

Thema

UMR 8184

Théorie Économique, Modélisation et Applications

Thema Working Paper n° 2008-23
University of Cergy Pontoise, France

**Economic Geography and Wages in Brazil:
Evidence from Micro-Data**

T. Fally
R. Paillacar
C. Terra

2008



Economic Geography and Wages in Brazil: Evidence from Micro-Data

Thibault Fally

PSE – Paris School of Economics

Rodrigo Paillacar

CES – Université de Paris 1

Cristina Terra

THEMA – Université de Cergy-Pontoise

EPGE – Fundação Getulio Vargas

Abstract

This paper estimates the impact of market and supplier access on wage disparities across Brazilian states, incorporating the control of individual characteristics to the new economic geography methodology. We estimate market and supplier access disaggregated by industry, and we compute separately access to international and internal markets. We find a strong correlation between market access and wages differentials, even after controlling for individual characteristics, firm productivity, the source of market access (international, national or local), and using instrumental variables. Furthermore, market access turns out to be more important than supplier access.

1. Introduction

Trade liberalization episodes over the past decades have been accompanied by increasing inequality in most developing countries. In a survey about the distributional effects of globalization, Goldberg and Pavcnik (2007) discuss several possible reasons for a positive association between trade openness and inequality. Among others, weak labor reallocation after reforms is pointed out as a common pattern for many developing countries. Conventional approaches like the standard version of the Stolper-Samuelson effect, or the

effects of Skill Biased Technological Change are often unable to capture all the complexity embodied in wage and labor reallocation responses to trade openness, especially from a spatial perspective. Recent empirical papers exploring regional wage inequality through alternative mechanisms, such as intra-national trade costs and spatial sorting, have provided important insights.

Two seminal works in this direction are Hanson (2005) and Redding and Venables (2004), applied to US counties and to a sampling of countries, respectively. Both studies test structural models of New Economic Geography, finding a significant impact of trade costs on wages. Inspired on this approach, intranational studies have been applied to European NUTS regions (Head and Mayer, 2006), US states (Knaap, 2006) and Chinese provinces (Hering and Poncet, 2008).¹ Using a different framework, Chiquiar (2008) explains wage disparities across Mexican states after trade liberalization by exploiting regional exposure to international trade.

This paper builds on this literature by incorporating individual and firm dimensions in the estimation of a structural model of New Economic Geography (NEG) to explain wage disparities across regions and industries in Brazil. The model is able to explain the impact of location on wage inequality across Brazilian states and sectors, and to separate it from other sources of wage inequality such as differences in the composition of the labor force or the local diversity of industries.

The central relationship proposed by the NEG framework is the negative impact of trade costs on firm profits, captured by two structural terms denominated in the literature “Market Access” and “Supplier Access”. The first one measures the access to potential consumers, while the latter refers to the access to intermediate inputs. Both measures have a positive impact on profits and are negatively related to trade costs. The assumption of zero profits yields, then, a positive relation between wages and access to markets and suppliers, known as the “NEG wage equation” (for a derivation of this equation, see Redding and Venables, 2004).

Our empirical framework brings two noteworthy methodological contributions. First, we control for individual characteristics, which could be correlated with NEG structural variables

leading to spurious results.² In fact, controlling for individual characteristics is particularly important when studying Brazil, as we argue below. Second, we estimate market and supplier access using trade flows at the industry level, while other studies use aggregate trade flows.³ This procedure mitigates the collinearity problem found in the literature when attempting to estimate these two variables. While it is true that demand and supply should be naturally correlated at the aggregate level, as workers are also consumers, at the industry level it should not be true. A region may be specialized in producing in a particular industry, but uses inputs from all industries in general. Hence, we are better equipped to disentangle the effects of market and supplier access.

We implement a three step procedure. First, wages are regressed on firm's and worker's characteristics, controlling for regional and industry fixed effects. Second, following the NEG literature, gravity equations are estimated in order to calculate market and supplier access differentials across states. We compute separately access to international, national and local markets, and we compute market and supplier access for each industry separately. Finally, the market and supplier access derived in the second step are used as explanatory variables to the cross-region wage disparities captured by the region and industry fixed effects in the first step.

Brazil is a particularly interesting country to study with the NEG model controlling for individual characteristics, as it faces huge inequalities at both individual and regional levels. A Gini index of 0.59 in 2001, according to UNDP (2005), gives an idea of the individual inequality. Concerning the spatial dimension, in a study on the evolution of the regional GDPs in Brazil for the 1939-1998 period, Mossi et al. (2003) identify two spatial clusters in the country: a low-income one in the Northeast and a high-income one in the Southeast. Per capita income in São Paulo, the wealthiest Brazilian state, is 7.2 times that of Piauí, the poorest northeastern state, according to Lall et al. (2004).⁴

¹ All these papers follow the methodology proposed by Redding and Venables (2004), performing a structural estimation. Reduced forms are more common in the literature (e.g. Mion and Naticchioni, 2005 for Italy, Combes et al., 2008 for France, and Lederman et al, 2004 and Da Mata et al, 2005 for Brazil).

² To our knowledge, only Hering and Poncet (2008) control for worker characteristics in a NEG framework, and no other study have introduced firm productivity. Mion and Naticchioni (2005) also control for individual characteristics, but in a different framework.

³ Head and Mayer (2004) also use industry level data, but they compute only market access.

⁴ These authors also show that spatial inequality is extended to social indicators like child mortality and life expectancy. They provide detailed information about the regional policies implemented in Brazil to address these problems.

Moreover, Brazil is a very large country (in territory, it is the fifth largest country in the world) with a rich spatial variation of market and supplier access across regions, suggesting the relevance of estimating the NEG wage equation. North states are geographically closer to richest international markets in the Northern hemisphere, but the greatest internal demand is in the Southeast. Moreover, Mercosur integration turns Buenos Aires into another important center, which benefits regions in the South. More recently, states in the center of the country have experienced important growth due to their increasing exports of agricultural products. Furthermore, Brazilian governments have subsidized for long periods the location of industries in lagging Northeastern regions, obtaining mixed results in terms of competitiveness (Lall et al., 2004). As spatial inequalities are also important within Brazilian states, we explore two levels of spatial aggregation: states and municipalities.

Additionally, the distribution of activities across regions varies largely across industries. As an example, chemicals are mainly produced in Bahia, whereas transportation industries are mostly located in São Paulo. Detailed Brazilian trade data disaggregated at the industry and regional level permits us to build different variants of Market and Supplier Access, capturing all these aspects. Consequently, we are able to establish which sectors and which kind of trade (international or intranational) has the largest impact on wages through a NEG mechanism.

Controlling for worker and firm heterogeneity is crucial for the estimation of the NEG wage equation for the case of Brazil for three main reasons. First, individual diversity is extremely large and an important determinant of the wage inequalities in the country. In a classical work, Langoni (1973) shows that the main source of income inequality in Brazil is the heterogeneity across workers with respect to educational attainment, age, gender, industry and location. More specifically, Barros et al. (2000) show that the distribution of education and its returns account for about half of the wage inequality from observed sources in Brazil.

Second, we observe large spatial differences in human capital across regions: workers from Southern regions are more educated than those from Northern regions. Duarte et al. (2004) show that over 55% of the difference in the return to labor between the Northeast and the Southeast regions is due to differences in education attainment. This substantial difference in the educational level of the work force across regions may be explained by sorting (Combes and Duranton, 2006) or endogenous differences in returns to schooling (Redding and Schott 2003). Controlling for education at the individual level permits to precisely correct for the

bias induced by the differences in work force composition across regions, which could potentially affect the relation between market access and wages.

Finally, all empirical studies on firm and employment adjustments after trade openness emphasize the substantial changes in productivity and labor reallocation inside Brazil. Ferreira and Rossi (2003) uncover a significant increase in total factor productivity and labor productivity due to the decrease of trade barriers. Schor (2004), on her turn, shows that the response to openness was heterogeneous across firms. Hence, controlling for differences in firms' productivity is also important to avoid biased results.

The paper is organized as follows. Section 2 describes the methodology, with a very brief summary of the theoretical background and a description of the empirical strategy pursued. The data is described in section 3, while section 4 discusses the results. Section 5 concludes.

2. Methodology

2.1. Theoretical framework

In economic geography models, the interaction between monopolistic competition among firms and transport costs renders the spatial distribution of demand an important determinant of profits. This will be the background framework for our empirical exercise. A brief description of its main hypothesis and the results, rather than a full fledged model, will be presented here, as such models are now standard in the literature (see, for instance, Head and Mayer, 2004 and 2006, Redding and Venables, 2004).

Departing from the Dixit-Stiglitz-Krugman model of trade, preferences are assumed to have constant elasticity of substitution across product varieties.⁵ Each variety is assumed to be produced by a single firm under monopolistic competition. Producers and consumers are spread over different regions, and *ad valorem* trade costs, τ_{rs} , are assumed between any two regions r and s .

⁵ We follow Head and Mayer (2006) and Redding and Venables (2004) by deriving market access from Dixit-Stiglitz preferences. Market access could also be derived from alternative models though (e.g. from a quadratic demand function as in Melitz and Ottaviano, 2008).

Given these assumptions, in a symmetric equilibrium with n_{ri} firms in region r and industry i , the value of total sales from region r to region s , in industry i , X_{rsi} , can be shown to be:

$$(1) \quad X_{rsi} \equiv n_{ri} p_{si} x_{rsi} = \frac{n_{ri} (p_{ri} \tau_{rsi})^{1-\sigma} E_{si}}{P_{si}^{1-\sigma}},$$

where x_{rsi} represents the sales from a firm in region r to region s , in industry i , p_{ri} is the price received by the firm, so that $p_{ri} \tau_{rsi}$ is the price paid by a consumer in region s for a good from region r in industry i , σ is the elasticity of substitution between product varieties, and E_{si} is total spending in region s . $P_{si}^{1-\sigma}$ is the price index, defined as:

$$(2) \quad P_{si}^{1-\sigma} \equiv \sum_r n_{ri} (p_{ri} \tau_{rsi})^{1-\sigma}.$$

As for the costs of production, we assume that the firms use as inputs labor and intermediate goods, and incur a fixed cost. More precisely, in industry i , intermediate inputs consist in a composite of goods from all industries where ϖ_{ji} is the share of expense devoted to inputs from industry j , where, for each industry i , $\sum_j \varpi_{ji} = 1$. The total price index of intermediate inputs is defined by SA_{ri} , which denotes the ‘supplier access’:⁶

$$(3) \quad SA_{ri} \equiv \prod_j (P_{rj}^{1-\sigma})^{\varpi_{ji}}$$

It is worth noting that the NEG literature traditionally defines supplier access as the price index for manufacturers. We adopt a more precise definition, where supplier access is calculated separately for each industry, taking into account the inter-industry linkages. This procedure helps disentangling supplier from market access.

Hence, total costs by a firm in region r and industry i may be represented by:

⁶ This specification of the price index of intermediate input may be derived from a Cobb-Douglas production function, using input from all other industries.

$$SA_{ri}^{\alpha(1-\sigma)} w_{ri}^{\beta} \left(f_i + \sum_s x_{rsi} \right),$$

where α and β are parameters, f_i indicates the fixed cost in industry i , and w_{ri} is wage in region r and industry i .⁷ The supplier access is a measure for the firm's access to intermediate inputs. The larger the supplier access, the smaller will be the cost of intermediate inputs.

In maximizing profits, prices are set as a constant mark-up over marginal cost. Profits, then, can be shown to equal:

$$(4) \quad \Pi_{ri} = \frac{1}{\sigma} \left(SA_{ri}^{\alpha(1-\sigma)} w_{ri}^{\beta} \right)^{1-\sigma} MA_{ri} - f_i SA_{ri}^{\alpha(1-\sigma)} w_{ri}^{\beta},$$

where MA_{ri} is the 'market access', or 'real market potential', as denominated by Head and Mayer (2006), defined as:

$$(5) \quad MA_{ri} \equiv \sum_s \left(\frac{\tau_{rsi}^{1-\sigma} E_{si}}{P_{si}^{1-\sigma}} \right).$$

The market access will be larger when trade costs are smaller and real expenditure of the importing region is larger. The higher the market access, the larger will be the potential demand for the region's products in industry i .

We are finally able to relate regional wages to market and supplier access (hereafter, MA and SA, respectively). With free entry, profits must be zero in equilibrium. Given the profit function in equation (4), this equilibrium condition yields:

$$(6) \quad w_{ri} = \left(\frac{MA_{ri}}{\sigma f_i} \right)^{\frac{1}{\beta\sigma}} SA_{ri}^{\frac{\alpha}{\beta(\sigma-1)}}.$$

Hence, wages are larger in regions with higher MA, that is, with low trade costs to regions with high spending. Also, wages are higher in regions with higher SA, that is, where inputs can be bought at low prices due to low transport costs to suppliers.

2.2. Empirical Strategy

Our empirical implementation of the theoretical framework described above involves a three step strategy. First, wages are regressed on firm's and worker's characteristics, including region and industry fixed effects. The wage premium captured by these fixed effects is the variable to be explained by market and supplier access. Second, following the new economic geography literature, gravity equations are estimated in order to calculate market and supplier access differentials across states. Finally, the market and supplier access derived in the second step are used as explanatory variables to the cross-region wage disparities captured by the region and industry fixed effects from the first step.⁸ We explain each step in turn.

First step

While the theoretical framework described in the previous subsections treats labor as a homogeneous factor of production, we know that it is not the case. There is an extensive literature explaining wage differences across individuals by their characteristics, such as education attainment, years of experience, gender, marital status, among many other variables. For Brazil, in particular, the seminal work of Langoni (1973) presents evidence of the importance of workers' heterogeneity in income inequality. Were the individual diversity of the labor force similar across regions, we could still explain average regional wages by regional market and supplier access differences, as proposed by equation (6). Previous empirical work, however, has identified substantial differences in the composition in the labor force across Brazilian regions, especially with respect to education attainment (see Duarte et al., 2004). In this context, our results would be biased if we did not to control for individual characteristics and sorting across regions and sectors. Hence, the first step of our empirical study consists in estimating the following equation:

$$(7) \quad \log w_{lh,ri} = \mu_1 ed_l + \mu_2 ex_l + \mu_3 f_h + p_{ri} + \xi_l,$$

⁷ We assume that labor migration across regions is not sufficiently high to arbitrage away all regional wage disparities.

where $w_{lh,ri}$ is the wage of worker l working in firm h , industry i , of region r . ed_l represents the vector of variables capturing educational level of the worker, the vector ex_l captures the worker's experience, f_h is the firm's productivity and p_{ri} represents region/industry fixed effects.

The region/industry fixed effects capture wage disparities that are not explained by workers' and firms' characteristics, and that is the variable that will be used to be explained by the region and industry market and supplier access.

Second step

The second step consists on estimating the MA and SA, which is done in the following way. Equation (1) can be written as:

$$(8) \quad \log X_{rsi} = \log n_{ri} p_{ri} + (1 - \sigma) \log \tau_{rsi} + \log \frac{E_{si}}{P_{si}^{1-\sigma}},$$

and it is estimated using bilateral trade data, through:

$$(9) \quad \log X_{rsi} = FX_{ri} + \sum_k \delta_{ki} TC_{k,rs} + FM_{si} + \varepsilon_{rsi},$$

where FX_{ri} and FM_{si} are exporting and importing regions fixed effects for industry i , respectively, used to capture the first and the last terms of equation (8). A set of k variables, represented by $TC_{k,rs}$, is used to capture trade costs and estimate the middle term of the equation. These variables include the distance between the two regions (in log), whether they share borders, language, if contiguous, or had a colonial link.⁹

⁸ We thank an anonymous referee for suggesting an empirical procedure where regional dummies are regressed on MA.

⁹ Several alternative set of variables may be specified, but changing the specification of the gravity equation yields very little changes in the final-step results. For example, results are similar whether we introduce a dummy for pairs of countries that belong to Mercosur, whether we introduce distances by road (for intra-national trade only) instead of physical distance, or whether we estimate differentiated coefficients distance for intra versus international trade. Finally, Paillacar (2007) shows that Gamma PML yields similar results to OLS.

Regression (9) is estimated for aggregate trade flows, as in Redding and Venables (2004), and, alternatively, it is also estimated separately for each industry. This second approach allows for differential impacts across industries of trade cost and of fixed effects from importer and exporter regions, and it is applied as follows.

From the definition of MA, in equation (5), the estimated coefficients in equation (9) can be used to compute it as in:

$$(10) \quad \hat{MA}_{ri} \equiv \sum_s \left[(\exp FM_{si}) \prod_k (\exp TC_{k,rsi})^{\delta_{ki}} \right].$$

We have, then, a market access measure for each industry separately, in each region.

As for the SA, defined in equation (3), its estimated value is computed in a similar fashion, but using the coefficient from the exporting region dummy variables. To account for the vertical linkages across industries, we use the coefficients from the input-output matrix, $\hat{\omega}_{ji}$, to weight the impact of each industry in supply access. We, then, compute:

$$(11) \quad \hat{SA}_{ri} \equiv \prod_j \left\{ \sum_s \left[(\exp FX_{sj}) \prod_k (\exp TC_{k,rsj})^{\delta_{kj}} \right] \right\}^{\hat{\omega}_{ji}},$$

which yields a SA measure for each industry in each region.

This paper is the first one to weight industry supplier access by an input-output matrix in the structural approach proposed by Redding and Venables (2004). Amiti and Cameron (2007) also apply this method, but in a somewhat different framework.

Third step

Finally, the estimated values for MA and SA are used to explain differences in wage premium across regions and industries. Equation (6) can be written as:

$$(12) \quad \log w_{ri} = -\frac{1}{\beta\sigma} \log \sigma f_i + \frac{1}{\beta\sigma} \log MA_{ri} + \frac{\alpha}{\beta(\sigma-1)} \log SA_{ri},$$

which is estimated through:

$$(13) \quad \log p_{ri} = \theta_0 + \theta_1 \log \hat{MA}_{ri} + \theta_2 \log \hat{SA}_{ri} + \theta_3 D_i + \zeta_{ri}.$$

where D_i represents industry dummies, and p_{ri} are the region/industry fixed effects estimated in the wage regression (7). Hence, MA and SA are used to explain wage differences across regions and industries that are not explained by workers' and firms' characteristics.

3. Data

We use three sets of data: on individual characteristics, trade flows and country characteristics. We focus on a cross sectional analysis for 1999, as the intra national trade data by industry is only available for that year (Vasconcelos and Oliveira, 2006).

Individual characteristics are drawn from the RAIS database (*Relação Anual das Informações Sociais* from the Brazilian Labor Ministry) that covers all workers in the formal sector¹⁰. We focus on the manufacturing sector for the compatibility with the trade data. As several jobs may be recorded for the same individual, we select the highest paying job¹¹. The data provides important individual characteristics (wages, educational level, age, gender, etc.) as well as worker and firm identification numbers which allows the matching of the RAIS database with the manufacturing survey.

The manufacturing survey, PIA (*Pesquisa Industrial Anual* from IBGE, *Instituto Brasileiro de Geografia e Estatística*), covers all firms above thirty employees from 1996 to 2003, representing most of the workforce in the manufacturing sector. This dataset provides a large range of variables on production, including sales, labor, materials, energy and investments,

¹⁰ Because of the very large number of observations, we run our regressions on random samples of 500000 or 800000 employees. Changing the size of the sample does not affect our coefficients and particularly the estimation of region and industry fixed effects.

which permits measuring productivity (see appendix). We complete the PIA using balance sheets data from IBRE-FGV (*Instituto Brasileiro de Economia – Fundação Getulio Vargas*) in 1995, from which we draw initial fixed capital. The two datasets can be matched thanks to the firm identification number. We refer to the appendix for a complete description of all variables.

In order to estimate market access we need three sets of trade data: (1) trade data between Brazilian states, which is drawn from Vasconcelos and Oliveira (2006) who processed the information of the value-added tax provided by the National Council of Financial Policy (CONFAZ, *Conselho Nacional de Política Fazendária*) from the Ministry of Finance (*Ministerio da Fazenda*); (2) trade data between Brazilian states and foreign countries (*Secretaria de Comércio Exterior*, Ministry of Trade); and (3) between foreign economies (BACI: *Base pour l'Analyse du Commerce International*, CEPII). Moreover, using total sales across regions and industry from the PIA database allows computing internal flows within regions by subtracting intra and international exports. These sets of data provide a complete and consistent picture of all trade flows, defined at the 2 digit ISIC Revision 3 level (corresponding to the Brazilian CNAE 2 digit industry classification).

We complete the trade and individual information with a variety of data on geography, infrastructures and regulations. Distances, colonial links, languages, coordinates, GDP, areas and demographic densities are provided by CEPII (*Centre d'Etudes Prospectives et d'Informations Internationales*) and IBGE. The distance between states in the geodesic distance between their respective capitals (computed in km using the coordinates).

We construct a dummy for international border that equals zero if both the origin and the destination of the trade are within the same country, and it equals 1 otherwise. Analogously, the dummy internal border equals zero if the trade is within one Brazilian state, and it equals 1 otherwise. The dummy international contiguity, on its turn, equals 1 if international border equals 1 and if both countries (or the country and the Brazilian state) share a border. Analogously, the dummy internal contiguity equals 1 when both Brazilian states share a border. The dummy for language equals one if the trade is between two different countries (that is, the international border dummy equals 1) and they share the same language (more

¹¹ For example, a worker may change occupation or establishment over time, or may even have two recorded jobs in the same time. To assess the robustness of our results, we alternatively chose the average wage, the total

precisely, if the official language in the same or if the same language is used by at least 20% of the population). The colonial link dummy, on its turn, equals 1 if the trade is between two different countries and one of them has been colonized by the other in the past.

Migration rates by municipalities are from the Census 2000 (IBGE). The input-output matrix is constructed by OECD and IBGE, across ISIC Rev3 2-digit industries. The cost to start a business has been measured by the World Bank for 13 Brazilian states (Doing Business database). An index of tax pressure across Brazilian states is constructed using the PIA data. Agricultural harvested area in 1996 is from the Agricultural Census. Regional shares of mineral production are taken from *Anuário Mineral Brasileiro 1999* (table 8 page 51). Municipality level data on natural endowments is from Timmins (2006).

4. Results

4.1. Preliminary regressions

We start by performing the first two steps of our empirical procedure described in section 3.2, that is, by estimating wage premium differentials across regions and computing estimated values of real market potential and supplier access.

First step: wage premium

Equation (7) is estimated using individual data for male workers between 25 and 65 years old. This group of workers was chosen to render the sample more homogeneous, thus eliminating possible effects from differences in variables such as early dropouts and female participation. For measuring education, we use dummy variables for 9 levels of education.¹² Age and age squared are used as proxies for experience. State and industry fixed effects are also included. Table 1 presents the results. The only difference between the regression results presented in columns (1) and (2) is that the latter includes firms' productivity as an additional control.

wage, in December or over the year, and the choice didn't affect the results.

¹² The description of the 9 education levels is in the appendix.

We should note that the R-squared (adjusted or not) is very high. This is explained by the very large wage inequality across industries and regions compared to the overall wage inequality among males in the formal manufacturing sectors. If region/industry dummies are excluded, 34.1% of the variance is explained by worker characteristics (not shown in the tables). If worker characteristics are excluded, region/industry dummies explain 83.1% of the variance, but a large fraction of this variance also corresponds to differences in age and educational attainment. These comparisons indicate that wage differentials across regions and industries constitute the main source of inequalities in Brazil, but a large part may be explained by systematic differences in individual skills, showing the importance of controlling for worker characteristics in our investigation.

Table 1: Wages and individual characteristics

Dependent variable: wages		
	(1)	(2)
Firm productivity		0,297** [0,002]
Age	0,072** [0,001]	0,084** [0,001]
Age squared / 100	-0,072** [0,001]	-0,084** [0,001]
Education (level 5 = 0):		
level 1	-0,365** [0,005]	-0,418** [0,006]
level 2	-0,239** [0,003]	-0,273** [0,003]
level 3	-0,149** [0,002]	-0,160** [0,003]
level 4	-0,075** [0,002]	-0,084** [0,002]
level 6	0,156** [0,003]	0,149** [0,003]
level 7	0,419** [0,002]	0,371** [0,003]
level 8	0,852** [0,004]	0,799** [0,005]
level 9	1,240** [0,003]	1,168** [0,004]
State x Industry FE	yes	yes
R-squared	0,880	0,899
Observations	798494	499144

Notes: OLS regressions with robust standard errors.

Statistical significance: ** 1% and * 5% level.

Second step: market potential and supplier access

For the computation of estimated values of market potentials and supplier access, we start by estimating the gravity equation (9). Each state of the country is defined as a region. We follow

two procedures. In the first one, the coefficients are taken to be the same for all industries, as usually done in the literature, while in the second a regression is run separately for each industry, estimating, thus, different coefficients for each one of them. Table 2 presents the results, where the first column displays the coefficients for the regression taking the aggregate trade flows, and the second presents the average value of the estimated coefficient across industries for the regressions by industry.

Table 2: Gravity equations

Dependent variable: trade flows		
	Aggregated ^(a)	By industry ^(b)
Physical distance	-1.448** [0.018]	-1.359 (0.180)
International border	-4.326** [0.116]	-4.534 (0.983)
International contiguity	1.001** [0.095]	0.785 (0.184)
Internal border	-2.594** [0.386]	-3.212 (0.968)
Internal contiguity	0.128 [0.225]	0.205 (0.469)
Language	0.839** [0.043]	0.604 (0.263)
Colonial link	0.832** [0.100]	0.903 (0.140)
Exporter FE	yes	yes
Importer FE	yes	yes
Industries		22
R-squared	0.982	
Observations	25315	total: 246833

Notes: (a) OLS regressions with robust standard errors between brackets, exporter and importer fixed effects. Dependent variable: aggregated trade flows. Statistical significance: ** 1% and * 5% level.

(b) OLS regressions by industry. Dependent variable: trade flows by industry. Mean and standard deviation of estimated coefficients across regressions.

With the estimated coefficients from regression (9), presented in Table 2, equation (10) was used to compute the estimated values for the market potential of each region. Notice that, when calculating MA, we add over all regions with which a particular region trades. It is then possible to construct a market access measure from a subgroup of trade partners, and that is exactly what we do to investigate the different impacts of local, national and international market access.

Before jumping to the estimation of the NEG wage equation, it is worth visualizing the relation between wages and MA estimated in steps 1 and 2. In Figure 1, four maps of Brazil

display the spatial distribution of wages and MA.¹³ Panel (a) presents the regional wages, *after controlling for individual characteristics*. It is clear that, even after skill sorting is taken into account, there are still substantial spatial differentials in wages across states, São Paulo being the region offering the highest wages, followed by nearby states (Rio de Janeiro, Paraná, Minas Gerais, among others). Interestingly, the State of Amazonas, a landlocked region in the less populated north, also exhibits a high wages.

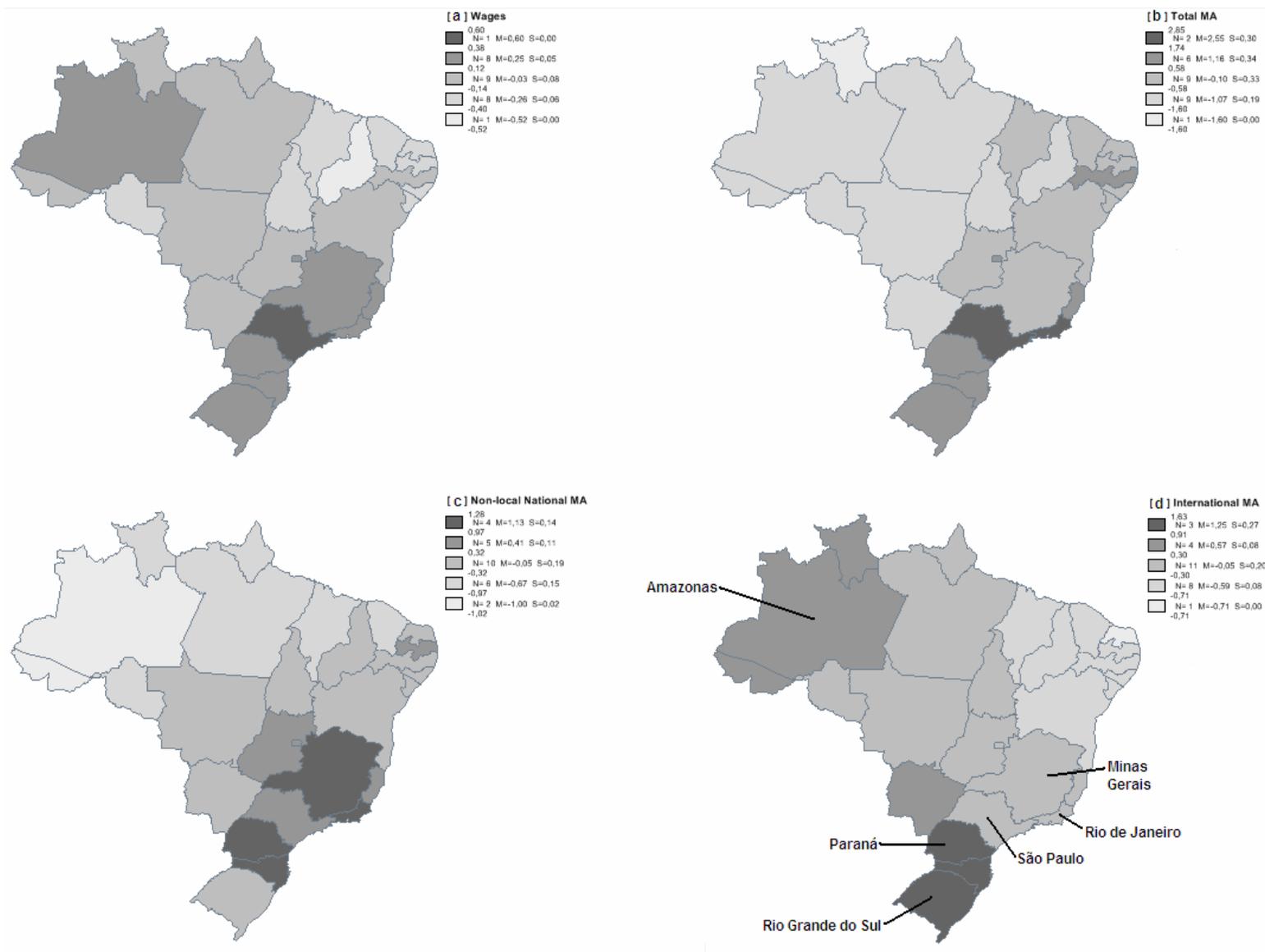
Panel (b) displays the total MA across regions. The visual impression from the figure is that MA and regional wages are indeed related. São Paulo is the state with largest MA (followed by Rio de Janeiro), which is compatible with the fact that wages are the highest in that state. For Amazonas, though, its MA is not differentiated from that of the rest of the North region. A more precise vision of the channels at work is obtained by decomposing MA into its national and international scopes.

In Panel (c) we highlight the role of inter-regional trade, by excluding international and local (i.e., own state) MA. As expected, the states neighboring São Paulo exhibit highest non-local national MA, while the value for São Paulo itself becomes smaller. More interestingly, this exercise shows that Amazonas and Rio Grande do Sul (the southern region closest to Argentina) are “too far” from the main national demands.

Panel (d) completes the picture by considering only international MA. We see that that these two regions have a different source of market potential to explain their high wages: its international dimension. Rio Grande do Sul is close to Buenos Aires, the other important economic center of MERCOSUR (besides São Paulo). The state of Amazonas, on its turn, is close to medially developed countries in South America (Colombia and Venezuela), and NAFTA members. Nevertheless, this is only one part of the explanation. Manaus, the capital of State of Amazonas is a Free Trade Zone, with important import and export volumes, explained by industries linked to assembling.

¹³ Values are normalized as deviations from the mean across regions, and they are grouped in 5 classes. The middle group is comprised between the mean +/- 0.5 standard deviation, and the next groups are delimited by 1 standard deviation.

Figure 1. Market Access and Wages across Brazilian States.



4.2. Wages and market access

Our baseline regression is presented in the first column of Table 3. In that regression, the wage differentials captured by the region/industry fixed effects in the first step of our empirical procedure is regressed on MA, as calculated in the second step. It shows that those wage differentials are positively and significantly correlated with MA.¹⁴

The coefficient of MA we find in our baseline regression is larger than the one estimated by Head and Mayer (2006) for Europe, and smaller than that in Redding and Venables (2004) in a cross country study. There are a few differences between our procedure and that of those two papers. For one, we control wage premium for individual characteristics. As Redding and Venables (2004) do not control wages for any individual variables, as we do, their estimated coefficient could be capturing different composition patterns of the labor force across countries, and hence their larger coefficient. Head and Mayer (2006) do control for education, but only at the aggregate level. Our estimated coefficient is closer to that found in Hering and Poncet (2008), which also controls wages disparities by individual characteristics in a study for Chinese regions. It should be noted that when we do not control for individual skills in the first stage (results in column 2 of Table 3), the coefficient for market access increases from 0.14 to 0.17.

Another difference from our work to that of Redding and Venables (2004) is that we construct market access departing from gravity equations by industry. When we construct market access using aggregate trade flows in the gravity equation instead (results in the third column), we find a coefficient around 0.08, which is lower and closer to that of Head and Mayer (2006) for European regions. This coefficient, however, is estimated less precisely than the one using trade flows disaggregated by industry.

¹⁴ Since the predicted values for market access and the wage premium are generated from prior regressions, we checked the sensibility of our results to bootstrap techniques. Results remained unchanged and bootstrapped standard errors were slightly lower than robust standard errors reported in the tables. Redding and Venables (2004), De Sousa and Poncet (2007) and Hering and Poncet (2007) also found bootstrapped standard errors that were close to the non-bootstrapped estimation.

In our baseline regression, MA explains 35% of wages disparities across regions and industries. If we use only industry dummies, instead of MA, only 17.5% of wages differentials are explained (regression not reported): MA has a stronger explanatory power than only industry dummies. Nevertheless, disaggregating by industry does help explaining wage differentials, as can be seen by the comparison of the results in the first and third column of Table 3.

We use our separate measures of MA to analyze the different impacts of local, national and international MA. When we drop the local market access and consider only the access to other Brazilian states and other countries (results in row 4), we still find a large and significant coefficient. The coefficient is even larger than the one in the first row, which also included local market access, although the difference is not statistically significant.

Rows 5 and 6 present the results when only national (excluding local) and when only international market access are considered, respectively. It is worth noting that international market access alone yields the highest impact on wages, and its coefficient is estimated with the highest precision, compared to the other subgroupings of market access. The R-square of the regression with international market access is also higher compared to the other subgroupings, although still smaller than that of total market access (column 1).

Table 3: Response of wage premium to market access

Dependent variable: wage premium						
Measure of MA	(1) Total MA	(2) Total MA	(3) Total MA (aggregate)	(4) Non-local MA	(5) National MA	(6) International MA
Market Access	0,140** [0,012]	0,168** [0,013]	0,079** [0,026]	0,185** [0,022]	0,162** [0,021]	0,228** [0,018]
Controlling for skills in 1st step	yes	no	yes	yes	yes	yes
By industry	yes	yes	no	yes	yes	yes
Industry FE	yes	yes	no	yes	yes	yes
R-squared	0,350	0,432	0,275	0,255	0,248	0,294
Observations	540	540	27	540	540	540

Notes: OLS regressions with standard errors robust to heteroskedasticity and industry fixed effects (except column 3).

Dependent variable: wage premium (see section 2.2, fixed effects from the regression of individual wages on individual characteristics)

Regressor: market access (see section 2.2, calculated from a gravity equation on intranational and international trade flows); "non local": excluding own state; "national": excluding foreign and local markets; "international": foreign countries with a common frontier with Brazil.

Statistical significance: ** at 1% level.

This interesting result may be explained by the trade liberalization that took place in the early 1990's. The lowering of trade barriers that occurred over the first half of that decade could have had differential impacts over the country precisely due to the differences in international MA across the regions. The impact of trade liberalization would be larger within regions with larger international MA. In the late 1990's, the period we study, it is possible that the labor mobility across regions had not yet been sufficient to arbitrage the differential impact on wages of the opening to trade. In a study for the Mexican trade liberalization, Chiquiar (2008) shows that, after the second stage of trade liberalization, "regions with a larger exposure to international markets exhibited a relative increase in wage level". We may be capturing a similar pattern for Brazil.

4.3. Market access and supplier access

So far we have studied the impact of MA on wages. As discussed in section 2.1, MA captures how close a firm in a given region is to consumers, whereas SA establishes the proximity to suppliers of intermediate goods. While MA has a positive impact on wages through the demand effect, the impact of SA is through lower costs and higher productivity.

A usual problem with MA and SA measures is that they tend to be strongly correlated. To address this issue, Redding and Venables (2004) add additional assumptions on the link between MA and SA. With our procedure, however, this problem is mitigated, without having to resort to supplementary restrictions. By calculating MA and SA for each industry as we do, market and supplier access are less likely to be correlated.

In the same fashion of what we did for the MA, we take a three step procedure for SA. In the first step we regress wages on individuals' characteristics and on region/industry dummies. Second, we compute the SA measure following equation (11), as described in section 2.2.¹⁵ Finally, we use SA as explanatory variable of the region/industry wage disparities estimated in the first step. The results are presented in Table 4.

¹⁵ As we need the exporter fixed effects for the industry's inputs, we cannot compute SA for industries using non-industrial inputs. Therefore, regressions in Table 4 exclude Food and Beverages, Tobacco, Wood and Fuel Refinement.

When only SA is used as explanatory variable (regression presented in the first column), its estimated coefficient has the same value as the one found for market access. The same is true when only non-local SA is used (results in the second column). This could be a sign that our MA and SA measures are actually correlated, so that both variables are capturing the same effect.

In order to investigate whether these two measures have independent impact on regional wages, we include both simultaneously as explanatory variables. The results (in the third column) show that both have a positive and significant impact on wages, and that the impact of MA is stronger than that of SA.

Using only non local SA and MA (results in the forth column), only MA has a positive and significant coefficient. Note the high standard deviations, indicating that co-linearity problems may be more important for non local market and supplier access than for local ones.

One concern with the SA measure is its possible correlation with its own industry. To deal with it, we create an instrument by computing SA excluding its own industry, that is, we set to zero the input-output matrix coefficient for own industry and normalize the other coefficients to have a sum equal to one. When this instrument is used in the place of SA (results in the fifth column), we still find positive and significant coefficients for SA and MA, and the distance between them increases.

Table 4: Market Access and Supplier Access

Dependent variable: wage premium					
	(1)	(2)	(3)	(4)	(5)
Supplier Access	0,140** [0,010]		0,066** [0,017]		0,049* [0,019]
Non local SA		0,193** [0,023]		-0,019 [0,078]	
Market Access			0,108** [0,021]		0,124** [0,021]
Non local MA				0,238** [0,079]	
SA instrumented	no	no	no	no	yes
Industry FE	yes	yes	yes	yes	yes
controlling for skills	yes	yes	yes	yes	yes
R-squared	0,347	0,222	0,384	0,241	0,382
Observations	441	441	441	441	441

Notes: OLS and 2SLS regressions with robust standard errors and industry fixed effects.

Regressors: s upplier and market access (see section 2.2, calculated from a gravity equation on intranational and international trade flows); "non local": excluding own state. Instrument for supplier access in column (5): supplier access excluding own industry in the input-output matrix.

Statistical significance: ** 1% and * 5% level.

4.4. Productivity

In this section we investigate whether the positive correlation found in the previous sections between MA and regional wage differential could be related to differentials in productivity that could have been produced by MA itself. Redding and Venables (2004) model emphasizes the labor demand effect through MA, but recent models on international trade and the selection of firms show that access to foreign markets may have a positive impact on average firm productivity, which in turn has a positive impact on wages (Melitz, 2003, and Melitz and Ottaviano, 2008). Baldwin and Okubo (2006) describe in a model how MA may impact productivity across regions.

It is, thereby, possible that the impact of MA on wage differentials is due to its impact on productivity. Hence, in the same way we have controlled for labor's individual characteristics, we also control for productivity when estimating wage differentials in the first step.¹⁶ Our dataset allows for this control as we are able to match the data on workers with the data on firms (see details in the Appendix). We find a positive and significant elasticity of wages to

¹⁶ Mion and Nattichioni (2005) also mean to control for productivity, but they only have firm size as a proxy for productivity.

productivity, close to 0.3, as shown in the results presented in the second column of Table 1 in section 4.1.

Table 5 presents the results of the impact of MA on region/industry wage differentials, controlling for productivity. The first column of Table 5 is the equivalent of that of Table 3, but using the wage regression in the first step that also controls for firm productivity, that is, using the region/industry fixed effects from the regression in the second column of Table 1. The wage premium corrected for productivity is still highly correlated with MA, though the coefficient is slightly smaller: 0.11 instead of 0.14.

The same comparison is true for the regression comprising only national and only international MA (in columns 2 and 3, respectively): controlling for productivity in the first step renders the coefficient for MA smaller in the third step. Additionally, the R-square of the regressions controlling for firms' productivity is larger than the regressions without such control.

Table 5: Controlling for productivity

Dependent variable: wage premium				
	(1)	(2)	(3)	(4)
Market Access	0,112** [0,011]			0,112** [0,010]
National, Non local MA		0,134** [0,018]		
International MA			0,201** [0,017]	
Productivity	\	\	\	0,294** [0,057]
Controlling for Productivity in 1st step	yes	yes	yes	no
Skills in 1st step	yes	yes	yes	yes
Industry FE	yes	yes	yes	yes
R-squared	0,403	0,328	0,388	0,510
Observations	466	466	466	419

Notes: OLS regressions with robust standard errors and industry fixed effects.

See section 2.2 and data appendix for the measure of productivity.

Statistical significance: ** 1% and * 5% level.

Alternatively to controlling for productivity in the first step, we control for it directly in the third step. The estimated coefficient for MA remains the same, as shown in the results

presented in the fourth column of Table 5. We have also controlled for productivity both in the first and third steps (regression not reported), and, interestingly, we have still found a positive and significant coefficient for productivity in the third step. This could be the working of either positive externalities of productivity or sorting effects.

The relations we find between MA and productivity, presented in Table 6, are in line with the predictions of Melitz (2003) model. According to that model, higher international access increases selection among firms, thus increasing average productivity and decreasing its dispersion. Conversely, local MA allows lower productivity firms to survive, yielding the opposite effects on productivity. According to the results in the first column of the table, we find a positive and significant correlation between productivity and global MA, although not strong. When splitting MA between national and international (results in the second column), we find a non-significant coefficient for national MA, while for international MA the coefficient is positive, significant, and of larger magnitude. Conversely, productivity dispersion has a positive relation with national and negative one with international MA, both coefficients being significant (results in the third column).

Table 6: Productivity and market access

Dependent variable:	Productivity		Productivity dispersion
	(1)	(2)	(3)
Market Access	0,022 α [0,012]		
National MA		-0,012 [0,019]	0,020 α [0,011]
International MA		0,062** [0,020]	-0,021* [0,009]
Industry FE	yes	yes	yes
R-squared	0,079	0,087	0,256
Observations	420	420	242

Notes: OLS regressions with robust standard errors and industry fixed effects.

See section 2.2 and data appendix for the measure of productivity;

Dispersion is the standard deviation of the log of productivity by region and industry.

Statistical significance: ** 1%; * 5% level and α 10% level.

4.5. Further robustness checks

Differences across skills

One of the underlying assumptions of our methodology is that returns to education are constant across states, that is, they are independent of MA. This assumption allows controlling for education in the first step independently from the final step regression. Theoretical papers have shown, though, that MA may affect the skill premium and returns to education (see for example Redding and Schott, 2003): on the one hand, skilled workers are more mobile; on the other hand, the concentration of activity may particularly increase the productivity of skilled workers through either increasing returns to scale or pervasive input-output linkages in skill-intensive sectors. The results in the first two columns of Table 7, however, indicate that this link is not relevant in the Brazilian case: the observed correlation between wages and MA does not seem to vary significantly across educational levels.

In column (1), where the wage premium is constructed only from data on skilled workers (completed high school or higher), the coefficient for MA is higher but not statistically different from the coefficient of the first column in Table 3 (same specification with all workers). In column (2), the wage premium is constructed using data on unskilled workers only (uncompleted high school or lower), and the coefficient obtained for MA is close to the baseline regression in Table 3. Hence, the impact of market access on wages seems stronger on skilled workers, but the difference is not significant.¹⁷ In other words, returns to education are not strongly correlated with market access, which validates our methodology in the first step.

When we consider only international MA, we obtain different and very interesting results. The coefficient for international MA on wage differences across region/industry is significantly larger among unskilled than skilled worker (results in columns (3) and (4) of Table 7).

This result means that higher international MA augments relatively more the wages of unskilled workers. Having in mind that our study corresponds to a period just a few years after a massive trade liberalization program, this result could actually be a sign of the working

¹⁷ Alternatively, we have also directly regress the skill premium on MA (results not reported). The coefficient has the expect sign but it is not significant.

of the Stolper-Samuelson mechanism. According to it, trade liberalization in Brazil, a country relatively abundant in unskilled labor, should increase relative returns to that factor of production. On an economic geography perspective, the impact would not be homogeneous across the country: it would be larger in regions with higher international MA. This interpretation is in line with the findings of Gonzaga et al. (2006), which present evidence for Brazil of relative wage changes compatible with Stolper-Samuelson predictions.

Table 7: Wage premium to market access - skilled versus unskilled workers

Dependent variable: wage premium				
	(1)	(2)	(3)	(4)
Workers:	Skilled	Unskilled	Skilled	Unskilled
Market Access	0,160** [0,014]	0,134** [0,011]		
International MA			0,196** [0,023]	0,229** [0,017]
Industry FE	yes	yes	yes	yes
R-squared	0,373	0,387	0,278	0,344
Observations	504	532	504	532

Notes: OLS regressions with standard errors robust to heteroskedasticity and industry fixed effects.
 Skilled workers: educational level above high school.
 Statistical significance: ** at 1% level.

Controlling for taxes and regulations

Whereas public infrastructures are generally better in the South, successive governments have tried to improve infrastructures in remote regions and have adopted fiscal incentives to promote industrial development in lagging regions, with various degrees of effectiveness. If infrastructures and tax rates are positively or negatively correlated with market access, not controlling for these variables would bias our results. In table 8, column (1), we regress the wage premium on market access and tax rates estimated from the firm-level data (sum of taxes paid by each firm divided by total sales). Although taxes have a negative impact on wages, controlling for taxes only marginally affects the coefficient for market access.

In column (2), we use data on firm’s entry costs from the Worldbank (Doing Business in Brazil), which provides an estimation of the cost to start a business across 13 states in Brazil. When we control for this variable, the coefficient for MA is still significant but lower, which means that part of the correlation between MA and wages might actually be explained by the negative correlation between entry costs and wages, and a negative correlation between entry

costs and MA.¹⁸ While both coefficients are statistically significant, we should still keep in mind that results from column (2) rely on variations across a small number of states.

Controlling by a dummy for São Paulo

São Paulo is the economic center of Brazil and attracts a large fraction of the Brazilian manufacturing industry. Also, it is located in a central position between Rio de Janeiro and Argentina. As both the market access and the wage premium are the largest for São Paulo, a reasonable concern is whether wages are high in São Paulo for reasons unrelated to MA, and, yet, the correlation between MA and wages found so far are driven by this state. As a robustness check, we estimate the response of wages to MA including a dummy for São Paulo. The coefficient for MA, in column (3) of Table 8, remains large and strongly significant, albeit smaller. Hence, our main result is not driven by the state of São Paulo. Additionally, the positive and significant estimated value for the São Paulo dummy reveals a larger wage premium for that state in addition to the impact of its highest market access.

Controlling for endowments

Endowments are unequally distributed across Brazilian states and play an important role in explaining wages differentials. In addition, endowments may be correlated with market access, thus biasing our coefficient. Notice that, in section 4.3, the correlation between wages and MA is not affected when we restrict our analysis to sectors that do not depend on natural resources (see footnote 15). We now perform another robustness check in which we directly control for endowments. In column (4) of table 8, we control for minerals, harvested land area, access to the sea and dummies for macro regions.¹⁹ As expected, wages are positively correlated with the presence of natural resources: the coefficient for harvested land is positive and significant; minerals (share of total national extraction) have a positive and significant coefficient; access to sea (minus landlocked) has a positive albeit not strongly significant effect on wages. Among macro-region dummy variables, only the Northeast is significant at the 1% level: its value is -0.22 (with an estimated standard error of 0.06). This may be

¹⁸ Alternative indices of entry costs (time to start a business and number of procedures) are positively correlated with market access. The resulting coefficient for MA is actually higher when alternative indices are used in the regression.

¹⁹ The Brazilian states are regrouped in five macro-regions based on geographical characteristics. They are North, Northeast, Southeast, South and Center-West. Our reference category is Southeast.

partially explained by its difficult climate (e.g. frequent droughts).²⁰ In spite of the inclusion of these controls, the coefficient for MA remains very large and significant.

As climate and land seem to have the strongest impact among the different types of endowments, we performed further robustness checks using more detailed data. This analysis is discussed in section 4.6 as these variables are available by municipality.

Table 8: Further robustness checks

Dependent variable: wage premium				
	(1)	(2)	(3)	(4)
Market Access	0,145** [0,011]	0,099** [0,013]	0,120** [0,013]	0,116** [0,016]
Taxes	-0,231** [0,037]			
Entry cost		-0,011** [0,001]		
Dummy São Paulo			0,280** [0,042]	
Harvested land				0,027* [0,011]
Minerals				0,028* [0,012]
Landlocked				-0,039 [0,055]
Industry FE	yes	yes	yes	yes
Macro-region FE	no	no	no	yes
controlling for skills	yes	yes	yes	yes
R-squared	0.384	0.522	0.364	0.875
Observations	540	277	540	540

Notes: OLS regressions with robust standard errors and industry fixed effects.

Controls: taxes: average tax/sales ratio of industrial firms in the state; entry cost: cost of creating a business / GDP for 13 Brazilian states (source: Worldbank); Minerals: regional share of mineral production (source: *Anuario Mineral Brasileiro 1999*); harvested land area in 1999 (source: Agricultural Census).

Statistical significance: ** 1% level; * 5% level.

Instruments for Market Access

Last but not least, one may be concerned about the endogeneity of market access. Wages might positively affect individual demand for goods, thus increasing the index of market access. Similarly, a productivity shock in a region would affect both wages and the index of market access if productivity also impacts the demand for goods. Considering “non local” market access constructed by excluding the demand from the own market, as we have already

²⁰ When we include only the Northeast dummy in the regression, other coefficients remain similar (with the effect of harvested land being more significant). Interestingly, the coefficient for “landlocked” becomes positive

done in column (4) of Table 3, mitigate such biases. Here, we instrument the index of MA by geographical and demographic variables which should impact market access but not directly affect the wage premium across regions.

An instrument proposed by Head and Mayer (2006) is the distance to the main economic center. In row (1) of Table 9, we thus instrument MA by the distance to São Paulo. In row (2) we run an alternative specification by instrumenting international MA by the distance to Buenos Aires (first stage regressions are provided in the appendix). Interestingly, estimated coefficients are stronger compared to the baseline OLS regression. One of the potential problems is the choice of the economic center, which may be itself endogenous: both centers exhibit high wages and high MA. More specifically, the distance to São Paulo may capture effects that are not related to MA: for example, proximity to firms' headquarters, and therefore managerial power, may have a positive impact on wages.

In order to deal with these issues, we propose three alternative instruments. First, we consider a “Harris Market Potential” (HMP, sum of other regions' GDP divided by the distance) constructed using GDP by states in 1939:

$$(14) \quad HMP_r = \sum_s GDP_s / Dist_{rs}$$

The main advantage is that it does not suffer from the (endogenous) choice of one economic center. Using HMP in 1939 as an instrument relies on the assumption that wages in 1939 are only indirectly related to current wages (which is a reasonable assumption given the evolution of technologies). In case this hypothesis is not be verified, we use a second instrument, which is the size of the population instead of GDP. Thus, we construct a “Stewart Market Potential” using population by state in 1940 divided by the distance. As shown in rows (3) and (4), both instruments yield significant and strong coefficient for MA, yet smaller than in the baseline OLS specification. As expected, HMP is a better predictor of the Market access index and yields a higher coefficient.

and significant when the dummy Northeast is also excluded.

Finally, we instrument market access by the average date of registration of municipalities in the region²¹ (row 5), which also provides similar coefficient for MA. If we use both HMP and the average date of registration as instruments, to test for over-identification, the Hansen J-test may not be rejected (as the P-value equals 0.229) and the coefficient remains unchanged.

Table 9: Response of wage premium to market access, instrumented

Market access variable:	Estimated coefficient	Robust std. error	Number of observations
MA, instrumented by the distance to Sao Paulo	0,206**	0,015	540
Inter MA, instrumented by the distance to Buenos Aires	0,319**	0,026	540
MA, instrumented by HMP in 1939	0,145**	0,015	540
MA, instrumented by SMP in 1940	0,082**	0,018	540
MA, instrumented by av. date of registration	0,119**	0,024	540
MA, instrumented by HMP and date of registration	0,144**	0,014	540

Notes: 2SLS regressions with standard errors robust to heteroskedasticity and industry fixed effects.

Wage premium regressed on Market Access. Instruments: distance to Sao Paulo or Buenos Aires; HMP: Harris Market Potential = $\text{sum}(\text{GDP}/\text{distance})$ in 1939; SMP: Stewart Market Potential = $\text{sum}(\text{population}/\text{distance})$ in 1940; average date of registration of municipalities within the state

Statistical significance: ** at 1% level.

4.6. By municipalities: local amenities and spillovers

The results found so far are consistent with the NEG explanation to regional wage disparities. However, it is possible that other explanations are in line with these results as well. More specifically, our MA measure could be also capturing short-distance interactions as modeled by the urban economics literature. If that is so, the relation between wages and MA found in this paper would be actually related to the urban economics explanations of wages disparities, rather than the NEG one. Additionally, natural endowments and local attractiveness could just as well play a role in explaining wage premium across regions, and we should check if our results still hold after controlling for these features.

The dataset we used based on individual data allows refining our analysis at the municipality level, which is the finest administrative unit in Brazil (the municipality refers to the location of the establishment). This refinement allows controlling for additional variables related to these alternative explanations.

²¹ We thank an anonymous referee for suggesting this instrument.

First, we estimate the wage premium across municipalities by running a first-step regression of wages on individual characteristics (education and age for males between 25 and 65 years old). The corrected wage premium is obtained by taking the mean of the residual for each municipality.²²

In order to estimate MA by municipalities, we would need to regress gravity equations of the trade flows between them, as specified in equation (9). Since we do not have this data, we use aggregate trade flows across states to estimate the importer fixed effects by state, and the coefficients for distance, language, colonial link, border effects (internal and international) and international contiguity.²³ Our estimation of trade costs involves the coefficients estimated in the gravity equation as well as physical distance between municipalities. If we further assume that price levels are relatively similar within states, we can construct pseudo-importer fixed effects by municipality, by multiplying the state importer fixed effects by the industrial GDP share of the municipality in the state. Formally, market access by municipality is estimated by the following expression:²⁴

$$(15) \quad \hat{MA}_r \equiv \sum_s \left[\left(\frac{GDP_s}{GDP_{S(s)}} \right) (\exp FM_{S(s)}) \prod_k (\exp TC_{k,rs})^{\delta_k} \right].$$

where s refers to municipalities or foreign countries and $S(s)$ stands for the state to which municipality s belong.

Besides MA, interactions between municipalities and spillovers, in particular, are likely to influence wages and its spatial correlation across municipalities. In order to correct for spatial autocorrelation, which induces underestimated standard errors via OLS, we employ the GMM methodology reported by Conley (1999). We specify a cutoff point for spatial interactions of 1.5 degree in latitude or longitude, that is, 100 miles²⁵ (it means that we neglect interactions between cities distant from more than 100 miles). This approach is robust to misspecification

²² This simplified method permits to avoid the estimation of thousands of fixed effects by municipalities, if we had rigorously followed the same methodology we used across states. This may lead to an underestimation of the correlation between wages and MA since we overestimate the effects of age and education. Nevertheless, the estimated coefficients for age and education are very close to the results obtained previously in Table 1.

²³ The estimated gravity equation is similar to the specification of column (1) in Table 1 but excluding the “internal” contiguity variable which is insignificant and has no meaning at the municipality level.

²⁴ We exclude the internal demand from the municipality in the calculation of MA. Excluding large cities from the regression does not affect the results

of the degree of spatial correlation among geographical units and allows obtaining robust standard errors for coefficients estimated through OLS.

Table 10 reports our results, where average wages by municipality are regressed on MA and controls. It confirms the correlation between wages and MA. The first column presents the results when wages are not controlled for skills in the first step, while in the second column the dependent variable is the wage premium corrected by skills. The coefficients obtained are very close to their corresponding ones in the results by state, in the second and third columns of Table 3. The standard error is three times higher than the estimated standard error using traditional OLS across municipalities, which confirms that OLS standard errors are underestimated due to spatial correlations. The corrected standard error is closer the one estimated from the regressions across states.

In the third column, wages are regressed on MA plus state fixed effects in order to capture within-state variations. The estimated coefficient is similar to that obtained from the equivalent regression at the state level (column 3 of Table 3), which is consistent with MA having a similar impact on wages at different geographical scales.

In order to investigate whether short-distance interactions are driving our results, that is, to attempt to disentangle the urban economics and the NEG explanations for the regional wage premium, we control for local interactions. As controls, we use demographic density, the average wage of workers and the proportion of workers with each level of educational attainment (reference is level 5: complete primary education). It is interesting to note that the coefficients for the highest levels of education in the final step are positive (not shown), which suggests positive externalities from workers with tertiary education (since we have already controlled for individual effects). The resulting coefficient for MA, in column (4), is slightly lower than in the first specification but remains large. This result suggests that, while it is true that local interactions play a role in explaining local wages, the NEG approach has also an important part in it.

In column (5) we control for the attractiveness of each municipality. We include in the regression the proportion of new residents in the municipality, using thus recent migration as

²⁵Specifying other cut-off points does not increase the standard error.

an indicator of revealed attractiveness.²⁶ The coefficient of market access remains unchanged. We are aware that recent migration may also be capturing moving decisions driven by differences in market access themselves that induce wage premium across regions. It is not easy to disentangle these two effects though. We refer to Hering and Paillacar (2008) for a study on the relation of market access differentials and migration.

Finally, we add a number of controls for local amenities and endowments (altitude, temperatures, rainfalls, soil quality, land by type of agriculture), and the coefficient for MA is still large and significant, as shown in column (6).

Table 10: By municipalities

Dependent variable: wage premium by municipality

	(1)	(2)	(3)	(4)	(5)	(6)
Market Access	0,162** [0,016]	0,086** [0,013]	0,091** [0,012]	0,107** [0,012]	0,091** [0,010]	0,095** [0,011]
Controls in final step			State fixed effects	Density Av. age Av. age ² Education (% workers by level)	New residents (%)	Erosion type Soil type Temperatures Precipitations Land by type of agriculture
controlling for individual skills (1st step)	no	yes	yes	yes	yes	yes
Spatially corrected SE	yes	yes	yes	yes	yes	yes
R-squared	0,242	0,109	0,301	0,151	0,170	0,255
Observations	3439	3439	3439	3439	3439	3439

Notes: OLS regressions with standard errors corrected for spatial dependence (Conley 1999).
The proportion of new residents is from the Census 2000; endowments are from Timmins (2006).
Statistical significance: ** 1%.

5. Concluding remarks

NEG models predict that migration within a country should offset in great part regional advantages derived from a market and supplier access, leaving wage differences mainly due to individual, industry and firm characteristics. Our results, however, indicate that labor mobility did not arbitrage away all cross regional wages differences. We find a positive and significant impact of market and supplier access on wages, and even stronger to that found for the European regions, despite higher migration levels in Brazil. Menezes-Filho and Muendler (2007) do find evidence of large labor displacements from import competing industries due to

²⁶The proportion of new residents refers to the proportion of males between 25 and 65 years old who have moved from another municipality within the past five years.

the Brazilian trade liberalization in the 1990's. Nevertheless, this does not mean that labor reallocation operated towards the expected direction.²⁷

While it is possible to propose NEG models with migration and spatial wage inequality (see Hanson, 2005), it seems that a more complex phenomenon is happening, that requires taking into account labor market frictions, migration dynamics (in particular taste preferences and spatial variation of skill rewards) and matching between worker heterogeneity and firm heterogeneity. A step in that direction is done in Hering and Paillacar (2008).

Acknowledgements

We thank Andrew Clark, Danilo Coelho, Pierre-Philippe Combes, Matthieu Crozet, Daniel Da Mata, João De Negri, William Foster, Gordon Hanson, Miren Lafourcade, Carolina Lennon, Aguinaldo Maciente, Philippe Martin, Thierry Mayer, Thierry Verdier, two anonymous referees and participants in the FAO Internal Seminar (Santiago), the CIDDES Seminar (Santiago), the XI SMYE Conference (Seville), RIEF Doctoral Meeting (Rennes), Sixth GEP Postgraduate Conference (Nottingham) for excellent suggestions and fruitful discussions. We are grateful to Christopher Timmins for providing data on natural endowments and to IPEA for access to the RAIS and PIA datasets, without which this study would not have been possible.

6. References

- Amiti, M. and L. Cameron (2007). "Economic Geography and Wages," *The Review of Economics and Statistics* 89(1): 15-29, 01.
- Baldwin, R. and T. Okubo (2006). "Agglomeration, Offshoring and Heterogenous Firms," CEPR Discussion Papers 5663.

²⁷ The detailed study on labor adjustment by Menezes-Filho and Muendler (2007) is particularly striking: "Brazil's trade liberalization triggers worker displacements particularly from protected industries, as trade theory predicts and welcomes. But neither comparative-advantage industries nor exporters absorb trade-displaced workers for years" (p. 2).

Barros, R., Henriques, R. and R. Mendonça (2000), “Education and Equitable Economic Development”, *Economia* 1: 111-144.

Barros, R. P. And R. S.P. Mendonça (1995). “Os Determinantes da Desigualdade no Brasil,” Texto para discussão n.377, IPEA.

Chiquiar, D. (2008). “Globalization, regional wage differentials and the Stolper–Samuelson Theorem: Evidence from Mexico,” *Journal of International Economics*. 74(1): 70-93.

Combes, P. and G. Duranton (2006). “Labor Pooling, Labor Poaching, and Spatial Clustering,” *Regional Science and Urban Economics* 36: 1-28.

Combes, P, G. Duranton and L. Gobillon (2008). "Spatial wage disparities; Sorting matters!," *Journal of Urban Economics* 63(2): 723-742.

Conley, T. (1999). “GMM Estimation with Cross Sectional Dependence.” *Journal of Econometrics* 92(1):1-45.

Da Mata, D., U. Deichmann, J. Vernon-Henderson, S. Lall, and H.-G. Wang (2005). “Determinants of City Growth in Brazil,” *Journal of Urban Economics* 62(2): 252-272.

De Sousa, J. and S. Poncet (2007). How are wages set in Beijing. CEPII Working paper 13.

Duarte, A. J, P. C. Ferreira and M. A. Salvato (2004). “Regional or educational disparities? A counterfactual exercise”, *Ensaio Econômicos EPGE* n.532.

Ferreira, P. and J. L. Rossi (2003). “New Evidence from Brazil on Trade Liberalization and Productivity Growth,” *International Economic Review* 44(4): 1383-1405.

Goldberg, P. and N. Pavcnik (2007). “Distributional effects of globalization in developing countries,” *Journal of Economic Literature* 45(1): 39-82.

Gonzaga, G., N.A. Menezes-Filho, and C. Terra (2006). "Trade liberalization and the evolution of skill earnings differentials in Brazil," *Journal of International Economics* 68(2): 345–367

Hanson, G. (2005). "Market potential, increasing returns, and geographic concentration," *Journal of International Economics* 67(1): 1-24.

Head, K. and T. Mayer (2004). "The Empirics of Agglomeration and Trade". In V. Henderson and J.F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Vol 4. pp. 2609-2669. North Holland.

Head, K. and T. Mayer (2006). "Regional wage and employment responses to market potential in the EU," *Regional Science and Urban Economics* 36(5): 573-594.

Hering, L. and R. Paillacar (2008). "On the Relation between market access, migration and wages: An empirical analysis," Mimeo.

Hering, L. and S. Poncet. (2008). Market access impact on individual wages: Evidence from China. Mimeo.

Knaap, T. (2006). Trade, location and wages in the United States. *Regional Science and Urban Economics* 36(5): 595-612.

Lall, S., R. Funderburg, and T. Yepes (2004). Location, concentration, and performance of economic activity in Brazil. World Bank Policy Research Working Paper 3268, World Bank.

Langoni, Carlos G. (1973). *Distribuição da renda e desenvolvimento econômico do Brasil*. Rio de Janeiro: Ed. Expressão e Cultura. 320 p.

Lederman, D., C. Bravo-Ortega, and P. Fajnzylber (2004). The Wage Geography of Brazil after the reforms. Working paper.

Melitz, M. (2003). "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71(6): 1695-1725.

Melitz, M. and G.I.P. Ottaviano (2008). "Market Size, Trade, and Productivity," *Review of Economic Studies* 75(1): 295-316.

Menezes-Filho, N.A. and M. Muendler (2007). "Labor reallocation in response to trade reform," Mimeo.

Menezes-Filho, N.A., M. Muendler and Garey Ramey (2008). "The structure of worker compensation in Brazil, with a comparison to France and the United States," *Review of Economics and Statistics*, 90(2): 347-368.

Mion, G. and P. Naticchioni (2005) "Urbanization Externalities, Market Potential and Spatial Sorting of Skills and Firms", CEPR Discussion Paper 5172.

Mossi, M., P. Aroca, I. Fernandez, and C. Azzoni (2003). Growth dynamics and space in Brazil. *International Regional Science Review* 26 (3): 393-418.

Muendler, M. (2004) "Trade, technology, and Productivity: A study of Brazilian Manufacturers, 1986-1998," CESIFO Working paper 1148.

Paillacar, R. (2007). Market Potential and Worker Heterogeneity as Determinants of Brazilian Wages. Mimeo.

Redding, S. and P. Schott (2003). "Distance, Skill Deepening and Development: Will Peripheral Countries Ever Get Rich?," *Journal of Development Economics* 72(2): 515-41.

Redding, S. and A.J. Venables (2004). "Economic geography and international inequality," *Journal of International Economics* 62(1): 53-82.

Schor, A. (2004). "Heterogeneous productivity response to tariff reduction. Evidence from Brazilian manufacturing firms," *Journal of Development Economics* 75 (2), 373-396.

Timmins, C. (2006). "Estimating spatial differences in the Brazilian cost-of-living with household location choices," *Journal of Development Economics* 80 (1), 59-83

Vasconcelos, J. R. and M. A. Oliveira, (2006). “Análise da Matriz por Atividade Econômica do Comércio Interestadual no Brasil,” IPEA, Texto para Discussão 1159.

A1. Data appendix

Education:

Educational variables are dummies for each grade. Grades are:

Level 1:	ANALFABETO	Illiterate
Level 2:	4.SER INCOMP	Primary School (incomplete)
Level 3:	4.SER COMP	Primary School (complete)
Level 4:	8.SER INCOMP	Middle School (incomplete)
Level 5:	8.SER COMP	Middle School (complete)
Level 6:	2.GR INCOMP	High School (incomplete)
Level 7:	2.GR COMP	High School (complete)
Level 8:	SUP. INCOMP	University (incomplete)
Level 9:	SUP. COMP	University (complete)

Measurement of Productivity:

Data by workers and firms are matched thanks to the firm identification number (CNPJ). The manufacturing survey (PIA after 1996) does not have information on capital, but we could impute the initial capital stock from IBRE data (*Fundação Getúlio Vargas*) for a large subset of firms. For the other firms we estimate the initial capital stock using capital stock data by industry obtained from the old PIA (corrected by the sampling rate in terms of labor) and other firm characteristics from the new PIA database, including investments and the depreciation of the capital stock. However, the large proportion of zero values in investments suggests that missing values may constitute an important issue and that measurement errors may be potentially large.

Productivity is estimated at the firm-level, using variables on sales, labor, capital stock, materials and energy. Assuming a Cobb-Douglas production function and perfect competition, the coefficient for each input should be given by its share in total costs (whose

measure should be independent of the wage level). We take the average of these coefficients across years within sectors.

We constructed alternative measures of productivity using either the residual of OLS regressions or the Levinsohn and Petrin (2003) methodology. Table A1 show how the main results are affected by the choice of the productivity measure. OLS regressions yield very similar results in the intermediary and final step regressions. Levinsohn and Petrin estimations yield lower correlations between productivity and wages but larger correlation between productivity and market access. As a result, the correlation between wages and market access is less affected by controlling for productivity with the OLS and the Levinsohn and Petrin measure.

We focused on the first measure, based on cost shares, because the other two measures are more prone errors. On the one hand, the OLS regressions suffer more from potential simultaneity biases, compared to the first method. On the other hand, the Levinshon and Petrin (2003) estimation appears to be sensible to the quality of the data as the coefficient for capital was very imprecisely estimated²⁸ (standard errors above 0.2). This might explain the low correlation between this measure of productivity and wages.

Table A1: Main results with alternative measures of productivity

Measure of Productivity	Cost share	OLS	LP
Productivity in first step	0.297 [0.002]**	0.227 [0.002]**	0.045 [0.001]**
MA in final step	0.112 [0.011]**	0.116 [0.012]**	0.120 [0.013]**
Productivity regressed on MA	0.022 [0.012]†	0.024 [0.011]*	0.058 [0.020]**

Notes: The first line corresponds to the coefficient of productivity in table 1; the second line corresponds to the coefficient of MA in table 5; the third line corresponds to the coefficient of MA in table 6. Columns correspond to different measures of productivity.

²⁸Our measure of productivity differs from Schor (2004) and Muendler (2004) because they based their analysis on firms present in the previous version of the PIA database (before 1995), which had more precise information on capital stock. Among the 23,933 firms in the current PIA in the year we focus our study, 1999, less than 10% were present in the in the old PIA database.

A2. First step regressions for the instrumental variable approach

Table A2 provide the first step estimations of the 2SLS regressions of table 9. The P-value for the test of excluded instruments is lower than 0.01 for all regressions.

Table A2: First stage regressions corresponding to table 9

Instruments for Market Access:	Coefficient	Std. error	Observations
Distance to Sao Paulo	-1.199**	0.051	540
Distance to Buenos Aires	-1.270**	0.037	540
HMP in 1939	1.707**	0.056	540
SMP in 1940	1.746**	0.078	540
Av. date of registration	-0.027**	0.002	540
{ HMP 1939 & Av. date of registration	{ 1.616** -0.005**	{ 0.064 0.001	540

Notes: First-stage regressions for table 9, with industry fixed effects.

Market Access instrumented by: distance to Sao Paulo or Buenos Aires; HMP: Harris Market Potential = $\text{sum}(\text{GDP}/\text{distance})$ in 1939; SMP: Stewart Market Potential = $\text{sum}(\text{population}/\text{distance})$ in 1940; average date of registration of municipalities within the state

Statistical significance: ** at 1% level.