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

Thibault Fally
PSE – Paris School of Economics

Rodrigo Paillacar
CES – Université de Paris 1
Paris School of Economics

Cristina Terra
Université de Cergy-Pontoise: THEMA
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 access to local, national and international markets separately. We find a strong correlation between market access and wages differentials, even after controlling for individual characteristics, firm productivity, market access source (international, national or local), and using instrumental variables.

1. Introduction

Brazil, the fifth largest country in the world in territory, is also among the most unequal ones. Its inequality is reflected not only at the individual level, but also in its geographic distribution. 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). Additionally, population density and market size vary substantially across regions. Most of the population lives in the coastal areas of the Northeast and the Southeast. While the average density in the Brazilian Southeast is over 150 inhabitants per square-kilometer, this number drops below 4 for the states in the North.

New Economic Geography (NEG) models, by emphasizing the impact of proximity to markets on economic outcomes, provide an interesting framework to study the regional wage inequalities in Brazil. An important relationship proposed by the NEG model is the negative impact of trade costs on firm profits. Trade costs are captured by two structural terms referred to in the literature as "market access" and "supplier access". The first measures access to potential consumers, the latter refers to the access to intermediate inputs, and both are negatively related to trade costs. Since market and supplier access have a positive impact on profits, maximal wages that firms can afford to pay are positively related to these variables.

This paper estimates a structural NEG model to study wage disparities across states and industries in Brazil. We use estimates of market and supplier access to explain regional wages, as in Redding and Venables (2004) and Head and Mayer (2006). We exploit industry level data and we control for individuals' characteristics in our estimations. Thereby, we are able to isolate the impact of location on wage inequality from other sources of wage inequality such as differences in the composition of the labor force or the local diversity of industries.

In two seminal works, Hanson (2005) and Redding and Venables (2004) test structural models of New Economic Geography. The first is applied to US counties and the second to a sampling of countries. Both find a significant impact of trade costs on wages. Inspired by this approach, intra-national studies have been applied to European NUTS regions (Head and Mayer, 2006), US states (Knaap, 2006) and Chinese provinces (Hering and Poncet, 2008).[1]

Our empirical framework brings two noteworthy methodological contributions. First, we control for individual characteristics. The spatial distribution of individuals could be such that their characteristics would be correlated with structural NEG variables, thus leading to spurious results in the estimation of the NEG wage equation.[2] Such control is particularly important in the case of Brazil, since individual diversity is vast and it is an important

[1] All these papers follow the methodology proposed by Redding and Venables (2004), performing a structural estimation of NEG models. There are also empirical studies on NEG using alternative frameworks, such as 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.

determinant of the wage inequalities in the country. For instance, Barros et al. (2000) show that the distribution of education and its return account for about half of the wage inequality from observed sources in Brazil. Additionally, we observe large differences in human capital distribution across regions: workers from Southern regions are on average 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 are 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, Duranton and Gobillon, 2008) or endogenous differences in returns to schooling (Redding and Schott, 2003). In any case, controlling for education allows for the correction for bias induced by the differences in work force composition across regions.

The second methodological contribution is an estimation of market and supplier access using trade flows at the industry level, while other studies use aggregate trade flows.[3] This procedure alleviates the collinearity problem found in the literature when attempts are made to estimate these two variables simultaneously. While it is true that demand and supply should be naturally correlated at the aggregate level, since workers are also consumers, it is less likely to be true at industry level. A region may be specialized in a particular industry production, while using inputs from all industries in general. Hence, by adopting this procedure we are better equipped to disentangle the effects of market and supplier access. As a matter of fact, in the case of Brazil, the distribution of economic activity across regions varies largely across industries. Chemicals, for example, are mainly produced in Bahia, whereas transportation industries are mostly located in São Paulo.

With data on intra- and international trade flows disaggregated at the industry level we are also able to isolate local, national and international market and supplier access. Consequently, we are able to establish which kinds of trade (intra- or international) have the greatest impact on wages through a NEG mechanism.

Our empirical strategy implements a three-step procedure. Firstly, wages are regressed on worker characteristics, controlling for state-industry fixed effects. Secondly, following the NEG literature, we estimate gravity equations in order to calculate market and supplier access for each industry in each state. We compute access to international, national and local markets

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

separately, and measure market and supplier access for each state and industry pair. Finally, market and supplier access derived in the second step are used as explanatory variables for the wage disparities captured by the state-industry fixed effects in the first step.

We find a positive and significant effect of market and supplier access on state-industry wage disparities, with the impact of market access being stronger than the one of supplier access. International market access turns out to have greater impact than national or local market access. The positive impact of market access on wage disparities is robust after controlling for several variables, such as firm productivity, taxes, regulation, endowments, and after using instrumental variables. The results are also unchanged in regressions at the municipality level where we are able to further control for local amenities and endowments.

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, transport costs render the geographic distribution of demand an important determinant of profits. We follow Head and Mayer (2006) and Redding and Venables (2004) and derive profits, market and supplier access from Dixit-Stiglitz preferences. We present a brief description of its main hypothesis and results, rather than a full-fledged model, since such models are now standard in the literature.

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

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:

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

where x_{rsi} represents sales of 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 the total region *s* spending on industry *i*. P_{si} is the price index for industry *i* in region *s*, defined as:

$$(2) \qquad P_{si} \equiv \left[\sum_r n_{ri}\left(p_{ri}\tau_{rsi}\right)^{1-\sigma}\right]^{1/1-\sigma}.$$

As for production costs, we assume that firms use labor and intermediate goods as inputs, 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*, and, for each industry *i*, $\sum_j \varpi_{ji} = 1$. The total price index of intermediate inputs is equal to $\prod_j P_{rj}^{\varpi_{ji}}$.[4] 'Supplier access' of a firm in region *r* and sector *i*, SA_{ri}, is defined as the price index of intermediate inputs, raised to the power $1-\sigma$, as in:

$$(3) \qquad SA_{ri} \equiv \prod_j \left(P_{rj}^{1-\sigma}\right)^{\varpi_{ji}}.$$

It is worth noting that in this paper we adopt a more precise definition of supplier access than the one provided in the NEG literature, by computing supplier access separately for each industry, and taking into account the inter-industry linkages. This procedure helps to disentangle supplier from market access.

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

Given the definition of supplier access, total costs of a firm in region r and industry i may be represented by $SA_{ri}^{\alpha/(1-\sigma)} w_{ri}^{\beta}\left(f_i + \sum_s x_{rsi}\right)$, 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.[5] Supplier access is a measure of the firm's access to intermediate inputs, and it is negatively related to trade costs. The larger the supplier access, the smaller 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 be given by:

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

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

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

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

We are 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:

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

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

Hence, wages are higher in regions with higher MA, that is, with low trade costs to importing 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 in a cross section analysis for 1999.[6] Firstly, wages are regressed on worker characteristics, including state-industry fixed effects. The wage premium captured by these fixed effects is the variable to be explained by market and supplier access. Secondly, in keeping with the new economic geography literature, we estimate gravity equations in order to calculate market and supplier access for each state and industry pair. Finally, market and supplier access derived in the second step are used as explanatory variables for wage disparities captured by state-industry fixed effects from the first step.[7] 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 through their characteristics, such as educational attainment, experience in years, gender, marital status, among many other variables. For Brazil, in particular, Langoni's seminal work (1973) presents evidence of the importance of worker heterogeneity in income inequality. If patterns of diversity among individuals in the labor force were similar across regions, we could still explain average regional wages by regional market and supplier access differences, as proposed in equation (6). Previous empirical work, however, has identified substantial differences in the composition of the labor force across Brazilian regions, especially with respect to educational attainment (see Duarte et al., 2004). Thus, our results would be biased if we did not to control

[6] We limit our analysis to 1999 due to the lack of intra national trade data for other periods in Brazil, as explained in section 3.

[7] We thank an anonymous referee for suggesting an empirical procedure where fixed effects from the wage equation are regressed on market and supplier access.

for individual characteristics and sorting across regions and sectors. The first step of our empirical study consists in estimating the following equation:

$$(7) \qquad \log w_{l,ri} = \lambda_1 age_{l,ri} + \lambda_2 age_{l,ri}^2 + \sum_{m=1}^{9} \mu_m ed_{l,ri}^m + \omega_{ri} + \xi_{l,ri} ,$$

where $w_{l,ri}$ is the wage of a male[8] worker l working in industry *i*, of region *r*, $age_{l,ri}$ is the worker's age, $ed_{l,ri}^m$ is a dummy variable for each of the 9 educational levels (see Appendix A1), and ω_{ri} are dummy variables for each state-industry pair.[9]

State-industry fixed effects capture wage disparities that are not explained by worker characteristics, and that is the variable that will be used to be explained through state-industry market and supplier access.

Second step

The second step consists in estimating MA and SA as follows. Total sales from region *r* to region *s* in industry *i*, from equation (1), can be written as:

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

The first term in right hand side equation (8) comprises variables related to the exporting region, while the third term involves variables exclusively from the importing region. Hence, these two terms are captured empirically by exporting and importing region fixed effects, FX_{ri} and FM_{si}, respectively. As for the second term, there is no single variable to seize trade costs between two regions. Trade costs will then be captured by a set of variables, $TC_{k,rs}$, such as the distance between the regions (in log), whether they share borders, a

8 As in keeping with most of the labor literature, we focus on male workers between the ages of 25 and 65, because the wage dynamics and labor supply of the female work force are often affected by non-economic factors, such as fertility decisions.

9 In section 4.4, on robustness checks, this equation will be estimated adding in productivity as explanatory variable. In that case, the regression will incorporate the firm dimension.

language, or whether they have a colonial link.[10] In sum, equation (8) is estimated though a gravity equation as follows:

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

where X_{rsi} stands for exports from region *r* to region *s* in industry *i*, and ε_{rsi} is an error term. A region may be defined as either a Brazilian state or one of the 210 countries in our dataset.

In order to render our results comparable to those in the literature, we have also estimated equation (9) for aggregate trade flows, instead of disaggregating by industry. In this way we can compute MA and SA measures comparable to the ones in Redding and Venables (2004), Knaap (2006), Head and Mayer (2006) and Hering and Poncet (2008).

We would like to note that an estimation based on gravity regressions has the advantage of using information related to the economic mechanism that our theoretical model intends to stress, namely, spatial interactions arising from trade. We would thus be less prone to capture other effects of proximity, such as technological or urban externalities. Nevertheless, we will perform several robustness checks to investigate a potential correlation between the trade channel and other covariates and competing explanations.

Despite their empirical success in explaining trade flows, gravity equations have an important caveat: they treat the size of regions as exogenous (Knaap, 2006). We acknowledge this limitation in explaining the long-term evolution of a country's economic geography, and we see our work as an effort to uncover the impact of market access on wages, taking the spatial distribution of economic activity as given.

From equation (5), the estimated coefficients in equation (9) can be used to compute market access as in:

[10] It is possible to choose a number of alternative sets of variables, but changing specifications of the gravity equation yields little change in results of the final-step. Similar results are obtained for example, 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 distance coefficients for intra versus international trade. Finally, Paillacar (2007) shows that Gamma PML yields similar results to OLS.

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

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

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 vertical linkages across industries, we use coefficients from the input-output matrix, $\hat{\varpi}_{ji}$, to weigh the impact of each industry in supply access. We, then, compute:

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

which yields a SA measure for each industry in each Brazilian state.

This paper is the first to weigh industry supplier access through 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 values for MA and SA estimated in the second step are used to explain differences in wage across states and industries. The equation for wages (6) can be written as:

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

As previously discussed in the beginning of this section, differences in human capital allocation across regions may distort the impact of market and supplier access on regional wages, and previous empirical studies suggest this to be a relevant issue for Brazil. Therefore, instead of adopting wages as a dependent variable, we use the state-industry fixed effects estimated in equation (7). They stand for the wage differentials across states and industries that are not explained by age and education, thus controlled for composition of labor force with respect to these variables. We estimate the equation as follows:

(13) $$\log \hat{\omega}_{ri} = \theta_0 + \theta_1 \log \hat{M}A_{ri} + \theta_2 \log \hat{S}A_{ri} + \theta_3 D_i + \zeta_{ri}\ .$$

where D_i represents industry dummies, $\hat{\omega}_{ri}$ are the state-industry fixed effects estimated in the wage regression (7), and ζ_{ri} is an error term.[11]

This three-step method involves the use of estimated values for the variables used in the NEG wage equation. Regarding the use of an estimated dependent variable, ω_{ri}, if an additive error term is assumed, ζ_{ri} will contain part of the variance of the error term in equation (7), and it could be heteroscedastic due to sampling error (Lewis and Linzer, 2005). This has led some researchers to use weighted least squares (WLS), with the inverse of the standard error of the wage premium estimates from the first stage as weights (see, for example, Pavcnik et al, 2004). Nevertheless, Monte Carlo experiments by Lewis and Linzer (2005) suggest that implementing a WLS can only surpass White standard error estimates in efficiency when a very high proportion (80% or more) of the residual in the final regression is due to sampling error on the dependent variable. Moreover, they found that WLS can also generate highly misleading standard error estimates if the contribution of the error term in the first stage is low. In our case, we have a very high number of individual observations, which result in very precise estimations of the state-industry fixed effects. Consequently, we chose to report regressions with robust standard errors.

A second issue is the use of MA and SA estimates from trade equations as independent variables, which means that trade equation residuals also affect ζ_{ri}. As Head and Mayer (2006) point out, this also invalidates standard errors, but it has no impact on the estimated coefficient. In this case, several researchers (Redding and Venables, 2004; Hering and Poncet, 2008) have implemented bootstrap to obtain unbiased confidence intervals in order to make inference. We, therefore, also computed bootstrapped standard errors.

[11] Combes et al (2008) employ similar methodology, but they estimate location and industry fixed effects separately, due to computational problems and insufficient data (they have 341 locations and 99 industries, see p. 727, Footnote 7). Our aggregation level, with 27 Brazilian States and 22 industries, precludes such problems. The only exception being in section 4.5, where we adopt municipalities and not states as regional units. For 3439 municipalities (instead of 27 Brazilian States), we only consider the spatial dimension.

3. Data

In this paper we use three sets of data: on individual characteristics, trade flows and country characteristics. We perform a cross sectional analysis for 1999, since intra-national trade data by industry for Brazil is only available for that year (Vasconcelos and Oliveira, 2006).

Individual characteristics are drawn from the RAIS database (*Relação Anual das Informações Sociais* issued by the Brazilian Labor Ministry) that covers all workers in the formal sector.[12] We focus on the manufacturing sector for compatibility with the trade data. When more than one job is recorded for the same individual, we select the highest paying one.[13] The database provides several individual characteristics (wages, educational level, age, gender, etc.), as well as worker and firm identification numbers which allows us to match the RAIS database with the manufacturing survey.

The manufacturing survey, PIA (*Pesquisa Industrial Anual* from IBGE, *Instituto Brasileiro de Geografia e Estatística*), includes all firms with thirty employees or more from 1996 to 2003, covering the majority of the workforce in the manufacturing sector. This dataset provides a large range of variables on production, which includes sales, labor, materials, energy and investments, which allows for measuring productivity (see Appendix A2). We complete the PIA with IBRE-FGV (*Instituto Brasileiro de Economia – Fundação Getulio Vargas*) balance sheet data from 1995, from which we draw initial fixed capital, and with patent data from INPI (*Instituto Nacional da Propriedade Industrial*). All datasets can be matched thanks to firm identification numbers.

In order to estimate the gravity equation we need three sets of trade data: (1) trade data among Brazilian states, which is drawn from Vasconcelos and Oliveira (2006) who processed value-added tax information provided by the National Council of Financial Policy (CONFAZ,

[12] Because of the very large number of observations, we ran our regressions on random samples of 500,000 or 800,000 employees (out of 2,786,852 employees in the full sample). Changing the size of the sample does not affect our coefficients nor does it particularly affect the estimation of state-industry fixed effects. Table A1 provides summary statistics of individual characteristics.

[13] For example, a worker may change occupation or place of work over time, or may even hold two recorded jobs at the same time. To assess the robustness of our results, we alternatively chose the average wage, the total wage, in December or over the year, and the choice did not affect the results.

Conselho Nacional de Politica Fazendaria) from the Ministry of Finance (*Ministerio da Fazenda*); (2) trade data between Brazilian states and foreign countries, from *Secretaria de Comércio Exterior*, Ministry of Trade; and (3) among foreign economies, from BACI: *Base pour l'Analyse du Commerce International*, CEPII. Moreover, using total sales by region and industry from the PIA database allows us to compute internal flows within state 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 (which corresponds to the Brazilian CNAE 2 digit industry classification).

We complete the trade and individual information with additional data on geography, infrastructure 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 is measured in 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 for internal border equals zero if the trade is within one Brazilian state, and it equals 1 otherwise. In addition, we construct a dummy for international contiguity that equals 1 if the dummy for international border equals 1 and if both countries (or the country and the Brazilian state) share a border. Analogously, the dummy for 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 precisely, if the official language is the same or if the same language is used by at least 20% of the population). Lastly, the colonial link dummy equals 1 if the trade is between two different countries and one of them has been colonized by the other in the past.

The Census 2000 (IBGE) provides us data on migration rates per municipality. The input-output matrix is constructed by OECD and IBGE, across ISIC Rev3 2-digit industries. The cost of starting 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. The data on harvested agricultural area in 1996 is from the Agricultural Census. The *Anuário Mineral Brasileiro 1999* (table 8 page 51) was our source for regional shares of mineral production. Municipality data on natural endowments is from Timmins (2006).

4. Results

We organize the results in five subsections. We start, in subsection 4.1, by presenting the results of the first and second steps of our empirical procedure, that is, the estimation of the state-industry wage differentials and of the market and supplier access through gravity equations. Subsection 4.2 presents the results of the regressions of MA and wage differentials, while 4.3 incorporates SA to the analysis. Robustness checks are made in subsection 4.4. Finally, subsection 4.5 replicates the MA results using municipality, rather than state, as the region unity in Brazil.

4.1. Preliminary regressions

First step: wage premium

The first step of the empirical procedure consists of estimating wage differentials across states and industries that are not driven by individual characteristics. We regress wages on education attainment, experience, and on state-industry fixed effects, as described in equation (7). We use individual data for male workers between the ages of 25 and 65. This group of workers was chosen to render the sample more homogeneous, thus eliminating possible effects from differences in variables such as early school dropouts and female participation. We measured education by dummy variables for nine levels of education (as described in Appendix A1). Age and age squared are used as proxies for experience. Table 1 presents the results.

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

Table 1: Wages and individual characteristics

Dependent variable: Wages	
	(1)
Age	0,072**
	[0,001]
Age squared / 100	-0,072**
	[0,001]
Education (level 5 = 0):	
level 1	-0,365**
	[0,005]
level 2	-0,239**
	[0,003]
level 3	-0,149**
	[0,002]
level 4	-0,075**
	[0,002]
level 6	0,156**
	[0,003]
level 7	0,419**
	[0,002]
level 8	0,852**
	[0,004]
level 9	1,240**
	[0,003]
State x Industry FE	yes
R-squared	0,880
Observations	798494

Notes: OLS regressions with robust standard errors.
Statistical significance: ** 1% and * 5% level.

Second step: market and supplier access

In order to compute estimated values of market and supplier access, we start by estimating the gravity equation (9), where bilateral trade flows are explained by exporter and importer fixed effects, and a set of variables capturing trade costs. We define each Brazilian state as a region, and follow two procedures. In the first one, we take coefficients to be the same for all industries, in keeping with the literature, and use them to compute aggregate measures of MA and SA. In the second procedure a regression is run separately for each industry, estimating, thus, different coefficients for each of them. We are thereby able to compute market and supplier access measures for each state-industry pair.

The first column of Table 2 present regression coefficients using aggregate trade flows, with the corresponding standard errors in the second column. The next three columns of the table shows some summary statistics for the 22 regressions by industry: average values of the estimated coefficient across industries (third column), average values of the standard errors of each regression in square brackets below each coefficient (fourth column), and the standard deviation of the 22 coefficients in parentheses (fifth column). Standard deviations of the coefficients are generally larger than average standard errors which indicate notable differences in transport costs coefficients across industries.

Table 2: Gravity equations

Dependent variable: trade flows

	Aggregated		By industry		
Statistics	Coefficient	Standard error	Average of coefficients	Average of standard errors	Standard dev. of coefficients
Physical distance	-1.448**	[0.018]	-1.359	[0.031]	(0.180)
International border	-4.326**	[0.116]	-4.534	[0.563]	(0.983)
International contiguity	1.001**	[0.095]	0.785	[0.249]	(0.184)
Internal border	-2.594**	[0.386]	-3.212	[0.224]	(0.968)
Internal contiguity	0.128	[0.225]	0.205	[0.118]	(0.469)
Language	0.839**	[0.043]	0.604	[0.071]	(0.263)
Colonial link	0.832**	[0.100]	0.903	[0.115]	(0.140)
Exporter FE	yes			yes	
Importer FE	yes			yes	
Industries				22 regressions	
R-squared	0.982				
Observations	25315			total: 246833	

Notes: OLS regressions with robust standard errors. Dependent variable: trade flows (aggregated or by industry). Statistical significance in the first column: ** 1% and * 5% level.

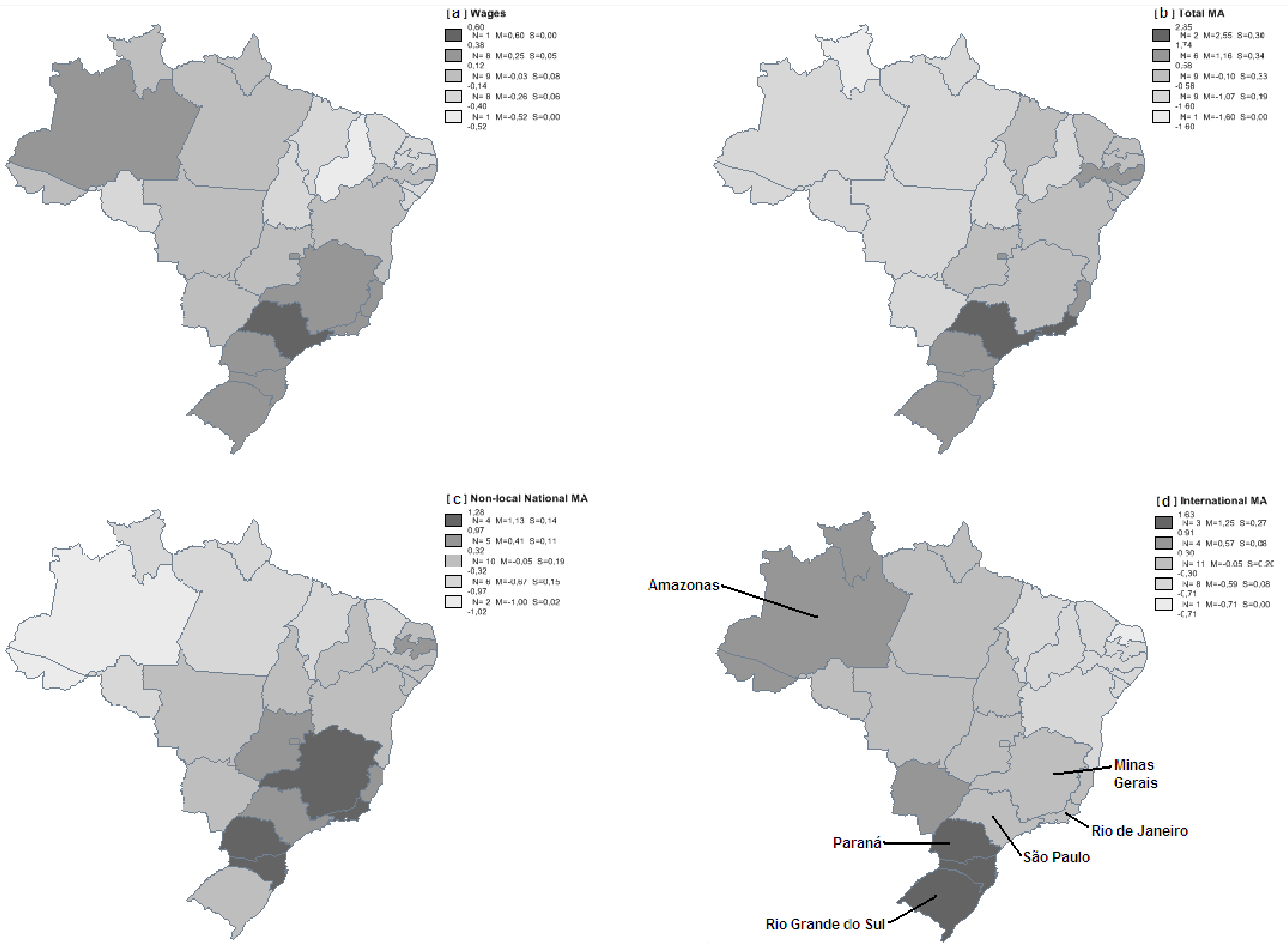
With estimated coefficients from equation (9) (presented in Table 2), we used equation (10) and (11) to compute estimated values for MA and SA, respectively, for each state-industry pair. Note that, when calculating MA, we sum over all states and countries with which a particular state trades. It is then possible to construct a market access measure from a subgroup of trade partners, and that is exactly what we do to investigate the varying impact of local, national and international market access.

Before proceeding 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 show the spatial distribution of wages and MA. Values are normalized as deviations from the mean across regions, and they are grouped in five classes. The middle group falls between the mean +/- 0.5 standard deviations, and the subsequent groups are delimited by 1 standard deviation.

Panel (a) presents 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 that offers 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 more sparsely populated north, also shows high wages. We expect these differentials to reflect exogenous regional characteristics, such as amenities and the availability of natural resources, as well as spatial externalities, such as knowledge spillovers and market access, among others.

Panel (b) displays the total MA across regions. If we look at Figure 1, we are given the impression 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 understanding of the factors at work is obtained by decomposing MA into its national and international scopes.

Figure 1. Market Access and Wages across Brazilian States.



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 is smaller. More interestingly, this exercise shows that Amazonas and Rio Grande do Sul (the southern region closest to Argentina) are remote from the main sources of demand within the country.

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

4.2. Wages and market access

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

The MA coefficient we find in our baseline regression is larger than the one estimated by Head and Mayer (2006) for Europe, and smaller than in Redding and Venables (2004) in a cross country study. Our procedure differs in a few aspects from the procedures adopted in those two papers. For one, we control wage premiums for individual characteristics. Since Redding and Venables (2004) do not control wages for any individual variables as we do, their estimated coefficient may 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

[14] Since the predicted values for market access and wage premiums are generated from prior regressions, we checked our results for sensibility 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.

Hering and Poncet (2008), which also controls for wage disparities per individual characteristics in a study of 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. This higher coefficient may capture part of the impact of spatial sorting of human capital.

Another difference between our work and that of Redding and Venables (2004) is that we construct market access 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 of approximately 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.

In our baseline regression, 35% of wage disparities across regions and industries is explained by MA and industry dummies. If we employ industry dummies exclusively, instead of MA, only 17.5% of wages differentials are explained (regression not reported): The explanatory power of the regression increases substantially with the inclusion of MA.

Table 3: Response of wage premium to market access

Dependent variable: wage premium

	(1)	(2)	(3)	(4)	(5)	(6)
Measure of MA	Total MA	Total MA	Total MA (aggregate)	Non-local MA	National MA	International MA
Market Access	0,140**	0,168**	0,079**	0,185**	0,162**	0,228**
	[0,012]	[0,013]	[0,026]	[0,022]	[0,021]	[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.

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

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

This interesting result may be explained by the trade liberalization that took place in the early 1990's. Trade barriers were lowered during the first half of that decade, and it may 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 in regions with larger international MA. It is possible that in the late 1990's, our period of study, labor mobility across regions had not yet been sufficient to arbitrage the differential impact on wages of the opening to trade. In a study of 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 due to the effect of demand, SA's impact on wages is associated to lower costs and higher productivity.

A common problem with MA and SA measures is that they tend to be strongly correlated. To address this issue, Redding and Venables (2004) incorporate additional assumptions on the link between MA and SA. In 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. Take for instance, a hypothetical industry N, whose output is consumed by all consumers or by all industries in general, while its inputs come from one particular industry Y. In this case there would be high supply access associated to regions with high production in industry Y, while its MA level would be independent of the production composition of the importing region. In practice, we found a high partial correlation between MA and SA (0.76), although lower than the correlation reported by Redding and Venables (0.88). Still, the correlation is apparently sufficiently low so as to allow for the inclusion of both variables in a single regression, without experiencing multicolinearity problems.

We follow the same three-step procedure adopted for the MA regressions. In the first step we regress wages on individual characteristics and on state-industry dummies. Secondly, we compute the SA measure following equation (11), as described in section 2.2. Finally, we use SA as an explanatory variable of the state-industry wage disparities estimated in the first step. The results are presented in Table 4.[15]

Table 4: Market Access and Supplier Access

Dependent variable: wage premium

	(1)	(2)	(3)	(4)	(5)
Supplier Access	0,140**		0,066**		
	[0,010]		[0,017]		
Non local SA		0,193**		-0,019	
		[0,023]		[0,078]	
SA (excl. own ind)					0,040*
					[0,017]
Market Access			0,108**		0,135**
			[0,021]		[0,019]
Non local MA				0,238**	
				[0,079]	
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 aregressions 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; "excl. own ind": supplier access excluding own industry in the input-output matrix.
Statistical significance: ** 1% and * 5% level.

[15] Since we need exporter fixed effects for industry 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 an explanatory variable (regression shown in the first column), the estimated coefficient has the same value as the one found for MA in our baseline regression as shown in the first column of Table 3. 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 impact on regional wages independently, 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. To be more precise, beta estimates of 0.242 for SA and of 0.467 for MA provides further evidence that MA is more important than SA in explaining wage differentials across states and industries in Brazil.

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

One concern with the SA measure is that it may be correlated to its own industry characteristics, such as productivity. To account for this possibility, we employ a SA that excludes its own industry, that is, we set to zero the input-output matrix coefficient for own industry and normalize the other coefficients so as to obtain a sum equal to one. When this proxy 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 difference between them increases.[16]

4.4. Robustness checks

Controlling for productivity and technology

We investigate whether the positive correlation found in the previous sections between MA and regional wage differential could be related to differences in productivity and technology.

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

[16] We also employed this measure as instrument for the SA, finding similar results. Coefficients for MA and SA are respectively, 0.124 and 0.049. MA is significant at 1%, while SA is significant at 5%.

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, thus, possible that the impact of MA on wage differentials is due to its impact on productivity, rather than through the NEG labor demand channel. Hence, in the same way we have controlled for the individual characteristics of laborers, we also control for productivity when estimating wage differentials in the first step. Our dataset allows for this control since we are able to match the data on workers with data on firms with more than 30 employees. In particular, data obtained at firm-level provide information on labor, wages, investment, capital, materials and energy (see details in Appendix A2).

Total factor productivity is measured in a simple way, using a cost share approach (see Foster, Haltiwager and Syverson, 2008, and Syverson, 2004, for similar measures of productivity using US data). Our index of productivity θ_{ih} for the firm *h*, in industry *i*, is defined by:[17]

$$\text{(14)} \qquad \log\theta_{ih} = \log Y_{ih} - sh_{Li}\log L_{ih} - sh_{Ki}\log K_{ih} - sh_{Ei}\log E_{ih} - sh_{Mi}\log M_{ih},$$

where Y refers to revenues, L, K, E and M refer to labor, capital, energy and materials, respectively, and sh_{zi} denotes the share of input Z in annual costs for firms in industry *i*, taken as the average of the period between 1996 and 2003 across all firms in the industry, for $Z = L, K, E, M$. Total costs equal the cost of labor (wages), capital (investments), energy (electricity, fuel and gas expenses) and materials (material expenses).

In addition to its simplicity, this methodology is very robust to measurement errors and misspecifications compared to alternative methods (Van Biesebroeck, 2007). Moreover, we find that our coefficient for MA is robust to alternative measures of productivity (see Appendix A2).

Table 5 presents the results of the impact of MA on state-industry wage differentials, controlling for productivity. The first column of Table 5 is the equivalent to the first column

---

[17] This formula can be derived from the optimization of a Cobb-Douglas production function with constant return to scale. Note that our results are not sensible to small changes in return to scale (multiplying the coefficients sh_{Zi} by the same factor, between 0.90 and 1.10). Moreover, we should note that our measure is robust to differences in wages across regions. The share of labor in total cost remains constant across firms in the same industry as long as the coefficient in the Cobb-Douglas production function is constant.

of Table 3, but using a version of the wage regression in the first step (equation (7)) that also controls for firm productivity. The wages corrected for productivity are 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): When we control for productivity in the first step, the MA coefficient becomes smaller in the third step. Additionally, the R-square of the regressions controlling for firm productivity is larger than in the regressions without such control.

Note that, in the first step, we find a positive and significant elasticity of wages to productivity, close to 0.3, as shown in the second part of Table 5. Alternatively to controlling for productivity in the first step, we can control for it directly in the third step, taking industrial and regional averages. The estimated effect of MA remains similar (results not reported).

Table 5: Controlling for productivity and technology

Dependent variable: Wage premium

	(1)	(2)	(3)	(4)
Market Access	0,112**			0,126**
	[0,011]			[0,010]
National, Non local MA		0,134**		
		[0,018]		
International MA			0,201**	
			[0,017]	
R-squared	0,403	0,328	0,388	0,343
Observations	466	466	466	534
First step regression:				
Firm productivity		0,297**		
		[0,002]		
Innovative firm				0,259**
				[0,002]
Patent stock				0,044**
				[0,001]
Controlling for skills		yes		yes
State x industry FE		yes		yes
R-squared in first step		0,899		0,886
Observations in first step		499144		499878

Notes: OLS regressions with robust standard errors and industry fixed effects.
See section 2.2 and data appendix for the measure of productivity and patent stock.
Statistical significance: ** 1% and * 5% level.

Productivity is measured using revenues since we do not have data on quantities and prices. One problem is that prices vary endogenously across regions depending on market access and competition. Thus, we use data on patents in order to control for technology in a way that is not affected by price levels. The data made available by INPI (*Instituto Nacional da Propriedade Industrial*) list all patents recorded in the 1990's. Our first variable is a dummy for innovative firms, which equals one when there is at least one patent recorded for a given firm. The second variable is the count of patents, in log.[18]

The first-stage regression in the fourth column of Table 5 shows that wages are strongly correlated with both variables on patents. Since the use of patents (access to technology) may also be correlated with market access, controlling for patents affects the coefficient for market access. The impact, however, is small: the coefficient of MA in the regression of the fourth column of Table 5 is not very different from that of the first column of Table 3. We find the same result if we use aggregate data on patents across states and industry. Hence, our results on MA do not seem to be driven by differences in technology across firms.

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 us to control for education in the first step independently of the final step regression. Theoretical papers have shown however, 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 6, 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 (workers who 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 for all workers). In column (2), the wage premium is constructed using data on unskilled workers only (workers who have not completed high school), and the coefficient obtained for MA is close to the baseline regression in Table 3. Hence, the impact of market access on

[18] The number of patents is discounted at a 15% yearly rate, but results are not sensible to moderate changes in the discount rate. This variable is normalized to zero for non-innovative firms.

wages seems stronger on skilled workers, but the difference is not significant.[19] 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 state-industry is significantly larger among unskilled than skilled workers (results in columns 3 and 4 of Table 6).

This result means that higher international MA increases relatively more the wages of unskilled workers. Given that our study corresponds to a period of just a few years after a massive trade liberalization program, this result could actually be a sign that the Stolper-Samuelson mechanism is at work. According to this mechanism, trade liberalization in Brazil, a country where unskilled labor is relatively abundant,[20] should increase relative returns to that factor of production. When viewed through the perspective of economic geography, such an 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) that present evidence for Brazil of relative wage changes compatible with Stolper-Samuelson predictions.

Table 6: 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**	0,134**		
	[0,014]	[0,011]		
International MA			0,196**	0,229**
			[0,023]	[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

[19] Alternatively, we have also directly regressed the skill premium on MA (results not reported). The coefficient has the expected sign but it is not significant.

[20] Muriel and Terra (2009) present evidence that Brazil is relatively abundant in unskilled labor.

Whereas public infrastructure is generally better in the South, successive governments have tried to improve infrastructure 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 7, column (1), we regress the wage premium on market access and tax rates estimated at 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 them only marginally affects the coefficient for market access.

In column (2), we control for firm entry costs, from the World Bank database (Doing Business in Brazil) that provides an estimate of the cost of starting a business across 13 states in Brazil. The MA coefficient is still significant but lower, which means that part of the correlation between MA and wages might actually be explained by a negative correlation between entry costs and MA.[21] 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 for São Paulo and Amazonas dummies

São Paulo is the economic center of Brazil and attracts a large fraction of the Brazilian manufacturing industry. It is also located in a central position between Rio de Janeiro and Buenos Aires. 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 estimated the response of wages to MA including a dummy for São Paulo. The MA coefficient in column (3) of Table 7 is still large and highly significant, albeit smaller. Thus, our main result is not driven by the state of São Paulo. Additionally, the positive and significant estimated value for the São Paulo dummy shows a larger wage premium for that state in addition to the impact of its higher market access.

Amazonas is also an interesting case: the state presents high wages and low aggregate market access. Its large wage premium is possibly associated to the state’s high international market

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

access, as shown in the Panel (d) of Figure 1. Amazonas is close to emerging economies in South America (Colombia and Venezuela) and to NAFTA members. More importantly, Manaus, the capital of the state of Amazonas, is a Free Trade Zone, with large import and export volumes, explained by industries linked to assembling. In any case, to isolate potential measurement errors or an outlier effect, we run regressions with a dummy controlling for the state of Amazonas and we still found a positive and significant impact of MA on wages (regressions not reported). Furthermore, the international MA retains its high elasticity (0.22, in Column 6 of Table 3) after controlling for this dummy.

Controlling for endowments

Endowments are unequally distributed across Brazilian states and play an important role in explaining wage 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 was not affected when we restricted 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 7, we control for minerals, harvested land area, access to the sea and dummies for macro regions.[22] As expected, wages are positively correlated to 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 partially explained by its harsh climate (e.g. frequent droughts).[23] In spite of the inclusion of these controls, the coefficient for MA remains 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 4.5 as these variables are available by municipality.

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

[23] When we include exclusively the Northeast dummy in the regression, other coefficients remain similar (the effect of harvested land being more significant). Interestingly, the coefficient for "landlocked" becomes positive and significant when the Northeast dummy is also excluded.

Table 7: Further robustness checks

Dependent variable: wage premium

	(1)	(2)	(3)	(4)
Market Access	0,145**	0,099**	0,120**	0,116**
	[0,011]	[0,013]	[0,013]	[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 taxe/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 market access index if productivity also impacts the demand for goods. If we consider "non local" market access constructed through excluding the demand from its own market, as we have already done in column (4) of Table 3, such biases are mitigated. In this section, we instrument the MA index by geographical and demographic variables that should impact market access but not directly affect the wage differentials across regions.

We propose two alternative instruments:[24] Firstly, we consider a "Harris Market Potential" (HMP, sum of other regions' GDP divided by the distance) constructed using GDP by states in 1939:

[24] We also tried the distance to the main economic centers as instruments, as proposed by Redding and Venables (2004). In particular, we estimated regressions using the distance to São Paulo and the distance to

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

This variable was first used in NEG empirical studies of the new economic geography literature by Hanson (2005), in his working paper version of 1998. 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 technological innovation). As shown in the first row of Table 8,[25] this instrument yields a significant and strong coefficient for MA, which is nevertheless smaller than in the baseline OLS specification. We also consider a second instrument, which employs population size (in 1940) instead of GDP in equation (15). It provides similar results to the HMP (result not reported here).

We also instrument market access by average registration dates of municipalities in the region[26] (second row of Table 8), and we still obtain a similar coefficient for MA. If we use both HMP and the average registration date 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 8: Response of wage premium to market access, instrumented

Market access variable:	Estimated coefficient	Robust std. error	Hansen J-test	Number of observations
MA, instrumented by HMP in 1939	0,145**	0,015	\	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	0,229	540

Notes: 2SLS regressions with standard errors robust to heteroskedasticity and industry fixed effects.
Wage premium regressed on Market Access. Instruments: HMP: Harris Market Potential = sum(GDP/distance) in 1939; average date of registration of municipalities within the state. Statistical significance: ** at 1% level.

Buenos Aires as instruments of national and international MA, respectively. Although the coefficients for MA in the wage equations support our results (0.20 for national MA, 0.32 for international MA, both highly significant), we share the concerns raised by Head and Mayer (2006): Economic centers may be themselves endogenous. More specifically, the distance to São Paulo may capture effects that are not related to MA, such as the proximity to firm headquarters, and therefore managerial power, which may have a positive impact on wages. Regressions are available on request and in a previous working paper (Fally et al, 2008).

25 Appendix A3 presents the first stage regressions corresponding to Table 8.

26 We thank an anonymous referee for suggesting this instrument.

4.5. Per municipality: Local amenities and spillovers

The results found so far are consistent with the NEG explanation for 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 the case, the relation between wages and MA found in this paper would actually reflect explanations for wage disparities from urban economics, rather than the explanations proposed by NEG. Additionally, natural endowments and local attractiveness could just as well play a role in explaining wage premiums across regions, and we should check if our results still hold after controlling for these features.

Our dataset on the individual workers allows us to refine our analysis at the municipal level, which is the finest administrative unit in Brazil (the municipality refers to the location of the firm). This refinement allows us to control for additional variables related to these alternative explanations.

Firstly, we estimate wage premiums across municipalities by running a first-step regression of wages on individual characteristics (education and age for males between 25 and 65 years of age). The corrected wage premium is obtained by taking the mean of the residual for each municipality.[27]

In order to estimate MA per municipality, 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 per state, and the coefficients for distance, language, colonial link, border effects (internal and international) and international contiguity.[28] Our estimation of trade costs involves the coefficients estimated in the gravity equation as well as the physical distance between municipalities. If we further assume that price levels are relatively similar within states, we can construct

[27] 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.

[28] 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.

pseudo-importer fixed effects per municipality, by multiplying state importer fixed effects by the industrial GDP share of the municipality in the state. Formally, market access per municipality is computed as follows: [29]

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

where s refers to municipality or foreign country and $S(s)$ stands for the state to which municipality s belongs in the first case or the foreign country itself in the latter.

Besides MA, other factors are likely to influence wages and their spatial correlation across municipalities, such as, in particular, the interactions between municipalities and spillovers. 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 at 1.5 degree in latitude or longitude, that is, 100 miles. It means that we neglect interactions between cities at a distance greater than 100 miles. Specifying other cut-off points does not increase the standard error. This approach is robust to misspecification of the degree of spatial correlation among geographical units and allows us to obtain robust standard errors for coefficients estimated through OLS.

Our results are reported in Table 9, where average wages by municipality are regressed on MA and controls. These results confirm the correlation between wages and MA. The first column shows 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 worker skills. This last result is very close to the corresponding one in regressions by state, presented in the third column of Table 3. It should also be noted that the standard error corrected for spatial autocorrelation is three times higher than the one estimated using traditional OLS across municipalities, which confirms that OLS standard errors are underestimated. The corrected standard error is closer to the one estimated in the regressions across states.

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

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

In order to investigate whether short-distance interactions are driving our results, that is, in an attempt to disentangle the urban economics and the NEG explanations for the regional wage premium, we control for local interactions. We use demographic density, the average age of workers and the proportion of workers at each level of educational attainment (our reference is level 5: Complete primary education) as our controls.

It is interesting to note that the coefficients for the highest levels of education in the final step are positive (not shown), even after controlling for this variable in the first step. This result suggests an additional impact of education on average wages, besides the one arising from its spatial composition of the labor force. A possible explanation could be the existence of positive externalities for workers with tertiary education. The resulting coefficient for MA in column (4) is slightly lower than in the first specification in column (1), but it remains significant. 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.

We control for the attractiveness of each municipality in column (5). A variety of regional amenities may influence the decisions of individuals to establish their location, and, ultimately, may be reflected in compensating wage differentials. Since the role of amenities is not easy to assess and constitutes a research topic in itself, we do not pretend to fully investigate it. In this article we simply use recent migration as a (raw) indicator of revealed attractiveness in the wage regression. In our regression we include the proportion of new residents in the municipality,[30] and the coefficient of market access remains unchanged. We are aware that recent migration may also capture decisions to change residence driven by differences in market access themselves that induce wage premium across regions. Nevertheless, it is not easy to disentangle these two effects. We refer to Hering and Paillacar (2008) for a study on the relation of market access differentials and migration.

[30] The 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.

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 9: By municipalities

Dependent variable: wage premium by municipality						
	(1)	(2)	(3)	(4)	(5)	(6)
Market Access	0,162**	0,086**	0,091**	0,107**	0,091**	0,095**
	[0,016]	[0,013]	[0,012]	[0,012]	[0,010]	[0,011]
Controls in final step			State fixed effects	Density Av. age Av. age² 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 market and supplier access, attributing wage differences mainly to individual, industry and firm characteristics. Our results, however, indicate that labor mobility did not arbitrage away all cross-regional wages differences in Brazil. We find that market and supplier access have a positive and significant impact on wages, and even stronger than has been found for the European regions, despite higher migration levels in Brazil. Menezes-Filho and Muendler (2007) do find evidence of large labor displacements out of import competing industries due to the Brazilian trade liberalization in the 1990's. Nevertheless, this does not mean that labor reallocation took place in the expected direction.[31]

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, one that requires us

[31] 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).

to take into account labor market frictions, migration dynamics (in particular preferences and spatial variation of skill rewards) and the match between worker heterogeneity and firm heterogeneity. A step in that direction has been carried out 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, Laura Hering, Miren Lafourcade, Carolina Lennon, Aguinaldo Maciente, Philippe Martin, Thierry Mayer, Sandra Poncet, 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), the EEA conference 2009 (Milan), the research seminar at University of Cergy-Pontoise and the workshop on Trade and Development at PSE (Paris) 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.

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

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, 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", Ensaios Econômicos EPGE n.532.

Fally, T., R.Paillacar and C.Terra (2008). "Economic geography and wages in Brazil: Evidence from micro-data," THEMA Working papers 23. U. de Cergy-Pontoise.

Foster, L., J.Haltiwanger and C.Syverson (2008). "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review* 98(1): 394-425.

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: 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). "How does market access shape internal migration? The Brazilian case," unpublished manuscript.

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

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," unpublished manuscript.

Levinsohn, J. and A.Petrin (2003). "Estimating Production Functions using Inputs to Control for Unobservables," *Review of Economic Studies* 70(2): 317-342.

Lewis, J. and D.Linzer (2005). “Estimating Regression Models in which the Dependent Variable is based on Estimates,” *Political Analysis* 13(4): 345-364.

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,” unpublished manuscript.

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

Muriel, B. and C. Terra (2009). “Sources of Comparative Advantages in Brazil,” *Review of Development Economics*, forthcoming.

Paillacar, R. (2007). “Market Potential and Worker Heterogeneity as Determinants of Brazilian Wages,” unpublished manuscript.

Pavcnik, N., A. Blom, P. Goldberg, and N. Schady (2004). Trade Policy and Industry Wage Structure: Evidence from Brazil. *World Bank Economic Review* 18(3): 319-344.

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.

Syverson, C. (2004). “Product Substitutability and Productivity Dispersion,” *Review of Economics and Statistics* 86(2): 534-550.

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

Van Biesebroeck, J. (2007). “Robustness of productivity estimates,” *Journal of Industrial Economics* LV(3):529-569.

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

A1. Data appendix

Education:

Educational variables are 9 dummies, one for each schooling level:

Level 1: Illiterate

Level 2: Primary School (incomplete)

Level 3: Primary School (complete)

Level 4: Middle School (incomplete)

Level 5: Middle School (complete)

Level 6: High School (incomplete)

Level 7: High School (complete)

Level 8: College (incomplete)

Level 9: College (complete)

Table A1: Summary statistics of individual characteristics

Variable:	Mean	Std. dev.
Log wage	1.508	0.852
Age	36.38	8.639
Age squared / 100	13.98	6.927
Educ. level 1	0.026	0.161
Educ. level 2	0.098	0.297
Educ. level 3	0.168	0.374
Educ. level 4	0.203	0.402
Educ. level 5	0.193	0.394
Educ. level 6	0.076	0.264
Educ. level 7	0.151	0.358
Educ. level 8	0.027	0.162
Educ. level 9	0.059	0.235
N obs	798494	

Notes: Summary statistics for the random sample; statistics for the full sample do not differ by more than 0.001.

A2. Measurement of Productivity

Data

Data by workers and firms are matched thanks to the firm identification number (CNPJ). Labor corresponds to the yearly average number of workers in the firm. Capital stock is estimated using the perpetual inventory method with a discount rate of 15% (results are not sensible to changes in discount rate between 5% or 25%).

The manufacturing survey (PIA after 1996) does not have information on capital stock, but we could impute the initial capital stock in 1995 from IBRE data (*Fundação Getulio 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.

Alternative measures

We constructed alternative measures of productivity using either the residual of OLS regressions or the Levinsohn and Petrin (2003) methodology. Table A2 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.

Table A2: Main resuts 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]**

Notes: The first line corresponds to the coefficient of productivity in the first step of Table 5; the second line corresponds to the main coefficient of MA in Table 5 (final step). Columns correspond to different measures of productivity.

Correlation between productivity and market access

As an additional result, we explore the relationship between our chosen measure of productivity and MA. According to the results in the first column of the table A3, there is no significant correlation between productivity and global MA at 5% (although it is significant at 10% level). 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 and significant at 1% level.

Table A3: Productivity and market access

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

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.

A3. First step regressions for the instrumental variable approach

Table A4 provide the first step estimations of the 2SLS regressions of Table 8. The P-value for the test of excluded instruments is lower than 0.01 for all regressions.

Table A4: First stage regressions corresponding to table 8

Instruments for Market Access:	Coefficient	Std. error	Observations
HMP in 1939	1.707**	0.056	540
Av. date of registration	-0.027**	0.002	540
{ HMP 1939 &	1.616**	0.064	
Av. date of registration	-0.005**	0.001	540

Notes: First-stage regressions for table 8, with industry fixed effects. *Market Access i* nstrumented by: HMP: Harris Market Potential = sum(GDP/distance) in 1939; average date of registration of municipalities within the state. Statistical significance: ** at 1% level.