Dong, Xiaowen, Alfredo Morales, Eaman Jahani, et al.2020. “Segregated interactions in urban and online space.” EPJ Data Science, vol. 9, no 1, p. 20.

Duranton, Gilles, Vernon Henderson, and William Strange, eds. 2015. “Handbook of regional and urban economics.” Elsevier, 2015.

Ellwood, David T. 1984. “The Hope for Self-Support.” *The State and the Poor in the 1980s*, 19.

Freeman, Richard B. and Martina Viarengo. 2014. “School and Family Effects on Educational Outcomes across Countries,” *Economic Policy*, (lead article), Vol. 79, No. 29, July, pp. 395-446.

Glaeser, Edward L. 1999. “Learning in Cities.” *Journal of Urban Economics* 46 (2): 254–77.

———. 2020. “Infrastructure and Urban Form.”

Glaeser, Edward L, and Joshua D Gottlieb. 2008. “The Economics of Place-Making Policies.” *Brookings Papers on Economic Activity*, no. 14373.

Glaeser, Edward L, Sari Pekkala Kerr, and William R Kerr. 2015. “Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines.” *Review of Economics and Statistics* 97 (2): 498–520.

Glaeser, Edward L, and David C Mare. 2001. “Cities and Skills.” *Journal of Labor Economics* 19 (2): 316–42.

Gollin, Douglas, Martina Kirchberger, and David Lagakos. 2021. “Do Urban Wage Premia Reflect Lower Amenities? Evidence from Africa.” *Journal of Urban Economics* 121: 103301.

Harris, John R, and Michael P Todaro. 1970. “Migration, Unemployment and Development: A Two-Sector Analysis.” *The American Economic Review* 60 (1): 126–42.

Hidalgo, César A. 2021. "Economic complexity theory and applications." *Nature Reviews Physics*, vol. 3, no 2, p. 92-113.

Hidalgo, César A., and Ricardo Hausmann. 2009. “The Building Blocks of Economic Complexity.” *Proceedings of the National Academy of Sciences* 106 (26): 10570–75.

Kain, John F. 1969. “Coping with Ghetto Unemployment.” *Journal of the American Institute of Planners* 35 (2): 80–83.

Kramarz, Francis and Martina Viarengo. 2015. “Ni en Emploi Ni en Formation - Des Jeunes Laissés pour Compte,” Paris: Sciences-Po Press.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E Waugh. 2018. “The Welfare Effects of Encouraging Rural-Urban Migration.” *National Bureau of Economic Research*, no. w24193.

Lee, Taehoon, Giovanni Peri and Martina Viarengo. 2022. “The Gender Aspect of Immigrants’ Assimilation in Europe,” *Labour Economics*, Vol. 78, pp. 102180.

Manea, Roxana, Patrizio Piraino and Martina Viarengo. 2023. “Crime, Inequality and Subsidized Housing: Evidence from South Africa,” *World Development*, Vol. 168, pp. 106243.

Marshall, Alfred. 1890. "Principles of Economics, 8th edn (1920)." London, *Mcmillan*.

Marx, Benjamin, Thomas Stoker, and Tavneet Suri. 2013. “The Economics of Slums in the Developing World.” *Journal of Economic Perspectives* 27 (4): 187–210.

Miller, Mark M, Lay James Gibson, and N Gene Wright. 1991. “Location Quotient: A Basic Tool for Economic Development Analysis.” *Economic Development Review* 9 (2): 65.

Musacchio, Aldo, André Martínez-Fritscher and Martina Viarengo. 2014. "Colonial Institutions, Trade Shocks, and the Diffusion of Elementary Education in Brazil, 1889-1930," *Journal of Economic History*, Vol. 74, No. 3, pp. 730-766.

Nilforoshan, Hamed, Wenli Looi, Emma Pierson, et al. 2023. "Human mobility networks reveal increased segregation in large cities." *Nature*, vol. 624, no 7992: 586-592.

Panizza, Ugo. 2001. "Public Sector Wages and Bureaucratic Quality: Evidence from Latin America." *Economía Journal*, The Latin American and Caribbean Economic Association - LACEA, vol. 0 (Fall 2001), pages 97-152, August.

Perlman, Janice E. 2008. "Urban Poverty and Politics in Rio de Janeiro." *Democratic Brazil Revisited*. p. 257.

Tóth, Gergő, Johannes Wachs, Riccardo Di Clemente, et al. 2021. "Inequality is rising where social network segregation interacts with urban topology." *Nature communications*, vol. 12, no 1, p. 1143.

World Bank 2023. "The Brazil of the Future - Towards Productivity, Inclusion, and Sustainability," Washington D.C.: World Bank Publication.

Figures

Figure 1. Microregions

A. São Paulo B. Rio de Janeiro C. Belo Horizonte

Note: Blue lines represent the state frontiers. Red lines represent the microregion frontiers.

Figure 2. Microregion Heterogeneity

A. 2010 GDP per Capita in REAL B. 2010 Population Size

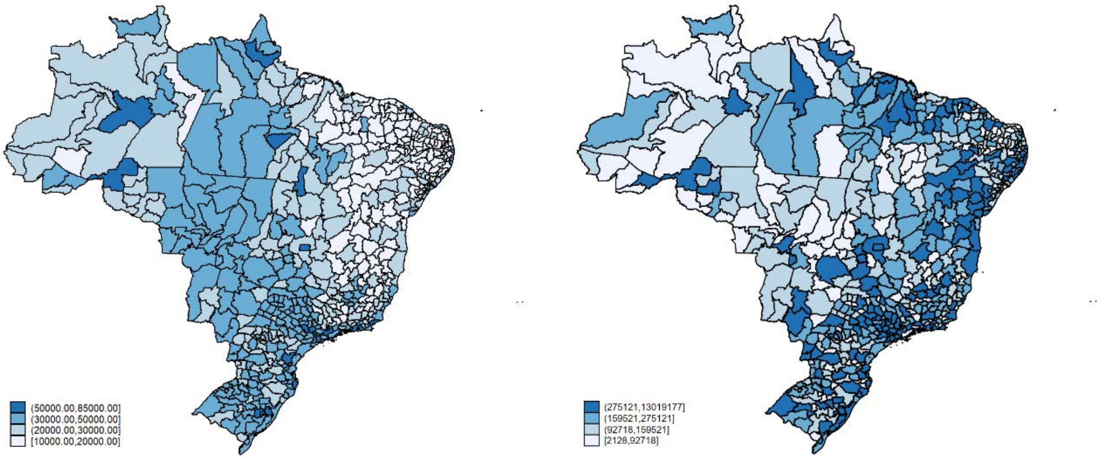
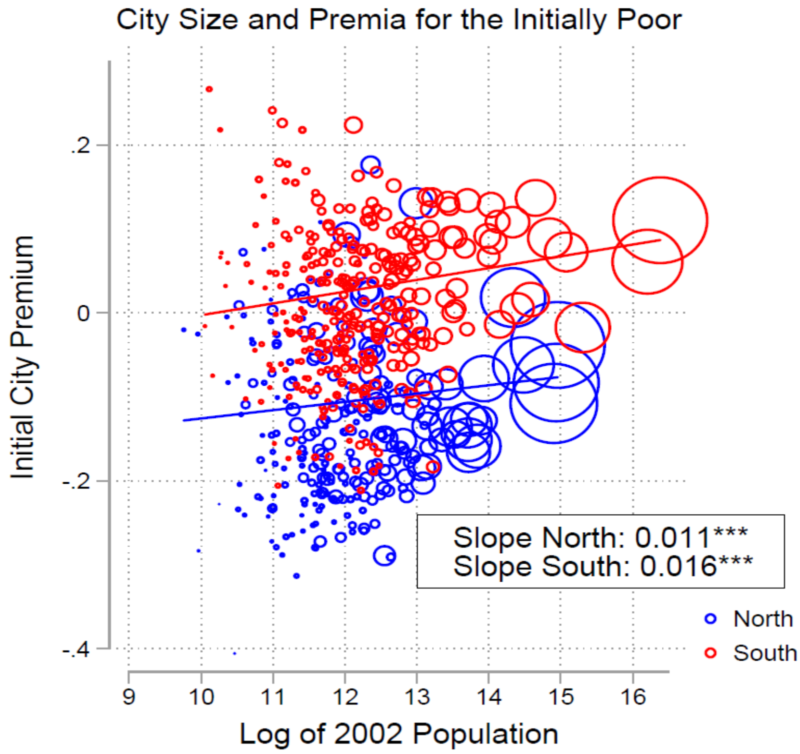
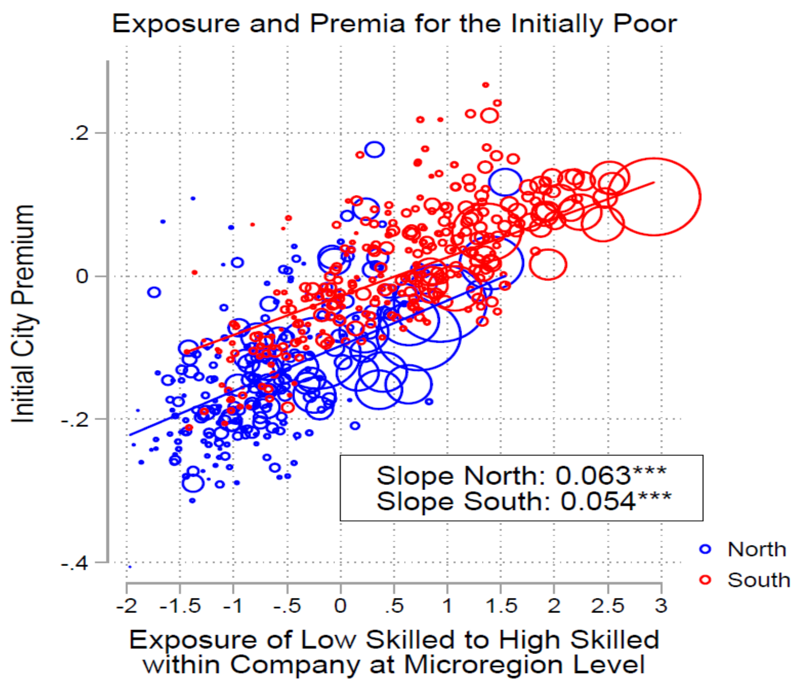


Figure 4. Wage Premia and City Size

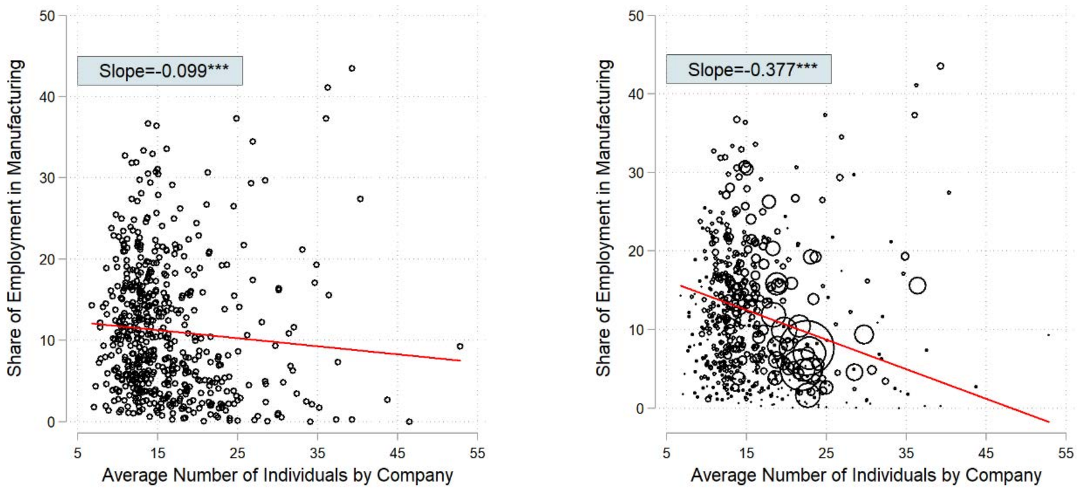




Note: Calculations for the exposure measure and the average number of individuals by company are done by the authors using the entire RAIS dataset, including the entire formally employed labor force available in the data.

Figure 4. Correlation between Average Establishment Size and the Share of Manufacturing.

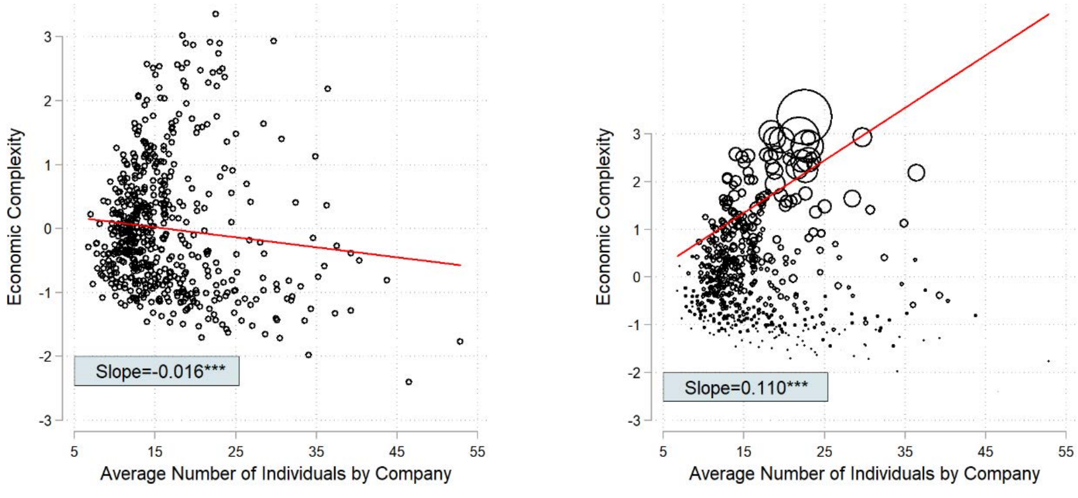
A. Unweighted Correlations **B. Weighting for the Local Employment**



Note: Calculations for the shares and the average number of individuals by company are done by the authors using the entire RAIS dataset, including the entire formally employed labor force available in the data.

Figure 5. Establishment Size and Complexity

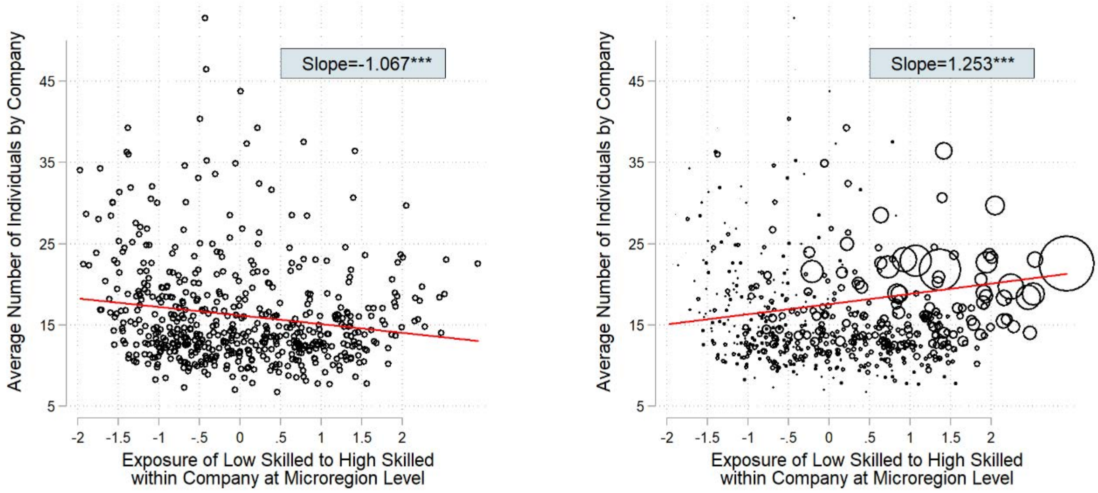
A. Unweighted Correlations **B. Weighting for the Local Employment**



Note: Calculations for the economic complexity and the average number of individuals by company are done by the authors using the entire RAIS dataset, including the entire formally employed labor force available in the data.

Figure 6. Correlation across Metropolitan Areas between Average Firm Size and Exposure between Low and High Skilled Workers

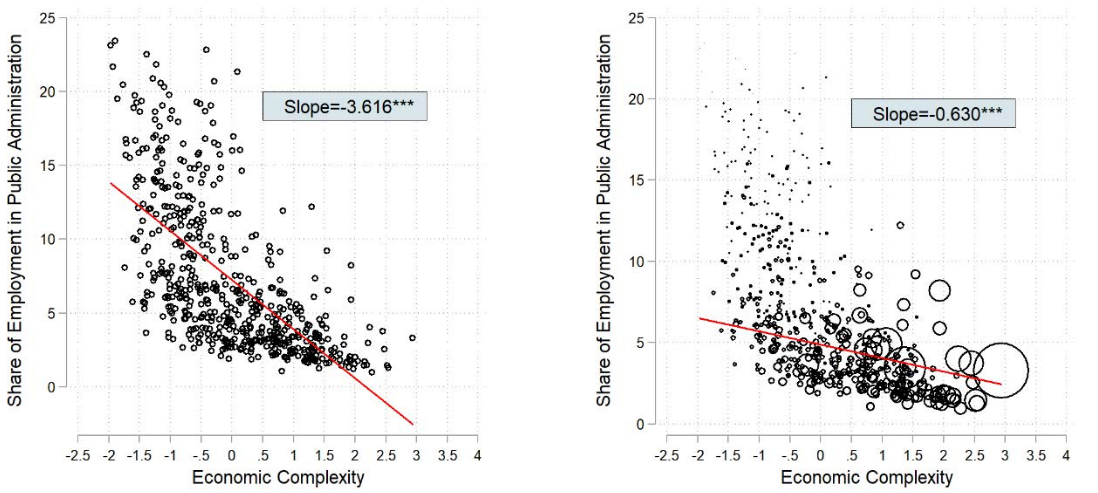
A. Unweighted Correlations **B. Weighting for the Local Employment**



Note: Calculations for the exposure measure and the average number of individuals by company are done by the authors using the entire RAIS dataset, including the entire formally employed labor force available in the data.

Figure 7. Correlations between the Size of the Public Sector and Industrial Complexity

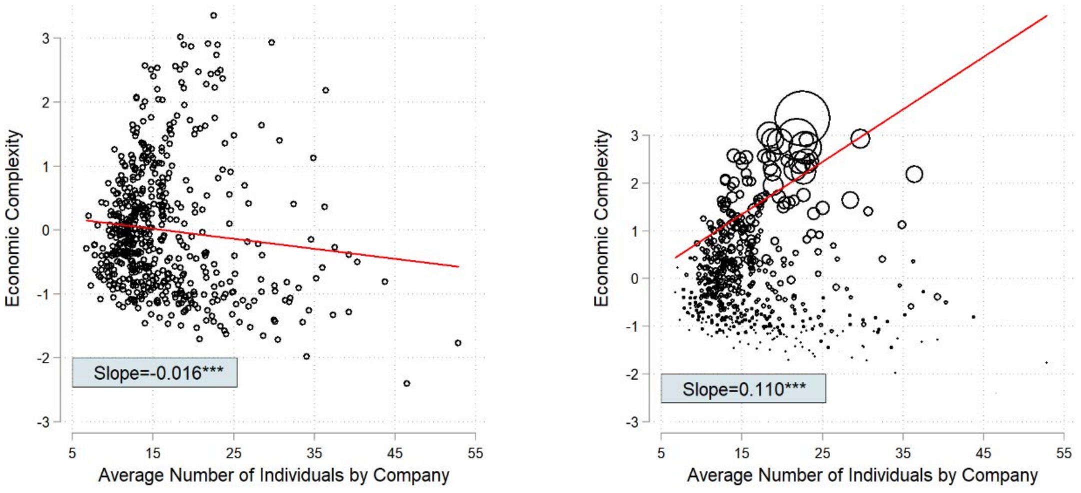
A. Unweighted Correlations **B. Weighting for the Local Employment**



Note: Calculations for the shares and the economic complexity index are done by the authors using the entire RAIS dataset, including the entire formally employed labor force available in the data.

Figure 8. Correlations between the Size of the Public Sector and Establishment Size

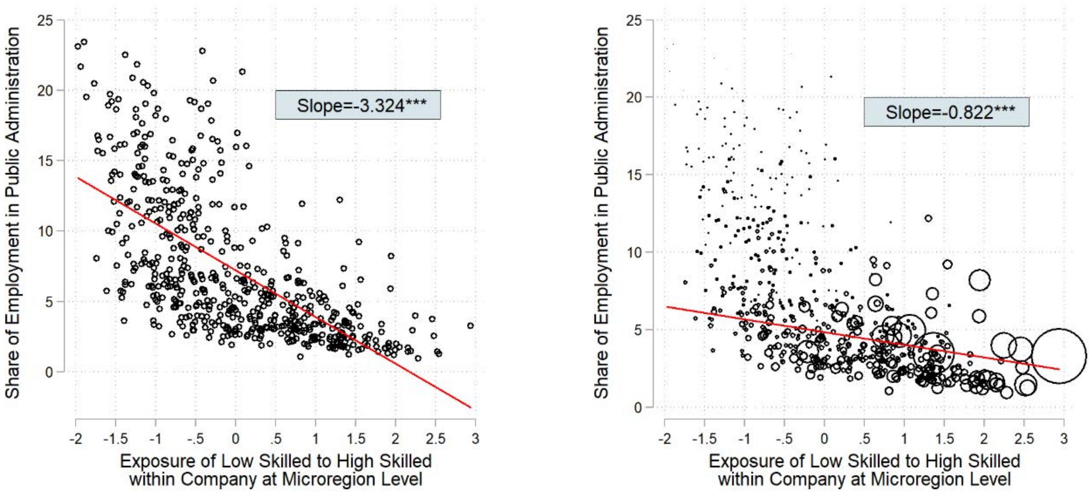
A. Unweighted Correlations **B. Weighting for the Local Employment**



Note: Calculations for the economic complexity index and the number of individuals by company are done by the authors using the entire RAIS dataset, including the entire formally employed labor force available in the data.

Figure 9. Correlations between the Size of the Public Sector Exposure between Low and High Skilled Workers

A. Unweighted Correlations **B. Weighting for the Local Employment**



Note: Calculations for the shares and the exposure measure are done by the authors using the entire RAIS dataset, including the entire formally employed labor force available in the data.

Tables

Table 1. Summary Statistics

	RAIS - Poor		RAIS - Males Aged 25 - 44	
	North	**South**	**North**	**South**
	(1)	(2)	(3)	(4)
Log of Mean Wages in 2010 REAL				
Under 25	6.35	6.54		
	(0.38)	(0.44)		
25 - 34	6.70	6.93	6.78	7.07
	(0.50)	(0.55)	(0.60)	(0.65)
35 - 44	6.84	6.99	6.95	7.26
	(0.46)	(0.52)	(0.71)	(0.76)
45 - 54			7.04	7.33
			(0.77)	(0.81)
All Age Groups	6.61	6.77	6.85	7.16
	(0.49)	(0.54)	(0.66)	(0.71)
Age	27.36	26.04	33.70	33.80
	(26.04)	(4.78)	(5.82)	(5.91)
Share Informality				
Share of Individuals with Earnings	0.93	0.94	0.96	0.98
Higher than the Min. Wage	(0.26)	(0.23)	(0.20)	(0.15)
Share of Individuals with Earnings	0.32	0.51	0.45	0.65
Higher than 1.5 x Min. Wage	(0.47)	(0.50)	(0.50)	(0.48)
Share of College Graduates	0.06	0.09	0.12	0.16
	(0.25)	(0.28)	(0.32)	(0.37)

Notes: This table reports the means of the main variables. Standard deviations are reported in parenthesis. RAIS includes data from 2006 to 2018. The initial poor sample includes all male individuals that joined the official labor market in 2006 at a wage lower than 1.5 times the minimum age. We follow these individuals over time until 2018 or until they leave the official labor market. Therefore, no individual in this sample is over 44 y.o.

Table 2. Replications de la Roca and Puga – Males in the Private Sector

	(1)	(2)	(3)	(4)	(5)	(6)
	RAIS Data, All Males in the Private Sector			**RAIS Data, Initial Poor Males in the Private Sector**		
	Log of Average Monthly Wage in 2010 Real	Initial Premium (city indicator coefficient column (1))	Medium - Term Premium (initial + 7.7 years local experience)	Log of Average Monthly Wage in 2010 Real	Initial Premium (city indicator coefficient column (1))	Medium - Term Premium (initial + 7.7 years local experience)
Log City Size		0.024*	0.030***		0.020***	0.039***
		(0.00001)	(0.00001)		(0.00002)	(0.00002)
City Indicator	YES			YES		
Company Tenure in Years	0.025***			0.013***		
	(0.0001)			(0.0002)		
Company Tenure Squared	-0.001***			0.0004***		
	(0.00001)			(0.000001)		
Experience in the labor market (years)						
The square of the experience in the labor market	-0.001***			-0.0049		
	(0.00001)			(0.00261)		
Experience first to second largest urban agglomerations	0.005***			0.016***		
	(0.0008)			(0.0010)		
Experience 3rd to 4th largest urban agglomerations	0.013***			0.019***		
	(0.0009)			(0.0011)		
Experience 5th to 7th largest urban agglomerations	0.016***			-0.003*		
	(0.0011)			(0.0012)		
…						
Experience first to second largest urban agglomerations x now in top 104	-0.001***			-0.001***		
	(0.0001)			(0.00009)		
Experience first to second largest urban agglomerations x experience	0.0005			0.0004		
	(0.0008)			(0.00093)		
Experience first to second largest urban agglomerations x experience x now top 104	-0.000			0.0002*		
	(0.0001)			(0.00009)		
Experience 3rd to 4th largest urban agglomerations x experience	-0.001***			-0.0012***		
	(0.0001)			(0.00009)		
Experience 3rd to 4th largest urban agglomerations x now in top 104	-0.002*			0.0001		
	(0.0009)			(0.00107)		
Experience 3rd to 4th largest urban agglomerations x experience x now top 104	0.000			0.0002		
	(0.0001)			(0.00010)		
Experience 5th to 7th largest urban agglomerations x experience	-0.001***			0.0004***		
	(0.0001)			(0.00011)		
Experience 5th to 7th largest urban agglomerations x now in top 104	-0.012***			-0.0077***		
	(0.0011)			(0.00121)		
Experience 5th to 7th largest urban agglomerations x experience x now top 104	0.001***			0.0005***		
	(0.0001)			(0.00011)		
…						
Experience outside top 104 urban agglomerations	-0.008***			-0.007***		
	(0.0004)			(0.00056)		
Experience outside top 104 urban agglomerations x now in top 104	0.008***			0.007***		
	(0.0004)			(0.00056)		
Experience outside top 104 urban agglomerations x experience x now top 104	-0.0001***			-0.0004***		
	(0.00001)			(0.00004)		
Unskilled Occupations Omitted						
Very high-skilled occupations	0.200***			0.482***		
	(0.0077)			(0.0213)		
High-skilled occupations	0.166***			0.347***		
	(0.0018)			(0.0028)		

Medium-high-skilled occupations	0.191***			0.246***		
	(0.0022)			(0.0033)		
Medium-low-skilled occupations	0.146***			0.189***		
	(0.0011)			(0.0015)		
Constant	6.598***	-0.339***	-0.400***	6.217***	-0.27***	-0.519***
	(0.0042)	(0.00021)	(0.00021)	(0.1079)	(0.00029)	(0.00033)
Observations	13,657,866	13,657,866	13,657,866	7,922,794	7,922,794	7,922,794
R-squared	0.809	0115	0.234	0.663	0.100	0.206
Adjusted R-squared	0.784	0.115	0.234	0.629	0. 100	0.206
Ethnicity FE	-	-	-	-	-	-
Literacy Level FE	-	-	-	-	-	-
Year FE	✓	-	-	✓	-	-
Economic Sector FE	✓	-	-	✓	-	-
Individual Fixed Effects	✓	-	-	✓	-	-

*Note: Data used here comes from RAIS. We use a 5 percent sample of the entire population on five strata: age, education, location, industry, and occupation. The sample is restricted to the legal working age population, i.e., 25 – 44 y.o. All results are obtained with linear regression models using fixed effects for the specified variables. All specifications include a constant term. We use indicator variables for cities with population over 350.000 people. We employ the Correia (2016) proposed methodology for the higher levels of fixed effects. Poor Sample includes all individuals that have joined the labor market in 2006 for a wage under 1.5 x the minimum wage. Standard errors are in parenthesis. We calculate them using a two-way clustering procedure at individual and microregion level. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Table 3. Estimates of the South - North City Premia Gap

Panel A: All Males in the Private Sector

	Initial Premium			**Medium - Term Premium (initial + 7.7 years local experience)**		
	(1)	(2)	(3)	(4)	(5)	(6)
Log of Population in 2002	0.024***		0.021***	0.030***		0.026***
	(0.00001)		(0.00001)	(0.00001)		(0.00001)
Is North		-0.142***	-0.120***		-0.135***	-0.123***
		(0.00005)	(0.00012)		(0.00006)	(0.00012)
North x Population			-0.008***			0.001***
			(0.00005)			(0.00005)
Constant	-0.339***	0.031***	-0.258***	-0.400***	0.050***	-0.312***
	(0.00021)	(0.00002)	(0.00018)	(0.00021)	(0.00002)	(0.00018)
Observations	13,657,86	13,657,866	13,657,866	13,657,866	13,657,866	13,657,866
R-squared	0.155	0.391	0.493	0.234	0.341	0.511
Adjusted R-squared	0.155	0.391	0.493	0.234	0.341	0.511

Panel B: Initial Poor Males in the Private Sector

	Initial Premium			**Medium - Term Premium (initial + 7.7 years local experience)**		
	(1)	(2)	(3)	(4)	(5)	(6)
Log of Population in 2002	0.020***		0.016***	0.039***		0.037***
	(0.00002)		(0.00002)	(0.00002)		(0.00002)
Is North		-0.133***	-0.118***		-0.208***	-0.143***
		(0.00006)	(0.00012)		(0.00007)	(0.00012)
North x Population			-0.005***			-0.032***
			(0.00005)			(0.00005)
Constant	-0.277***	0.038***	-0.184***	-0.519***	0.086***	-0.434***
	(0.00029)	(0.00003)	(0.00027)	(0.00033)	(0.00004)	(0.00028)
Observations	7,922,794	7,922,794	7,922,794	7,922,794	7,922,794	7,922,794
R-squared	0.100	0.406	0.463	0.206	0.524	0.674
Adjusted R-squared	0.100	0.406	0.463	0.206	0.524	0.674

*Notes: Please refer to notes of Table 2 for detailed notes on the estimates and variables used. These estimates represent the second stage regression in de la Roca and Puga methodology. Standard errors are in parentheses, clustered at microregion level. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Table 4. Learning while Working in Large Cities Across Sectors

Panel A: North - All Males

	Dependent Variable: City Premia					
	(1)	(2)	(3)	(4)	(5)	(6)
	Private		Public		All	
	Initial	Medium	Initial	Medium	Initial	Medium
Log of Population in 2002	0.013***	0.026***	0.070***	0.124***	0.026***	0.047***
	(0.00005)	(0.00005)	(0.00017)	(0.00017)	(0.00005)	(0.00005)
Constant	-0.284***	-0.443***	-1.003***	-1.650***	-0.458***	-0.705***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Observations	3,009,781	3,009,781	724,447	724,447	3,753,442	3,753,442
R-squared	0.028	0.108	0.165	0.395	0.086	0.236
Adjusted R-squared	0.028	0.108	0.165	0.395	0.086	0.236

Panel B: North - Initial Poor Males

	Dependent Variable: City Premia					
	(1)	(2)	(3)	(4)	(5)	(6)
	Private		Public		All	
	Initial	Medium	Initial	Medium	Initial	Medium
Log of Population in 2002	0.011***	0.005***	0.030***	0.087***	0.020***	0.020***
	(0.00004)	(0.00005)	(0.00033)	(0.00034)	(0.00005)	(0.00005)
Constant	-0.238***	-0.187***	-0.497***	-1.210***	-0.372***	-0.390***
	(0.001)	(0.001)	(0.005)	(0.005)	(0.001)	(0.001)
Observations	2,263,156	2,263,156	279,466	279,466	2,553,588	2,553,588
R-squared	0.026	0.004	0.025	0.164	0.072	0.054
Adjusted R-squared	0.026	0.004	0.025	0.164	0.072	0.054

Panel C: South - All Males

	Dependent Variable: City Premia					
	(1)	(2)	(3)	(4)	(5)	(6)
	Private		Public		All	
	Initial	Medium	Initial	Medium	Initial	Medium
Log of Population in 2002	0.021***	0.026***	0.006***	0.011***	0.023***	0.029***
	(0.00001)	(0.00001)	(0.00007)	(0.00007)	(0.00001)	(0.00001)
Constant	-0.258***	-0.312***	-0.057***	-0.065***	-0.284***	-0.347***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Observations	10,648,085	10,648,085	1,448,005	1,448,005	12,139,318	12,139,318
R-squared	0.230	0.325	0.007	0.017	0.249	0.341
Adjusted R-squared	0.230	0.325	0.007	0.017	0.249	0.341

Panel D: South - Initial Poor Males

	Dependent Variable: City Premia

	(1)	(2)	(3)	(4)	(5)	(6)
	Private		Public		All	
	Initial	Medium	Initial	Medium	Initial	Medium
Log of Population in 2002	0.016***	0.037***	-0.004***	0.067***	0.018***	0.042***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-0.184***	-0.434***	0.129***	-0.698***	-0.209***	-0.500***
	(0.00002)	(0.00002)	(0.00017)	(0.00022)	(0.00002)	(0.00002)
Observations	5,659,638	5,659,638	374,139	374,139	6,055,391	6,055,391
R-squared	0.124	0.431	0.002	0.110	0.142	0.446
Adjusted R-squared	0.124	0.431	0.002	0.110	0.142	0.446

*Notes: Initial city premia and medium city premia are calculated using the city fixed effects estimated in regression equivalent to the regressions reported in Table 2 for each of the specific sample. Please refer to notes of Table 2 for detailed information on the first stage regressions. Standard errors are in parentheses. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Table 5. Why is Northern Brazil Different - Gap in Initial Premia

Panel A: All Males in the Private Sector

	Initial Premium		
	(1)	(2)	(3)
Is North (= 1 if it is North)	-0.134***	-0.106***	-0.048***
	(0.00005)	(0.00006)	(0.00005)
Log of Population in 2002	0.025***	-0.015***	-0.008***
	(0.00002)	(0.00002)	(0.00001)
Share of College Graduates	-0.217***		
	(0.00065)		
Complexity Measure		0.051***	
		(0.00004)	
Exposure of Low Skilled to High Skilled			0.069***
within Company at Microregion Level			(0.00002)
Constant	-0.285***	0.137***	0.044***
	(0.00024)	(0.00027)	(0.00017)
Observations	13,657,866	13,657,866	13,657,866
R-squared	0.495	0.562	0.760
Adjusted R-squared	0.495	0.562	0.760

Panel B: Initial Poor Males in the Private Sector

	Initial Premium		
	(1)	(2)	(3)
Is North (= 1 if it is North)	-0.127***	-0.098***	-0.047***
	(0.00006)	(0.00006)	(0.00005)
Log of Population in 2002	0.018***	-0.026***	-0.014***
	(0.00003)	(0.00003)	(0.00002)
Share of College Graduates	-0.122***		
	(0.00086)		
Complexity Measure		0.058***	
		(0.00005)	
Exposure of Low Skilled to High Skilled			0.072***
within Company at Microregion Level			(0.00002)
Constant	-0.193***	0.287***	0.139***
	(0.00030)	(0.00039)	(0.00022)
Observations	7,922,794	7,922,794	7,922,794
R-squared	0.463	0.554	0.772
Adjusted R-squared	0.463	0.554	0.772

*Notes: Initial city premia and medium city premia are obtained using the city fixed effects estimated in regression equivalent to the regressions reported in Table 2 for each of the specific sample. Please refer to notes of Table 2 for detailed information on the first stage regressions. Standard errors are in parentheses. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Table 6. Why is Northern Brazil Different - Gap in Medium Term Premia

Panel A: All Males in the Private Sector

	Medium Term Premia		
	(1)	(2)	(3)
Is North (= 1 if it is North)	-0.123***	-0.092***	-0.038***
	(0.00005)	(0.00006)	(0.00005)
Log of Population in 2002	0.031***	-0.013***	-0.002***
	(0.00002)	(0.00002)	(0.00001)
Share of College Graduates	-0.217***		
	(0.00064)		
Complexity Measure		0.057***	
		(0.00004)	
Exposure of Low Skilled to High Skilled within Company at Microregion Level			0.069***
			(0.00002)
Constant	-0.354***	0.116***	-0.027***
	(0.00024)	(0.00026)	(0.00016)
Observations	13,657,866	13,657,866	13,657,866
R-squared	0.515	0.599	0.770
Adjusted R-squared	0.515	0.599	0.770

Panel B: Initial Poor Males in the Private Sector

	Medium Term Premia		
	(1)	(2)	(3)
Is North (= 1 if it is North)	-0.194***	-0.167***	-0.111***
	(0.00007)	(0.00007)	(0.00006)
Log of Population in 2002	0.033***	-0.007***	0.002***
	(0.00003)	(0.00004)	(0.00002)
Share of College Graduates	-0.064***		
	(0.00095)		
Complexity Measure		0.055***	
		(0.00005)	
Exposure of Low Skilled to High Skilled within Company at Microregion Level			0.076***
			(0.00003)
Constant	-0.369***	0.080***	-0.034***
	(0.00032)	(0.00043)	(0.00024)
Observations	7,922,794	7,922,794	7,922,794
R-squared	0.656	0.700	0.834
Adjusted R-squared	0.656	0.700	0.834

*Notes: Initial city premia and medium city premia are obtained using the city fixed effects estimated in regression equivalent to the regressions reported in Table 5 for each of the specific sample. Please refer to notes of Table 2 for detailed information on the first stage regressions. Standard errors are in parentheses. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Table 7. Company Size and the North Dummy

	Number of Individuals by Company		Share of One-Employee Companies to More-than-one Employee Companies at Microregion		Share of Small (< 5 Employees) Companies to Large Companies at Microregion	
	(1)	(2)	(3)	(4)	(5)	(6)
Log of Population in 2002	4.440***		-0.052***		-0.335***	
	(0.041)		(0.001)		(0.009)	
Is North (North = 1)		-2.927***		0.057***		0.410***
		(0.156)		(0.003)		(0.019)
Constant	-44.418***	17.888***	1.053***	0.392***	5.949***	1.693***
	(0.569)	(0.069)	(0.018)	(0.002)	(0.115)	(0.013)
Observations	25,209,826	25,209,826	8,928	8,928	8,923	8,923
R-squared	0.029	0.029	0.673	0.643	0.533	0.494
Adjusted R-squared	0.029	0.029	0.672	0.642	0.532	0.493
Year FE	✓	✓	✓	✓	✓	✓

*Notes: Standard errors are in parentheses, clustered at microregion level. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Appendix – Replications without weights

Table A.1. Replications de la Roca and Puga – Males in the Private Sector

	(1)	(2)	(3)	(4)	(5)	(6)
	RAIS Data, All Males in the Private Sector			**RAIS Data, Initial Poor Males in the Private Sector**		
	Log of Average Monthly Wage in 2010 Real	Initial Premium (city indicator coefficient column (1))	Medium - Term Premium (initial + 7.7 years local experience)	Log of Average Monthly Wage in 2010 Real	Initial Premium (city indicator coefficient column (1))	Medium - Term Premium (initial + 7.7 years local experience)
Log City Size		0.014*	0.022***		0.015**	0.020***
		(0.005)	(0.005)		(0.005)	(0.006)
City Indicator	YES			YES		
Company Tenure in Years	0.025***			0.013***		
	(0.0001)			(0.0002)		
Company Tenure Squared	-0.001***			0.0004***		
	(0.00001)			(0.000001)		
Experience in the labor market (years)						
The square of the experience in the labor market	-0.001***			-0.0049		
	(0.00001)			(0.00261)		
Experience first to second largest urban agglomerations	0.005***			0.016***		
	(0.0008)			(0.0010)		
Experience 3rd to 4th largest urban agglomerations	0.013***			0.019***		
	(0.0009)			(0.0011)		
Experience 5th to 7th largest urban agglomerations	0.016***			-0.003*		
	(0.0011)			(0.0012)		
…						
Experience first to second largest urban agglomerations x now in top 104	-0.0004***			-0.0008***		
	(0.0001)			(0.00009)		
Experience first to second largest urban agglomerations x experience	0.0005			0.0004		
	(0.0008)			(0.00093)		
Experience first to second largest urban agglomerations x experience x now in top 104	-0.000			0.0002*		
	(0.0001)			(0.00009)		
Experience 3rd to 4th largest urban agglomerations x experience	-0.001***			-0.0012***		
	(0.0001)			(0.00009)		
Experience 3rd to 4th largest urban agglomerations x now in top 104	-0.002*			0.0001		
	(0.0009)			(0.00107)		
Experience 3rd to 4th largest urban agglomerations x experience x now in top 104	0.000			0.0002		
	(0.0001)			(0.00010)		
Experience 5th to 7th largest urban agglomerations x experience	-0.001***			0.0004***		
	(0.0001)			(0.00011)		
Experience 5th to 7th largest urban agglomerations x now in top 104	-0.012***			-0.0077***		

	(1)	(2)	(3)	(4)	(5)	(6)
	(0.0011)			(0.00121)		
Experience 5th to 7th largest urban agglomerations x experience x now in top 104	0.001***			0.0005***		
	(0.0001)			(0.00011)		
…						
Experience outside top 104 urban agglomerations	-0.008***			-0.007***		
	(0.0004)			(0.00056)		
Experience outside top 104 urban agglomerations x now in top 104	0.008***			0.007***		
	(0.0004)			(0.00056)		
Experience outside top 104 urban agglomerations x experience x now in top 104	- 0.0001***			- 0.0004***		
	(0.00001)			(0.00004)		
Unskilled Occupations Omitted						
Very high-skilled occupations	0.200***			0.482***		
	(0.0077)			(0.0213)		
High-skilled occupations	0.166***			0.347***		
	(0.0018)			(0.0028)		
Medium-high-skilled occupations	0.191***			0.246***		
	(0.0022)			(0.0033)		
Medium-low-skilled occupations	0.146***			0.189***		
	(0.0011)			(0.0015)		
Constant	6.598***	-0.251***	-0.345***	6.217***	-0.238***	-0.299***
	(0.0042)	(0.065)	(0.065)	(0.1079)	(0.062)	(0.067)
Observations	13,657,866	558	558	7,922,794	558	558
R-squared	0.809	0.012	0.029	0.663	0.015	0.024
Adjusted R-squared	0.784	0.010	0.027	0.629	0.013	0.022
Ethnicity FE	-	-	-	-	-	-
Literacy Level FE	-	-	-	-	-	-
Year FE	✓	-	-	✓	-	-
Economic Sector FE	✓	-	-	✓	-	-
Individual Fixed Effects	✓	-	-	✓	-	-

*Note: Data used here comes from RAIS. We use a 5% sample of the entire population on four strata: gender, education, location, and age. We follow them over time until 2018 or until they leave the labor market if this happens prior to 2018. The sample is restricted to the legal working age population, i.e., 25 – 45 y.o. All results are obtained with linear regression models using fixed effects for the specified variables. All specifications include a constant term. We employ the Correia (2016) proposed methodology for the higher levels of fixed effects. Poor Sample includes all individuals that have joined the labor market in 2006 for a wage under 1.5 x the minimum wage. Standard errors are in parenthesis. We calculate them using a two-way clustering procedure at individual and microregion level. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Table A.2. Estimates of the South - North City Premia Gap

Panel A: All Males in the Private Sector

	Initial Premium			**Medium - Term Premium (initial + 7.7 years local experience)**		
	(1)	(2)	(3)	(4)	(5)	(6)
Log of Population in 2002	0.014*		0.012	0.022***		0.019**

	(1)	(2)	(3)	(4)	(5)	(6)
	(0.005)		(0.006)	(0.005)		(0.006)
Is North		-0.137***	-0.136***		-0.137***	-0.134***
		(0.008)	(0.008)		(0.009)	(0.009)
North x Population			-0.013			-0.011
			(0.009)			(0.009)
Constant	-0.251***	-0.024***	-0.164*	-0.345***	-0.019**	-0.252***
	(0.065)	(0.006)	(0.073)	(0.065)	(0.006)	(0.074)
Observations	558	558	558	558	558	558
R-squared	0.012	0.321	0.325	0.029	0.312	0.326
Adjusted R-squared	0.010	0.319	0.322	0.027	0.311	0.322

Panel B: Initial Poor Males in the Private Sector

	Initial Premium			**Medium - Term Premium (initial + 7.7 years local experience)**		
	(1)	(2)	(3)	(4)	(5)	(6)
Log of Population in 2002	0.015**		0.008	0.020***		0.026***
	(0.005)		(0.006)	(0.006)		(0.006)
Is North		-0.142***	-0.141***		-0.159***	-0.157***
		(0.008)	(0.008)		(0.008)	(0.008)
North x Population			-0.003			-0.032***
			(0.008)			(0.009)
Constant	-0.238***	0.008	-0.094	-0.299***	0.017**	-0.294***
	(0.062)	(0.005)	(0.068)	(0.067)	(0.005)	(0.070)
Observations	558	558	558	558	558	558
R-squared	0.015	0.374	0.377	0.024	0.407	0.429
Adjusted R-squared	0.013	0.373	0.374	0.022	0.406	0.426

*Notes: Please refer to notes of Table 5 for detailed notes on the estimates and variables used. These estimates represent the second stage regression in de la Roca and Puga methodology. Standard errors are in parentheses, clustered at microregion level. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Table A.3. Learning while Working in Large Cities Across Sectors

Panel A: North - All Males

	Dependent Variable: City Premia					
	(1)	(2)	(3)	(4)	(5)	(6)
	Private		Public		All	
	Initial	Medium	Initial	Medium	Initial	Medium
Log of Population in 2002	-0.001	0.008	0.052***	0.083***	0.021*	0.033***
	(0.008)	(0.008)	(0.014)	(0.014)	(0.008)	(0.009)
Constant	-0.147	-0.251**	-0.848***	-1.199***	-0.451***	-0.584***
	(0.092)	(0.093)	(0.169)	(0.173)	(0.101)	(0.102)
Observations	252	252	251	251	252	252
R-squared	0.0001	0.004	0.052	0.116	0.024	0.056
Adjusted R-squared	-0.004	0.0002	0.048	0.113	0.021	0.052

Panel B: North - Initial Poor Males

	Dependent Variable: City Premia					
	(1)	(2)	(3)	(4)	(5)	(6)

	Private		Public		All	
	Initial	Medium	Initial	Medium	Initial	Medium
Log of Population in 2002	0.005	-0.007	-0.009	-0.006	0.018*	0.006
	(0.006)	(0.007)	(0.018)	(0.019)	(0.007)	(0.007)
Constant	-0.195*	-0.061	-0.042	-0.084	-0.372***	-0.242**
	(0.077)	(0.078)	(0.220)	(0.229)	(0.086)	(0.088)
Observations	252	252	251	251	252	252
R-squared	0.003	0.004	0.001	0.000	0.023	0.003
Adjusted R-squared	-0.001	0.0003	-0.003	-0.004	0.019	-0.001

Table A.3. (Continuation)

Panel C: South - All Males

	Dependent Variable: City Premia					
	(1)	(2)	(3)	(4)	(5)	(6)
	Private		Public		All	
	Initial	Medium	Initial	Medium	Initial	Medium
Log of Population in 2002	0.012*	0.019***	0.019	0.048***	0.016**	0.025***
	(0.005)	(0.005)	(0.010)	(0.010)	(0.005)	(0.005)
Constant	-0.164*	-0.252***	-0.252*	-0.584***	-0.220***	-0.324***
	(0.063)	(0.064)	(0.121)	(0.123)	(0.066)	(0.067)
Observations	306	306	306	306	306	306
R-squared	0.016	0.042	0.012	0.070	0.028	0.064
Adjusted R-squared	0.013	0.039	0.009	0.067	0.025	0.061

Panel D: South - Initial Poor Males

	Dependent Variable: City Premia					
	(1)	(2)	(3)	(4)	(5)	(6)
	Private		Public		All	
	Initial	Medium	Initial	Medium	Initial	Medium
Log of Population in 2002	0.008	0.026***	0.006	0.037**	0.012*	0.029***
	(0.005)	(0.006)	(0.013)	(0.014)	(0.006)	(0.006)
Constant	-0.094	-0.294***	0.010	-0.366*	-0.132	-0.332***
	(0.066)	(0.068)	(0.162)	(0.174)	(0.069)	(0.071)
Observations	306	306	306	306	306	306
R-squared	0.008	0.065	0.001	0.022	0.013	0.073
Adjusted R-squared	0.005	0.061	-0.003	0.019	0.010	0.070

*Notes: Initial city premia and medium city premia are calculated using the city fixed effects estimated in regression equivalent to the regressions reported in Table 5 for each of the specific sample. Please refer to notes of Table 5 for detailed information on the first stage regressions. Standard errors are in parentheses. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Table A.4. Why is Northern Brazil Different - Gap in Initial Premia

Panel A: All Males in the Private Sector

	Initial Premium		
	(1)	(2)	(3)
Is North (= 1 if it is North)	-0.121***	-0.082***	-0.034***
	(0.009)	(0.010)	(0.008)
Log of Population in 2002	0.010*	-0.042***	-0.023***
	(0.004)	(0.007)	(0.004)
Share of College Graduates	-0.629***		
	(0.116)		
Complexity Measure		0.066***	
		(0.007)	
Exposure of Low Skilled to High Skilled within Company at Microregion Level			0.090***
			(0.004)
Constant	-0.084	0.462***	0.201***
	(0.053)	(0.079)	(0.042)
Observations	558	558	558
R-squared	0.357	0.412	0.636
Adjusted R-squared	0.353	0.409	0.634

Panel B: Initial Poor Males in the Private Sector

	Initial Premium		
	(1)	(2)	(3)
Is North (= 1 if it is North)	-0.129***	-0.090***	-0.047***
	(0.008)	(0.009)	(0.007)
Log of Population in 2002	0.010*	-0.039***	-0.019***
	(0.004)	(0.006)	(0.003)
Share of College Graduates	-0.491***		
	(0.108)		
Complexity Measure		0.063***	
		(0.007)	
Exposure of Low Skilled to High Skilled within Company at Microregion Level			0.083***
			(0.004)
Constant	-0.068	0.457***	0.198***
	(0.050)	(0.073)	(0.039)
Observations	558	558	558
R-squared	0.400	0.466	0.665
Adjusted R-squared	0.396	0.463	0.664

*Notes: Initial city premia and medium city premia are obtained using the city fixed effects estimated in regression equivalent to the regressions reported in Table 5 for each of the specific sample. Please refer to notes of Table 5 for detailed information on the first stage regressions. Standard errors are in parentheses. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Table A.5. Why is Northern Brazil Different - Gap in Medium Term Premia

Panel A: All Males in the Private Sector

	Medium Term Premia		
	(1)	(2)	(3)
Is North (= 1 if it is North)	-0.119***	-0.078***	-0.031***

	(0.009)	(0.010)	(0.008)
Log of Population in 2002	0.019***	-0.036***	-0.015***
	(0.005)	(0.007)	(0.004)
Share of College Graduates	-0.631***		
	(0.118)		
Complexity Measure		0.068***	
		(0.007)	
Exposure of Low Skilled to High Skilled within Company at Microregion Level			0.091***
			(0.004)
Constant	-0.180***	0.387***	0.109*
	(0.054)	(0.080)	(0.043)
Observations	558	558	558
R-squared	0.358	0.418	0.637
Adjusted R-squared	0.354	0.415	0.635

Panel B: Initial Poor Males in the Private Sector

	Medium Term Premia		
	(1)	(2)	(3)
Is North (= 1 if it is North)	-0.148***	-0.101***	-0.059***
	(0.008)	(0.009)	(0.007)
Log of Population in 2002	0.014**	-0.039***	-0.016***
	(0.004)	(0.006)	(0.003)
Share of College Graduates	-0.373**		
	(0.113)		
Complexity Measure		0.069***	
		(0.007)	
Exposure of Low Skilled to High Skilled within Company at Microregion Level			0.087***
			(0.004)
Constant	-0.113*	0.464***	0.168***
	(0.052)	(0.075)	(0.041)
Observations	558	558	558
R-squared	0.426	0.506	0.688
Adjusted R-squared	0.422	0.504	0.686

*Notes: Initial city premia and medium city premia are obtained using the city fixed effects estimated in regression equivalent to the regressions reported in Table 5 for each of the specific sample. Please refer to notes of Table 5 for detailed information on the first stage regressions. Standard errors are in parentheses. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*

Table A.6. Why is Northern Brazil Different – Gaps in Wages – Mincerian Regressions

Panel A: All Males in the Private Sector

	Log of Average Monthly Wage in 2010 Real				
	(1)	(2)	(3)		(4)
Is North (= 1 if it is North)	-0.141***	-0.130***	-0.130***	-0.103***	-0.045***
	(0.0012)	(0.0012)	(0.0012)	(0.0013)	(0.0014)
Log of Population in 2002		0.013***	0.014***	-0.018***	-0.017***

		(0.0002)	(0.0003)	(0.0004)	(0.0003)
Share of College Graduates			-0.030**		
			(0.0102)		
Complexity Measure				0.048***	
				(0.0006)	
Exposure of Low Skilled to High Skilled within Company at Microregion Level					0.385***
					(0.0029)
Constant	6.451***	6.269***	6.263***	6.601***	6.339***
	(0.0065)	(0.0072)	(0.0075)	(0.0083)	(0.0072)
Observations	13,657,601	13,657,601	13,657,601	13,657,601	13,657,601
R-squared	0.802	0.802	0.802	0.802	0.803
Adjusted R-squared	0.776	0.776	0.776	0.777	0.777
Age, Company Tenure, Labor Market Experience	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Highest Education Level Achieved FE	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓

Panel B: Initial Poor Males in the Private Sector

	Log of Average Monthly Wage in 2010 Real				
	(1)	(2)	(3)		(5)
Is North (= 1 if it is North)	-0.155***	-0.142***	-0.142***	-0.107***	-0.048***
	(0.0012)	(0.0012)	(0.0012)	(0.0013)	(0.0014)
Log of Population in 2002		0.014***	0.013***	-0.026***	-0.019***
		(0.0002)	(0.0004)	(0.0005)	(0.0004)
Share of College Graduates			0.012		
			(0.0109)		
Complexity Measure				0.059***	
				(0.0006)	
Exposure of Low Skilled to High Skilled within Company at Microregion Level					0.409***
					(0.0031)
Constant	6.124***	5.927***	5.929***	6.372***	6.036***
	(0.0037)	(0.0051)	(0.0055)	(0.0072)	(0.0052)
Observations	7,922,781	7,922,781	7,922,781	7,922,781	7,922,781
R-squared	0.645	0.646	0.646	0.647	0.648
Adjusted R-squared	0.609	0.610	0.610	0.610	0.612
Age, Company Tenure, Labor Market Experience	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Highest Education Level Achieved FE	✓	✓	✓	✓	✓
Individual Fixed Effects	✓	✓	✓	✓	✓

*Notes: Please refer to notes of Table 5 for the full description of variables used. The estimates are calculated with using the high level of fixed effects estimates proposed by Correia (2016). Standard errors are in parentheses. We calculate them using a two-way clustering procedure at individual and microregion level. Significance level: * p <0.1; ** p<0.05; *** p<0.01.*