The General Equilibrium Wage Impact of Trade-Induced Shifts in Industrial Compositions of Employment in Brazilian Cities, 1991-2000

Mardukhi, Jian

University of British Columbia, Department of Economics

1 October 2010

The General Equilibrium Wage Impact of Trade-Induced Shifts in Industrial Compositions of Employment in Brazilian Cities, 1991-2000*

Jian Mardukhi†

Abstract

Conventionally, it is presumed that restructuring of industrial composition of employment only modestly affects the average wage. This is because in a partial equilibrium setting such a restructuring affects the calculation of the average wage only through changes in employment shares of industries used as weights on *constant* industry wages. On the contrary, this paper brings substantial evidence indicating that aside from such partial equilibrium shift-share effects, a change in industrial composition sizably impacts all industry wages through general equilibrium (G.E.) feedbacks from the average wage – as a reservation wage in all industries in a search and bargaining framework – onto *all* industry wages. In particular, this paper uses Brazilian census data for years 1991 and 2000 to study the G.E. wage impacts of exogenous shifts in industrial compositions in cities of Brazil induced by substantial trade liberalization in this country during the 1990s. A restructuring of industrial composition in a city favouring high-wage industries that modestly raises the average wage in this city by only 1% through shift-share accounting, is estimated here to increase all industry wages in the city in average by at least twice as much – between 2 to 4 percent – in the long-run through the G.E impacts, resulting in an overall increase of 3 to 5 percent in the average wage. Concerns about endogeneity is address by using an IV strategy that exploits distance of a city from major international commercial ports as an indicator of how the change in trade policy impacted its industrial composition. The result is also robust to correcting for sample selection bias generated by regional migrations and to the presence of alternative explanatory mechanisms. The finding here highlights the importance of considering G.E. interactions in policy evaluations. It also indicates that major changes in national industrial or trade policies in developing countries such as Brazil, with already non-uniform distribution of economic development across regions, create geographical winners and losers depending on how the impacts are distributed across different localities sub-nationally. If the distribution of impacts is such that the losers-to-be regions are those already suffering, then balancing measures are necessary to avoid spatially uneven sub-national economic development.

Keywords: Industrial Composition, Wage Structure, International Trade, Sub-national Economic Development, Spatial Distribution of Policy Impacts, Brazil

JEL Classification: O18, R58, O24, R11, O25, J31, O54, O11, O12, R23

* I am very grateful to my supervisor, Paul Beaudry, and my advisers, David Green and Giovanni Gallipoli, for their support and guidance. I am also very grateful to Joseph Potter for providing me with the geographic aggregation formulas for Brazil and to Thomas Fujiwara for introducing me to Joseph Potter's work. The usual disclaimer applies. This paper is a part of my PhD thesis at the University of British Columbia.

† University of British Columbia, Department of Economics, 997 - 1873 East Mall Vancouver, B.C. V6T 1Z1. jmardukh@interchange.ubc.ca

1 Introduction

Conventionally, it is presumed that restructurings of industrial composition do not have important impacts on average wages other than the direct shift-share impact through changes in industries' employment shares (see Bound and Johnson, 1992). Average wage in a city is the local-employment-share weighted sum of local industry wages. Keeping local wages constant, an increase in the share of a high-paying industry will change the average wage figure only by the *wage premium* paid in that industry multiplied by the change in its share. As a result, in the event of a 5% increase in the share of an industry that pays 20% premium relative to average in other industries, the shift-share accounting predicts that it results in only 1% increase in the average wage. This is because an increase in an industry's share of local employment is a decrease in other industries' shares. As a result, the net direct impact on average wage from a shift in industrial composition is deemed to be modest. However, in a general equilibrium search and bargaining model of a labour market, the average wage can play the role of an outside option for the bargaining unemployed workers in all industries (see Beaudry, Green, and Sand, 2009), so that even a modest increase in average wage results in wages in all sectors to increase, which will bring about a further rise in average wage that will impact all industrial wages again through the same bargaining mechanism, and so on and so forth. This cycle will continue until it eventually dies off and new steady-state equilibrium takes shape. As a result of these cycles, the G.E. wage impacts of a shift in industrial composition could turn out to be large even though the initial direct impact on average wage may be small.

Contrary to the conventional presumption, this paper brings substantial evidence indicating that, aside from the wage impact of changes in within-industry labour demand, a shift in industrial composition has sizable between-industry impacts on wages through the general equilibrium (G.E.) feedbacks from the average wage onto *all* industrial wages. Industrial composition in a city is measured here as local-employment-share weighted sum of national industrial wage premia. Using Brazilian census data for years 1991 and 2000, this paper empirically identifies the G.E. wage impact of exogenous shifts in industrial composition of cities in Brazil during the 1990s that were brought about by substantial international trade liberalization in this country during the decade.

A restructuring of industrial composition in a city favouring high-wage industries that raises the average wage in this city by only 1%[3] through shift-share accounting[4] is estimated to increase all industry wages in the city by at least twice as much – between 2 to 4 percent – in the long-run only through the G.E impacts. In other words, the total increase in average wage from such a change in industrial composition is at least 3% – between 3 to 5 percent – with 67% of it being only due to the G.E. wage effects of shifts in industrial composition. The G.E. impacts are interpreted as spillovers from high-wage industries to other industries within a city; i.e., a shift in local industrial composition favouring higher paying industries improves the chances of getting hired in those industries for the wage-bargaining unemployed workers in the city, which results in higher wages being paid within each skill group by all industries in that locality.

It should be emphasized that the G.E. wage impacts sought after here are different from the wage effects of changes in within-industry demand for labour. A shift in industrial composition is made up of changes in within-industry labour demands. However, the within-industry changes in labour demand are expected to only affect the within-industry wages. What is identified here is the between-industry spillover of wages. An industrial composition shifts in favour of high-wage industries increases the average wage at first only through the shift-share effect. Such an increase in the average wage is then transmitted to all other industries through the G.E. mechanism in which the average wage plays the role of a reservation wage. In other words, through this G.E. mechanism, the high wages paid by the high-wage industries spill over onto all other industries. As a result, what is identified here is the between-industry spillover of wages and in terms of the empirics these are the estimated wage impacts associated with changes in industrial composition while keeping the city-industry or city level labour demands constant.[5]

The size of the G.E. wage impact estimated here is at least twice the conventional measures. To clarify this comparison, it helps to draw on some level of formality. The average wage in a local economy can be calculated as employment share weighted sum of industry wages:

$$\overline{w}_c = \sum_i \eta_{ic} w_{ic}, \tag{1.1}$$

[3] This is about 0.002 units increase in the measure of industrial composition in a typical city in Brazil. The sample mean of the measure of industrial composition over 1991-2000 is 0.25 with a standard deviation of 0.046.
[4] That is, keeping the wages constant.
[5] In the empirics, this is done by controlling for city-industry or city employment rates.

where η_{ic} is the employment share of industry *i* in city *c* and w_{ic} is the wage paid by industry *i* in city *c*. The effect of a shift in industrial composition on the local average wage is *conventionally* measured under the assumption that such shifts do not impact the structure of wages[6]:

$$A_c \equiv \sum_i (\Delta\eta_{ic}) w_{ic}. \tag{1.2}$$

In (1.2), A_c measures how much the local average wage changes as the local industrial composition of employment shifts, assuming the structure of industrial wages to be invariant to shifts in industrial composition. However, if the shift in industrial composition did affect the structure of local wages, then the total impact on average wage would have to be calculated according to the following decomposition of (1.1):

$$\begin{aligned} \Delta\overline{w}_c &= \sum_i (\Delta\eta_{ic}) w_{ic} + \sum_i \eta_{ic}(\Delta w_{ic}) \\ &= A_c + \sum_i \eta_{ic}(\Delta w_{ic}) \end{aligned} \tag{1.3}$$

This could be the case, for instance, if local wages (w_{ic}) were linked to the local average wage ($\overline{w}_c$). Consider a shift in industrial composition that generates a pure[7] increase in the average wage as measured by A_c in (1.2). If all local wages were in average linked to the local average wage by a multiplier, say γ, then such a shift in industrial composition would in average raise local wages by $\gamma \times A_c$. This would increase the local average wage by the same amount, which will result in a second round of increase in local wages in average by $\gamma^2 \times A_c$, and so on and so forth. For $\gamma < 1$, by the time a new equilibrium arises, local wages will in total increase in average by $\frac{1}{1-\gamma} \times A_c$. This means the second term in (1.3), which is the part of the total increase in local average wage that is other than (but generated by) the initial increase induced by the shift in industrial composition (which is the G.E. wage impact associated with the initial change in industrial composition), is equal to $\frac{\gamma}{1-\gamma} \times A_c$:

[6] The conventional measure is based on a standard shift-share analysis in which the impact of changes in industrial composition on average wages is obtained by multiplying the industrial wage premium associated with each industry in a base year by changes in the corresponding proportion of employment in that industry and then summing across industries.

[7] Pure in the sense that the wage structure is held fixed and the average wage changes solely by the shift in the weights used in calculation of the average wage.

$$\Delta\bar{w}_c = \sum_i (\Delta\eta_{ic})w_{ic} + \sum_i \eta_{ic}(\Delta w_{ic}) \\ = A_c + \frac{1}{1-\gamma} \times A_c. \tag{1.4}$$

Therefore, estimating $\frac{\gamma}{1-\gamma}$ over a long enough period of time will reveal if such spillovers exist and are sizable, and give an estimate of the magnitude by which the G.E. effect is larger than the conventional measures.[8]

A scenario based on which average wage could impact industry wages is defined in Beaudry, Green, and Sand (2009). They develop a search and bargaining model of a labour market that incorporates a general equilibrium channel through which changes in industrial composition of employment impacts wages in all industries. In this model, when an unemployed worker is matched with a firm in a specific industry in a city, they start bargaining over the wage. Unemployed workers use their outside option as leverage for bargaining. The outside option is to leave the match and search for another job in that city. The value of this option depends on the distribution of employment opportunities as it is assumed that an unemployed worker finding a job in another industry will find it in proportion to the relative size of that industry in the city. Thus, it turns out that improvements in the composition of employment in a city in favour of the higher paying industries[9] is an improvement in the value of the outside option of the bargaining, unemployed workers everywhere within the city and will consequently result in higher wages across all industries.

The current study builds on the work of Beaudry et al. (2009) and uses it as a guide for implementing the empirics. The Brazilian census data[10] is then used to measure the city compositions of industrial employment[11]. Geographical variation in this measure over time are then exploited to see whether there is general equilibrium effects associated with shifts in local industrial compositions that systematically affect local wages across all industries within cities.

The finding of this paper is important in two respects. Firstly, the sizable G.E. wage impact associated with shifts in industrial composition suggests that ex-ante evaluation of trade or

[8] The term $\frac{\gamma}{1-\gamma}$ is the ratio between the second part of the decomposition to the first part in (1.4).

[9] So that employment shares of higher-paying industries in the city's total employment increase that raises the chance of finding a job in these industries.

[10] IPUMS-International, a project in the Minnesota Population Centre Data Projects at https://international.ipums.org/international, is to be highlighted as the provider of the data used in this study.

[11] As employment-share weighted sum of national industrial wage premia or essentially similar to (1.1), only using innate industry wages rather than city-industry wages: $R_c = \sum_i \eta_{ic} w_i$.

industrial policies should not be carried out based on partial equilibrium assumptions and analysis (such as shift-share accounting here). Instead a proper evaluation of the impacts of such policy changes should consider general equilibrium interactions such as the one found here. Secondly, the finding of this paper highlights the role that changes in national trade or industrial policies play in creating (geographic) winners and losers depending on how the impact of such policy changes are transmitted sub-nationally to different regions. A change in the policies may induce different patterns of shift in industrial composition in different regions – favouring high-paying industries in some and low-paying industries in others. Given the sizable G.E. wage impacts found here, an un-even pattern of shifts in composition of employment could significantly contribute to the worsening of regional wage disparities and formation of wide spatial wage gaps. Both of these aspects are especially important in developing countries given that they are relatively more prone to major policy changes and sub-nationally egalitarian spatial economic development is of major importance to their overall developmental progress. Realizing that in most developing countries already the distribution of economic development across localities is non-uniform, such distribution of national policy impacts could especially be worrying if the losers-to-be regions are the already less-developed ones.

In the case of Brazil, the pattern of shifts in local industrial compositions in fact helped reduce the regional wage gap between the poor and rich areas in this country, probably due to the fact that Brazil benefits from geographically widespread major international commercial seaports (see Figure 1.1). Drawing on the finding of economic geography models of trade[12] (see Hanson, 2005; Redding and Venables, 2004; Head and Meyer, 2006; Knaap, 2006; Hering and Poncet, 2008; Mion and Naticchioni, 2005; Combes et al., 2008; Lederman et al., 2004; Da Mata et al., 2005; Fally et al., 2010) the distance of cities from these ports is effectively their distance from trade partners of Brazil and should partly determine how trade liberalization in Brazil during the 1990s were distributed across different cities and regions. Thus, even though the Northeast region was still the poorest in Brazil at the end of the 1990s[13] (see Figure 1.2) given the high share of agricultural employment in this region (see Graph 1.1), probably due to the existence of major seaports in this region and the fact that the northeast tip of Brazil on the side of the

[12] That distance as a trade barrier determines the size of trade between two economies.

[13] A World Bank report calls the Northeast region the region with the "... most remaining income poverty ..." in Brazil (World Bank, 2001, p. 1). In a study on the evolution of the regional GDP's 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, was 7.2 times that of Piauí, the poorest North Eastern state (Lall et al., 2004).

Atlantic ocean is the closest to the major trade partner of Brazil in North America, Europe, and Asia (see Figures 1.1, 1.2, and 1.3), majority of cities in this region benefited from shifts in their industrial compositions that favoured high-paying industries. As a result, not only the average wages increased in majority of these cities due to the favourable shifts in industrial composition, but because of adverse shifts in the composition of employment in cities in the richest region of Brazil in Southeast, the spatial wage gap between these two densely populated extremes in fact shrank during the 1990s (see Graphs 1.1, 1.2, 1.3, and 1.4). Of course, as is clear from Graph (1.5), there were also exceptions to this rule in both regions among cities.

Comparing the case of Brazil with Mexico is helpful here. Mardukhi (2009) studies the same G.E. wage impacts discussed here for the case of Mexican cities during the 1990s, the decade in which Mexico became a member of NAFTA and substantially increased its trade relations with the US. He estimates the G.E. wage impacts of shifts in industrial composition in Mexican cities to be almost twice as much as what is estimated here for Brazilian cities. He further finds that the pattern of shifts in compositions was inversely related to distance from Mexico-US border so that in South of Mexico, far away from the US border, the composition of employment shifted in favour of the low-paying industries while in the northern regions employment shifted toward the high-paying industries. Given that most southern cities in Mexico were already among the poorest in Mexico, he finds that the geographic losers of the trade liberalization of the 1990s in Mexico – as far as the impacts on industrial compositions and consequently wages are concerned – were regions that already were suffering from lower levels of economic development. This regressive spatial pattern of shifts in industrial compositions created a north-south disparity in growth rates of wages and helped increase the spatial wage gap in Mexico. In Brazil, however, due to a progressive pattern of shifts in compositions of employment across different regions, the wage impacts of trade policy-induced shifts in industrial composition were also progressively distributed spatially among regions and helped reduce the spatial wage gap between the poorest and richest regions in the country during 1991-2000.

To estimate the G.E. impact of the shifts in industrial composition consistently, it is essential to make sure that the estimating equation is not suffering from endogeneity. Aside from the simultaneity between employment shares of industries and industry wages within cities, which makes it possible that unobserved city wide improvements systematically move the measure of industrial composition and wages together, another likely source of endogeneity is a reflection problem between wages (see Manski, 1993; Moffitt, 2001). In a search and bargaining framework, industrial wages act as strategic complements; that is, high wages in one industry are associated with high wages in other industries as a result of bargaining within skill groups. As a result, it is likely that OLS estimates cannot disentangle the relationship between the average wage and wages that is solely generated by the G.E. impacts associated with shifts in industrial composition, from that of a simple correlation between the two due to the reflection problem. Both these sources are likely to affect the OLS estimates, which makes using an identification strategy necessary.

Concerns about the endogeneity of the measure of industrial composition are addressed by using an instrumental variable strategy that uses distance of a city from major commercial seaports as an indicator of how the change in international trade policy impacted different cities' industrial compositions. Specifically, the physical distance between Brazilian cities and 15 major international commercial seaports on the Atlantic side is used to predict the impact of Brazil's trade openings during the 1990s on local industrial compositions across different cities. This is essentially the impact of fast increases in trade with Europe, Japan, China and the US. The time horizon of current study incorporates major changes in trade policy in Brazil, which resulted in a significant boom in the country's international trade and major restructuring of the country's industrial composition. Under the assumption that the long term changes in industrial compositions during 1991-2000 in Brazil are at least partly due to the changes in trade policy in this period, and that geographic distance from major commercial ports of entry and exit into Brazil explains how the impacts of the change in trade policy are distributed across different cities, an instrumental variable (IV) strategy is devised.[14] As a result, identification in the IV strategy comes from differences across cities in terms of the impact of the change in trade policy on their industrial compositions during the horizon of this study and what is identified here is in fact the average local wage impact of the *trade-induced* shifts in local industrial compositions in Brazil, which are deemed to be exogenous.

A city's distance from major international commercial ports of entry and exit is expected to be exogenous to wages in that city conditional on controlling for city or city-industry level labour demands (employment rates). In general, time invariant geographic attributes of a city, such as geographic distance, are not expected to be part of the wage determination process in different sectors in the city. For instance, two cities with the same physical distance to international commercial ports are not expected to necessarily have the same wage structure or

[14] These assumptions will be justified and tested in the form of the first-stage regressions.

experience the same wage growth trajectories over time. If geographic distance of a city from international commercial ports has any importance in determining wages in a locality, it has to be through the within city-industry impacts on labour demand and supply, the impacts of which are captured here in the empirical model by controlling for city or city-industry employment rates. A similar strategy to that of Blanchard and Katz (1992) is used for instrumenting for employment rates based on national growth rates of employment within industries.

Over 1991-2000, Brazil's total merchandize export increased by 74.4% from US$31.6 billion to US$55.1 billion and its total merchandize import increased by 97.0% from US$30.0 billion to US$59.1 billion.[15] During this period, Brazil's total trade (export plus import) in manufacturing increased by 224.8% (67.8% and 157%, respectively). The full enactment of the regional free trade agreement between Argentina, Brazil, Paraguay, and Uruguay (MERCOSUR)[16] in 1995 falls in the middle of the horizon of this study. Although this agreement made Argentina as a major trade partner of Brazil (in terms of its share in total value of manufacturing exports and imports of Brazil, which places it second only to the US), it was not the only change resulting from Brazil's change in trade policies during 1991-2000. Within manufacturing, Brazil's imports from the US increased by 153% percent and its exports to the US increased by 102%. Respectively, the same statistics for Argentina is 313% and 305%, for Germany is 127% and 8.46% and, for Japan is 131% and -5.84%, for Italy is 170% and 30.1%, for France is 205% and 36.3%, for UK is 167% and 34.1%, and for China is 1,783% and 370%.[17]

Such trade boom and group of partners have amplified the importance of sea transportation and see ports in Brazil. Sea ports handle 95% of Brazil's trade by volume (and 85% by value).[18] This provides the rational for using distance from major commercial sea ports in Brazil as an indicator of how the impact of the change in trade policy was distributed across different cities in Brazil. It is to be mentioned that although Argentina's share of total Brazilian Trade increased significantly during this period as a result of MERCOSUR, the size of Argentina-Brazil trade was not large enough to generate major impacts on industrial compositions in cities far from the Argentina-Brazil border. As a result, distance for Argentina-Brazil border is not a good indicator of the geographic distribution of the impacts of change in trade policy during this period. For instance, respectively in 1991 and 2000 and within manufacturing, the US, France, Germany,

[15] UNCTAD Handbook of Statistics, 2009. (Calculated at current prices and current exchange rates.)
[16] MERCOSUR or the Common Market of the South was founded in 1991 and was fully enacted in 1995. For more information see www.mercosur.int.
[17] NBER-United Nations Trade Data, http://cid.econ.ucdavis.edu/.
[18] "An Ocean of Opportunities," *Brazil Now*, Ministry of External Relations, Brazil, November 2008.

Italy, UK, Japan, and China together imported 52% and 48% of Brazil's exports and exported 47% and 48% of Brazil's imports, with the US alone having an average share of 22% and 23% in Brazil's exports and imports in this period. This is while Argentina increased its share in Brazil's manufacturing exports and imports to 10.1% and 12.2% in 2000.[19]

The estimates of the relationship of interest are shown to be robust to correcting for the likely endogeneities discussed above and the sample selection bias that is caused by the migration of workers across cities within Brazil. Although it is verified that the sample suffers from selection bias, after correcting for it according to the approach in Dahl (2002), the estimates of the effects of changes in industrial compositions on wages remain significant and do not significantly change in magnitude. The findings are also shown to be robust in significance and size to the introduction of other alternative explanations for differences in wage changes across cities such as those related to diversity of employment in a city (Glaeser, Kallal, Scheinkman, and Shleifer, 1992), and levels of education (Moretti, 2004; Acemoglu and Angrist, 1999).

The structure of the paper is as follows. Section 2 briefly describes the theoretical model that is used as a guide in the empirical section. Section 3 explains the empirical strategy and the necessary steps required before moving on to the estimation. Section 4 introduces the data and section 5 reports the results of the estimations. Section 6 concludes the paper.

2 Theory

This section briefly explains the theoretical model in Beaudry et al. (2009), which is used a guide for implementing the empirics later on. The model shows how in a general equilibrium search and bargaining framework a change in industrial composition of employment affects industrial wages, even in industries that are not part of the change. When a worker is matched with a firm in a specific industry, they start bargaining over the wage. Workers use their outside option as leverage for bargaining. The outside option is the likelihood of leaving an industry to find a job in other industries that pay higher wages. If composition of employment in a city changes in favour of the higher paying industries, it is an improvement in the outside option of the bargaining unemployed workers and results in higher wages in all industries. Whether this composition effect is sizable or not is of course an empirical question.

It is important to note that the G.E. wage impacts discussed here are different from the wage effects of changes in within-industry demand for labour. A shift in industrial composition is

[19] NBER-United Nations Trade Data, http://cid.econ.ucdavis.edu/.

made up of changes in within-industry labour demands. The within-industry changes in labour demand are expected to affect the within-industry wages. What is discussed here, however, is the between-industry spillover of wages.

The economy is characterized by C local economies (cities), in which firms produce goods and individuals seek employment in I industries. To produce and make profits, firms create new jobs and seek to fill the costly vacancies. They weigh up the discounted costs of keeping those vacancies versus discounted expected profits they make by employing workers, and paying a wage that is city-industry specific, to produce an industry-specific product sold at an industry-specific price. Similarly, individuals compare the discounted expected benefits from being unemployed with being employed and receiving the city-industry wages. There is a random matching process through which workers are matched with firms. In a steady-state equilibrium of this economy, the value functions must satisfy the standard Bellman relationship. All throughout the model it is assumed that workers are not mobile across cities, an assumption that if relaxed is not going to change the key result because before migration between cities equalizes wages everywhere no matter what the industrial compositions are, increases in the cost of amenities within cities (for example price of housing) will bring migration to a halt. Furthermore, to avoid corner solutions in which all production concentrates in one city or in several cities but in one industry, it is as well assumed that cities have different advantages in different industries (denoted by ϵ's in terms of performance and in terms of costs of entry by Ω's). These city-industry advantages are defined by exogenous distributions that determine each city's advantages and disadvantages in all industries in terms of profits earned and costs of entry within each industry, and ultimately dictate the equilibrium values of all the variables in the model. In equilibrium all the variables of the model, including industrial compositions of cities, are functions of these exogenous city-industry advantage terms. Empirically, this will be the source of a likely endogeneity at the time of estimation as these city-industry cost advantage terms appear as the error term in estimating equations.

Without going into the details, solving the model for city-industry wages gives the following equilibrium relationship:

$$w_{ic} = \gamma_{c0} + \gamma_{c1} p_i + \gamma_{c2} \sum_j \eta_{jc}\, w_{jc} + \gamma_{c1} \epsilon_{ic}, \tag{2.1}$$

where w_{ic} is city c industry i's innate wage, p_i is price of the industry specific product, η_{jc} represents the fraction of city c's vacant jobs that are in industry j, and ϵ_{ic} is the exogenous advantage of city c in terms of performance (profits) in industry i. The parameters in this equation are all implicit functions of city-industry employment shares, city level employment ratios, a measure of the bargaining power of the workforce, and exogenous city-industry cost advantage terms.

The derived equation for city-industry wages captures the notion that in a search and matching framework, industrial wages act as strategic complements; that is, high wages in one industry are associated with high wages in other industries (for more details on the classic reflection or social interaction problem see Manski, 1993 and Moffitt, 2001). According to equation (2.1), increase in wages in one industry increases the average wage in city, and the latter would increase wages in all industries in the city as a result of the bargaining mechanism and improvements in the outside option of unemployed workers. This will be a source of endogeneity in estimating this equation. The strength of this strategic complementarity is captured by γ_{c2}. If workers are immobile across industries, γ_{c2} becomes zero, this effect disappears, and wages are determined solely by the value of marginal product.

According to equation (2.1), a pure shift in industrial composition that causes a one unit increase in the average city wage, $\sum_j \eta_{jc} w_{jc}$, increases industry wages by γ_{c2} in all industries. But these increases in all within industry wages cause the average wage to increase by another γ_{c2} units, inducing a further round of adjustments. The total effect of the pure change in industrial composition on the average wage would therefore be $\frac{1}{1-\gamma_{c2}}$ $(= 1 + \frac{\gamma_{c2}}{1-\gamma_{c2}})$. It is important here to explain that in this example the conventional accounting measure of the wage impact of such a shift in industrial composition is the same one unit by which the average wage is increased. That is, the conventional measures of the wage impact of shifts in structure of industrial employment ignore the link between local wages and the average local wage. As a result, the total wage impact of such a shift in industrial composition $(\frac{1}{1-\gamma_{c2}})$ can be decomposed as $1 + \frac{\gamma_{c2}}{1-\gamma_{c2}}$, where the first term is the conventional impact and the second term is the remaining general equilibrium impact. In this way, $\frac{\gamma_{c2}}{1-\gamma_{c2}}$ captures how big the spillover effect associated with a shift in industrial composition is relative to the simple accounting measure in a general equilibrium setting.[20]

The reflection problem inherent in equation (2.1) due to w_{ic} appearing on both side of the equation, is problematic since an increase in average wage will impact the wage in industry *i* but at the same time this will increase the average wage again. To get around this problem, and to directly show the impact of employment rate, manipulation[21] of this equation results in the following equation:

$$w_{ic} = d_i + \frac{\gamma_2}{(1-\gamma_2)} \sum_j \eta_{cj}\left(w_j - w_1\right) + \gamma_{i5} ER_c + \xi_{ic}, \tag{2.2}$$

where d_i is an industry level term that can be captured in an empirical specification by including industry dummies, w_j is the nationally innate wage in industry j and $w_j - w_1$ is the national level wage premium[22] in industry j relative to industry 1, ER_c is the city level employment rate and the added coefficient, γ_{i5}, reflects the effect of a change in the employment rate within an industry on wage determination in that industry. This coefficient may vary across industries since the effects of a tighter labour market may affect the bargaining power of firms in an industry with a high value-added product differently from the bargaining power of firms in an industry with a low value added product.

Equation (2.2) shows how city-industry wages depend on the industrial composition of a city's employments captured by the term $\sum_j \eta_{cj}\left(w_j - w_1\right)$. From here on, this term is denoted by R_c and is referred to it as *the measure of industrial composition*[23]:

$$R_c = \sum_j \eta_{cj}\left(w_j - w_1\right). \tag{2.3}$$

[20] In fact, the same ratio will be estimated later on as an average across cities.
[21] Taking linear approximation at the point where cities have identical industrial composition ($\eta_{ic} = \eta_i = 1/I$) and employment rates ($ER_c = ER$), which arises when there is no city-sector advantages in the model, and assuming similar matching probabilities across cities and sectors so that all the γ coefficients are nothing but the average of γ_c's across cities at these similar matching probabilities.

[22] Note that the theory is silent about attributes of workers, and specifically their skills. One should think of the wages and wage premia as calculated for one skill group so that an increase in the measure of industrial composition is not an increase of skill. In the empirics these will be obtained controlling for skills and other attributes of the workers.
[23] Notice that a high value for the measure of industrial composition indicates that the city's employment is concentrated in higher paying sectors.

Differencing the structural equation in (2.2) within a city-industry cell across two steady state equilibria, gives the following estimating equation:

$$\Delta w_{ic} = \Delta d_i + \frac{\gamma_2}{(1-\gamma_2)} \Delta R_c + \gamma_{i5} \Delta ER_c + \Delta \xi_{ic}, \qquad (2.4)$$

where Δd_i is an industry specific effect that can be captured in an empirical specification by including industry dummies, and $\Delta\xi_{ic} = \gamma_1 \Delta\epsilon_{ic} + \gamma_1 \frac{\gamma_2}{(1-\gamma_2)} \sum_j \frac{1}{I} \Delta\epsilon_{jc}$ is the error term, with I being the total number of industries.

This study is interested in estimating the coefficient on the changes in the measure of industrial composition in (2.4); $\frac{\gamma_2}{(1-\gamma_2)}$. Consistent estimates of this coefficient would provide an estimate of the extent of city-level strategic complementarity between wages in different industries by backing out γ_2. The coefficient $\frac{\gamma_2}{(1-\gamma_2)}$ is of interest in its own right as it provides an estimate of the total – direct and feedback – effect of a one unit increase in the measure of industrial composition on within industry wages, as opposed to γ_2, which provides the partial unidirectional effect. Specifically, it indicates how big the average wage spillover associated with a shift in industrial composition is relative to the conventional accounting measure associated with it.

Examining wages in one industry in different cities, a positive value for $\frac{\gamma_2}{(1-\gamma_2)}$ implies that for example agriculture wages will be higher in cities where employment is more heavily weighted toward high-rent industries, where high-rent industries are defined in terms of the national level wage premia. This arises in the model because unemployed workers in agriculture (or any other low paying industry) have better outside options in cities with distribution of employment more in favour of higher paying industries.

Endogeneity of the variables in the estimating equation (2.4) puts the success of the identification strategy in danger. As explained before, one source of endogeneity here is the kind of unobserved city wide overall improvements that may systematically move the measure of industrial composition and wages together. The success of the estimation strategy relies upon the properties of the error term in (2.4). As far as the change in measure of industrial composition is concerned, the requirement for OLS to give consistent estimates of the coefficients in (2.4) can be expressed as follows:

$$\underset{C,I\to\infty}{\text{plim}}\frac{1}{I}\frac{1}{C}\sum_{i=1}^{I}\sum_{c=1}^{C}\Delta R_c\Delta\xi_{ic} = \underset{C,I\to\infty}{\text{plim}}\frac{1}{I}\frac{1}{C}\sum_{c=1}^{C}\Delta R_c\sum_{i=1}^{I}\Delta\xi_{ic} = 0, \tag{2.5}$$

recognizing that a similar condition is required for the change in employment rate. Given that this error term is $\Delta\xi_{ic} = \gamma_1\Delta\epsilon_{ic} + \gamma_1\frac{\gamma_2}{(1-\gamma_2)}\sum_j\frac{1}{I}\Delta\epsilon_{jc}$, this condition effectively reduces to the properties of ϵ_{ic}. Both the measure of industrial composition and employment rates may be endogenous because they are functions of η_{ic}'s, which are correlated with the ϵ's.

It can be shown that an instrumental approach could be devised to consistently estimate the parameters in (2.4) if the ϵ's are assumed to follow a random walk process – i.e., if the increments are independent of the past. Intuitively, under the random walk assumption the residuals in (2.4) aggregated at the city level (averaged across industries within each city) are independent of past values. Thus, a useful instrumental variable could be a suitable function of the initial period local employment shares that varies only across cities and is highly correlated with the changes in the measure of industrial composition. The decomposition of changes in the measure of industrial compositions into two parts, one based on changes in the city-industry employment shares and the other based on changes in the national wage premia, will be used as a guide in choosing the suitable functional form in generating the instruments[24]. Chapter 3, section 3.2 presents a more detailed exposition of generating instruments. Also, a detailed discussion of the source of the endogeneity and the assumptions required for the OLS or IV approach to work are left for section 3.2 below.

3 Empirical Strategy

The aim of this section is to explain the empirical strategy, potential issues, and necessary steps and approaches devised for a proper estimation of the relationship between the measure of industrial composition and city-industry wages. Briefly, the empirical strategy here is to explore geographical variation of changes in the measure of industrial composition over two far enough points in time to see whether changes in city-industry wages are systematically related to the shifts in local industrial compositions. Several preliminary steps are required to prepare the data for estimating the ultimate relationship of interest. Also, endogeneity and sample selection bias

[24] $\Delta R_c = \Delta\sum_j \eta_{cj}\cdot(w_j - w_1) = \sum_j \Delta\eta_{cj}\cdot(w_j - w_1) + \sum_j \eta_{cj}.\Delta(w_j - w_1) = \Delta\mathrm{R}_\mathrm{c}^1 + \Delta\mathrm{R}_\mathrm{c}^2$

are potential issues that may jeopardize the success of the estimation strategy. This section explains the preliminary steps and addresses several estimation issues.

The empirical estimating equation that closely matches equation (2.4) is specified as:

$$\Delta w_{cit} = \Delta d_t + \Delta d_{it} + \beta \Delta R_{ct} + \gamma \Delta ER_{cit} + \Delta \xi_{cit} \,. \tag{3.1}$$

Since the data used in this study is only available for two distinct points in time ten years apart from each other, Δd_t, the change in time dummy, becomes a constant playing the role of an intercept and Δd_{it}, the change in industry-time dummies, becomes nothing but a full set of industry dummies excluding the base industry. The left hand side variable is the time change in wages paid in industry i, city c, ΔR_{ct} is the change in the measure of industrial composition over time, ΔER_{cit} is the change in city-industry employment rate over time, and $\Delta \xi_{cit}$ is the error term.

The parameter of interest is β that captures the relationship between the measure of industrial composition and city-industry wages in isolation from the impacts of the within industry or city level demands for labour. In estimation of equation (3.1) and conducting inferences, consistency of the estimates is crucial. Assuming consistent estimation, the goal would be to test the null hypothesis that $\beta = 0$. If the null cannot be rejected, one can disregard the inter-industry wage interactions in the process of wage determination in local economies. On the other hand, however, a statistically significant and sizable coefficient is indicative of a general equilibrium mechanism through which local industrial composition of employment in each local economy has a significant impact on wages in all sectors in that economy. If this mechanism is estimated to be sizable, disregarding the general equilibrium impact on wages could turn out to be costly in developmental policy making.

To be used in equation (3.1), industrial composition will be measured as the city-industry employment share weighted sum of the national industrial wage premia (as a fraction of wages in a base industry):

$$R_{ct} = \sum_i \frac{e_{cit}}{\sum_i e_{cit}} \cdot \left(\frac{w_{it}}{w_{1t}} - 1 \right). \tag{3.2}$$

Measuring wage premia as a fraction of the base industry wage rather than a level difference will allow for using logarithm of wages instead of the levels all throughout the empirics. In (3.2), e_{cit} indicates employment in industry i, city c, in year t, w_{it} indicates industry i's *innate* wage at the national level in year t, and in the same way w_{1t} measures the intrinsic wage in industry one in the national economy. The share of industry i in total employment in city c (or $\frac{e_{cit}}{\sum_i e_{cit}}$) is the weight associated with the national wage premium in industry i indicated by $(\frac{w_{it}}{w_{1t}} - 1)$. So, the measure of industrial composition is a proxy for average wages in a city and measures how much wage premium a city is generating given its distribution of employment across different sectors.

To illustrate how this way of measuring the industrial composition (using national rather than local wage premia) is useful, consider the following example. In a given year, between two otherwise similar cities, the one with higher concentration of employment in sectors that intrinsically offer higher wages is expected to have a higher measure of industrial composition. This is because national wage premia, rather than local wage premia, are used and wage realization in each city has an insignificant role in the construction of the measure of industrial compositions. Therefore, a city with relatively higher wages in all sectors is not necessarily going to have a higher measure of industrial composition by construction. If the spillover mechanism from good jobs is at work, then, it is expected to see higher wages across all sectors in the city with a higher measure of industrial composition. Of course, as explained before, if there are city wide improvements that systematically move the measure of industrial composition and wages together, because we cannot control for city fixed effects, the OLS estimates are no more reliable due the endogeneity of regressors. I will later explain that under some assumptions an IV approach can be devised to deal with the likely endogeneity of the regressors.

The city-industry employment shares, $\frac{e_{cit}}{\sum_i e_{cit}}$'s, can directly be calculated from the data with no problem, but the national wage premia needs to be estimated. Comparing equations (3.2) with (3.1), in (3.2) the formula has been modified by dividing the measure of industrial composition by w_1 to allow for using the log-wages in estimating the industrial wage premia.[25] To empirically estimate the national wage premia in different sectors from the observed wage and employment data, the following specification is used and estimated separately for each year:

[25] The estimates of β will become dependent on the choice of the base sector as a result of this modification. However, it can be shown that this dependence is corrected for by using the logarithm of wages as the dependent variable. It can be shown that the estimated β in this way is different from the true β by a factor of $\frac{ln(\bar{w}_1)}{ln(\bar{w})}$, where $\bar{w}$ is the geometric average of wages at the national level and $\bar{w}_1$ is the same but only within the base industry. The closer that logarithm of the geometric average wage in the base sector is to the geometric average of wages at the national level, the less would be the difference between the estimated β and the true β. Different choices for the base sector were considered. The estimates are robust with respect to the the choice of the base sector.

$$ln(W_{kci}) = \alpha + \boldsymbol{X}_k'\boldsymbol{\gamma} + \sum_i \varpi_i d_i + \sum_c \alpha_c d_c + \varepsilon_{kci}, \tag{3.3}$$

where W_{kci} is the observed wage received by person *k* in city *c* working in industry *i*, $\boldsymbol{X}_k$ denotes an array of worker attributes, d_i and d_c indicate industry and city dummies, and *ln(.)* is the natural logarithm function. In equation (3.3), estimates of ϖ_i's in each year by definition capture the national level industrial wage premia relative to the base industry and can be used to replace the term $(\frac{w_{it}}{w_{1t}} - 1)$ in equation (3.2):

$$R_{ct} = \sum_i \frac{e_{cit}}{\sum_i e_{cit}} \cdot \left(\frac{w_{it}}{w_{1t}} - 1\right) \equiv \sum_i \frac{e_{cit}}{\sum_i e_{cit}} \varpi_{it}. \tag{3.4}$$

In (3.4) ϖ_{it} is the coefficient of the respective industry dummy in equation (3.3) estimated for year *t*.

The next variable that requires attention is the left-hand-side variable in equation (3.1); w_{cit}. In the theoretical model, the worker is abstracted from all attributes or that the wages considered in the model are independent of the attributes of the workers and are *intrinsic* to the industry and city where they work.[26] It is therefore necessary to adjust the data on individual wages for all the attributes for which information is available and properly aggregate the wages from individuals to the city-industry level since after all individual attributes are important determinants of the wages. It is further necessary in the case of Brazil since individual diversity is an important determinant of spatial wage differences in the country (see Fally et al., 2010). Duarte et al. (2004) show that differences in wages between the Northeast and Southeast regions in Brazil can be explained by differences in workers' educational attainment. If such regional difference in the education levels of the work force across regions can be explained by sorting (Combes and Duranton, 2006) or endogenous differences in returns to schooling (Redding and Schott, 2003), not controlling for the demographic differences across individuals may result in finding an artificial relationship between industrial composition of employment and wages due to the relationship between composition of workers and industrial composition of employment across different regions.

[26] Another way to interpret this, is to say that the model is for a person from a given skill group.

The coefficients of city-industry dummies in the following estimating equation can be considered as regression adjusted wages for the attributes of workers averaged across individuals within each city-industry cell:

$$ln(W_{kci}) = \alpha + \boldsymbol{X}_k'\boldsymbol{\gamma} + \sum_c \sum_i w_{ci} d_{ci} + \omega_{kci}. \tag{3.5}$$

Equation (3.5) can be estimated separately for each year using the sampling weights in the data so that each round of estimation generates the appropriately aggregated city-industry wages for that year, which will be used as the left-hand-side variable in equation (3.1).

In the same way as city-industry employment shares, the city level employment rates can also be computed directly from the data. Having generated all the appropriate dependent and explanatory variables, equation (3.1) can now be estimated to see whether changes in industrial composition of employment in Mexican cities systematically relay externalities on all local wages across all sectors. It remains in this section to address the concerns about endogeneity and sample selection as follows.

3.1 Selection

This section addresses the concerns about selection bias that the empirical strategy may suffer from. If in practice workers are mobile across cities and choose where to live and work by comparing different cities in terms of their personal priorities, then individuals currently observed living in a city are not a random sample of the population. An individual's wage is not observed in any city other than the one they choose to be a resident of (born there and not moved anywhere else or born somewhere else and moved to this city). This will compromise one of the conditions required for the consistency of OLS estimates of these regressions being the zero mean residual. In practice, in equations (3.3) and (3.5) a conditional residual mean term, conditioned on the wage figure being observable, is of concern. For example, equation (3.3) can be written as:

$$\begin{aligned} &E[ln(W_{kci})|\boldsymbol{X}_k,\ d_i, and\ W_{kci}\ being\ observed] \\ &\qquad = \alpha + \boldsymbol{X}_k'\boldsymbol{\gamma} + \sum_i \varpi_i d_i + E[\varepsilon_{kci}|\boldsymbol{X}_k,\ d_i, and\ W_{kci}\ being\ observed]. \end{aligned} \tag{3.6}$$

It is not clear if the conditional error mean term in (3.6) is actually zero in a self selected sample. If this is not the case, then the conditional residual mean term is correlated with other regressors and OLS is no more consistent.

Intuitively, if suddenly a group of individuals move from a city to another city in expectation of higher wages for reasons not observable to us but related to the structure of wages ($\Delta\xi_{ic}$), the change in the measure of industrial composition in equation (3.1) will also capture the impact of this sort of movements and the OLS estimation of this equation may give significantly-different-from-zero estimates of the relationship of interest, without it really existing. Thus, it is very important to adjust the empirical strategy to correct for this possibility.

In addressing this issue, the approach in Dahl (2002) is implemented. Dahl (2002) develops an econometric approach to correct for sample selection bias. He builds a multi-market model of mobility and earnings, in which individuals choose where in any of the 50 U.S. states to live and work, and proposes a semi parametric methodology to correct for sample selection bias in such a choice model. He shows that the bias correction is an unknown function of a small number of selection probabilities, which are calculated without making any distributional assumptions simply by classifying similar individuals into cells and estimating the proportion of movers and stayers for each place of birth and cell combination. His work essentially shows that in order to correct for the selection bias, under some sufficiency conditions, the conditional error mean term in (3.6) can be replaced by an unknown function of the relevant migration probabilities in the outcome regression, which can then be estimated with simple OLS.

Modifying the approach in Dahl (2002) for the setting here, the mean error term can be identified as a function of the relevant migration probabilities:

$$E[\varepsilon_{kci}|\boldsymbol{X}_k,\ d_i, and\ W_{kci}\ being\ observed] = \sum_b d_{kbc} \cdot f_{bc}(P_{kbc}, P_{kbb}) + \vartheta_{kci}\ , \tag{3.7}$$

where d_{kbc} is an indicator that takes one only if person k born in state b has actually moved to city c, $E[\vartheta_{kci}|\boldsymbol{\kappa}_{kci},\ d_i, and\ W_{kci}\ being\ observed] = 0$, and $f_{bc}(\cdot)$ is an unknown function of the probability that person k, born in state b, is observed in city c (P_{kbc}) and probability that person k, born in state b, remains in the same state (P_{kbb}). The function $f_{bc}(\cdot)$ is chosen to be quadratic in each of the probabilities separately. In this way, equation (3.3) can be written as:

$$ln(W_{kci}) = \alpha + \boldsymbol{X}_k'\boldsymbol{\gamma} + \sum_i \varpi_i d_i + \sum_b d_{kbc} \cdot f_{bc}(P_{kbc}, P_{kbb}) + \vartheta_{kci}. \tag{3.8}$$

Notice that for non-movers, the correction terms are only functions of the probability of staying since for individuals who do not move from their state of birth $c = b$.

In the same way, equation (3.5) can also be corrected for selection:

$$ln(W_{kci}) = \alpha + \boldsymbol{X}_k'\boldsymbol{\gamma} + \sum_c \sum_i w_{ci} d_{ci} + \sum_b d_{kbc} \cdot f_{bc}(P_{kbc}, P_{kbb}) + \varsigma_{kci}, \tag{3.9}$$

where $E[\varsigma_{kci} | \boldsymbol{X}_k,\ and\ W_{kci}\ being\ observed] = 0$.

In a given city c, the identification for the movers (P_{kbc}) comes from the variation in the state of birth and the distance between the state of birth and the city which determines the probability that people make a given move. So, here the underlying assumption is that the state of birth and the distance between the state of birth and the city the person is observed in for the case of movers are not directly related to the wage a person receives. In other words, two individuals with exactly similar characteristics, living and working in the same city, but born in different states with different distances from this city, will not necessarily receive different amounts. For the stayers, however, identification comes from the differences in family status and hence is the assumption that family status is not directly related to the wage the person receives.

3.2 Endogeneity

Both the measure of industrial composition and employment rate are likely to be endogenous in a general equilibrium framework. As was indicated in section two above and is reviewed in detail in the appendix, as far as the change in the measure of industrial composition is concerned the consistency of estimates of the parameters in equation (3.1) relies partly[27] on the following condition:

$$\underset{C,I\to\infty}{plim} \frac{1}{I}\frac{1}{C}\sum_{i=1}^{I}\sum_{c=1}^{C} \Delta R_c \Delta\xi_{ic} = \underset{C,I\to\infty}{plim} \frac{1}{I}\frac{1}{C}\sum_{c=1}^{C} \Delta R_c \sum_{i=1}^{I} \Delta\xi_{ic} = 0, \tag{3.10}$$

where from the theory $\Delta\xi_{ic} = \gamma_1 \Delta\epsilon_{ic} + \gamma_1 \frac{\gamma_2}{1-\gamma_2} \sum_j \frac{1}{I} \Delta\epsilon_{jc}$. Following Beaudry et al. (2009) city-industry performance advantage terms can always be decomposed as $\epsilon_{ic} = \hat{\epsilon}_c + \upsilon_{ic}^{\epsilon}$ into absolute

[27] A similar condition is required for the employment rate, ΔER_c.

city advantage, $\hat{\epsilon}_c$, and comparative city-industry advantage term, υ_{ic}^{ϵ}, where by definition $\sum_i \upsilon_{ic}^{\epsilon} = 0$. Condition (3.10) depends primarily on properties of the absolute advantage component $\hat{\epsilon}_{ct}$. It can be shown that the condition for *consistency of OLS estimates relies on the assumption that absolute advantages are independent of comparative advantages in all periods* (see the appendix for details). Intuitively, this requirement means that shifts in the measure of industrial composition should not depend on average city-wide improvements in wages in that city. In other words, it implies that whatever drives general city performance is not related to a particular pattern of industrial structure. Since city fixed effects cannot be controlled for and this condition may not hold, using instrumental variables is necessary for consistent estimation of equation (3.1).

Under a weaker assumption that changes in the absolute advantages are independent of the initial set of comparative advantage factors for that city – i.e., if absolute advantages follow a random walk process with increments independent of past values – an instrumental variable approach can be devised. Under this assumption, one can use national growth rates of employment within each industry to reconstruct the changes in city level employment shares based on the initial level of employment and the national growth rate of employment within each industry. Essentially, this approach looks at what the change in the measure of industrial composition would have been had the employment in each industry across cities had grown at the same rate as national level employment in that industry. This approach essentially assumes one growth rate for each industry across all cities. This could be an effective strategy for predicting local variations in industrial compositions that closely follow the average national variations.

However, changes in employment within an industry at the national level may mask sub-national heterogeneities across different cities, which are important in understanding how a change in national policy (such as trade liberalization) is spatially distributed and especially when this distribution is not spatially uniform. Using one growth rate for employment in an industry across all cities may not capture the important spatial heterogeneities in terms of different patterns of shifts in industrial compositions that are induced by changes in national policies. In a developed country, where infrastructures are more or less similar everywhere and there are no major trade or industrial policy changes, economic opportunities are more or less similar everywhere. But both of these are probably not true for most less developed countries including Brazil.

Under the assumption that the long term changes in industrial compositions during 1991-2000 in Brazil are at least partly due to the major changes in trade policy in this period, and that geographic distance from major international commercial ports of entry and exit in Brazil explains how the impact of the change in policy is distributed across different cities, an instrumental variable (IV) strategy is devised here.[28] In particular, the distance between Brazilian cities and 15 major international commercial sea ports on the Atlantic side is used to predict the impact of Brazil's trade openings during the 1990s on local industrial compositions across different cities (essentially the impact of fast increases in trade with Europe, Japan, China and the US as well as Argentina).

As a result of this identification strategy used, what is identified here is in fact the average local wage impact of the trade policy-induced shifts in local industrial compositions in Brazil. Identification in the IV strategy comes from differences across cities in how the change in trade policy impacted their industrial compositions during the horizon of this study. The time horizon of this study incorporates major changes in trade policy in Brazil, which resulted in a significant boom in the country's international trade and major restructuring of the country's industrial composition.

Total merchandize export of Brazil increased by 74.4% from US$31.6 billion in 1991 to US$55.1 billion in 2000 and total merchandize import of Brazil increased by 97.0% from US$30.0 billion to US$59.1 billion.[29] During this period, Brazil's total export and import in manufacturing increased by 67.8% and 157%, respectively. The full enactment of the regional free trade agreement between Argentina, Brazil, Paraguay, and Uruguay (MERCOSUR)[30] in 1995 falls in the middle of the horizon of this study. Although this agreement resulted in major increase in the importance of Argentina as a trade partner of Brazil (in terms of its share in total exports and imports of Brazil, which places it second only to the US), it was not the only change resulting from Brazil's change in trade policies during 1991-2000. Within manufacturing, the Brazil's imports from the US increased by 153% percent and its exports to the US increased by 102%. Respectively, the same statistics for Argentina is 313% and 305%, for Germany is 127%

[28] These assumptions will be justified and tested by looking at the performance of instruments used here in the first-stage regressions.

[29] UNCTAD Handbook of Statistics, 2009. (Calculated at current prices and current exchange rates.)

[30] MERCOSUR or the Common Market of the South was founded in 1991 and was fully enacted in 1995. For more information see www.mercosur.int.

and 8.46% and, for Japan is 131% and -5.84%, for Italy is 170% and 30.1%, for France is 205% and 36.3%, for UK is 167% and 34.1%, and for China is 1,783% and 370%.[31]

Such a trade boom and group of partners has amplified the importance of sea transportation and see ports in Brazil. Sea ports handle 95% of Brazil's trade by volume (and 85% by value). This provides the rational for using distance from major commercial sea ports in Brazil as an indicator of how the impact of the change in trade policy was distributed across different cities in Brazil. It is to be mentioned that although Argentina's share of total Brazilian Trade increased significantly during this period as a result of MERCOSUR, the size of Argentina-Brazil trade was not large enough to generate major impacts on industrial compositions in Brazilian cities far from Argentina-Brazil border, and therefore distance from Argentina on its own is not an indicator of changes in local industrial compositions outside of the neighbouring regions. For instance, within manufacturing the US, France, Germany, Italy, UK, Japan, and China together imported 52% and 48% of Brazil's exports and exported 47% and 48% of Brazil's imports, respectively in 1991 and 2000, with the US alone having an average share of 22% and 23% in Brazil's exports and imports in this period. This is while Argentina increased its share in Brazil's exports and imports to 10.1% and 12.2% in 2000.[32]

The validity of instruments partly relies on the assumption that wage determination process in each city is independent of the city's distance from the major international commercial ports, conditional on controlling for employment rates. In general, time invariant geographic attributes of a city such as its distance from major commercial ports are not expected to be part of the wage determination process in different sectors within that city in ordinary times. Since two cities with the same distance to a commercial port are not expected to necessarily have the same wage structure or experience the same wage growth trajectories over time, geographic distance is in general expected to be orthogonal to the unobserved determinants of wages in each city. If geographic distance has any importance in determining wages in a locality, it has to be through its impacts on labour demand and supply within each city-industry, which are captured here by city-industry employment rates. It crosses mind that because of choosing a specific time period, distance from major commercial ports in Brazil may have become a relevant factor in part in the wage determination process in different sectors in cities; while being close to the major ports were not an advantage before the 1990s, during this decade being close to the major ports might

[31] NBER-United Nations Trade Data, http://cid.econ.ucdavis.edu/.
[32] Ibid.

have become an advantage (or disadvantage). Nevertheless, as far as the wage determination process is concerned, what is conceived as advantage (or disadvantage) in cities closer to the major international commercial ports is only due to the demand and supply effects of the trade boom that are mediated through distance from major commercial ports and affects the wage determination process through impacting demand and supply of labour within industries in cities (e.g. as a result of firm relocations or worker migration). These are controlled for by the changes in employment rate in equation (3.1) and by correcting for selection bias as explained before. In other words, if there is a relationship between distance from border and changes in city-industry wages, it should be through mechanisms that are already controlled for by the employment rates in equation (3.1).

Due to specific functional forms that will be used here, the validity of instruments hinges on the same weaker assumption mentioned before that the city absolute advantage terms follow a random walk process with increments independent of past values. The instruments are constructed based on the decomposition of ΔR_{ct} into a part that captures the change in the measure of industrial composition resulting from changes in employment shares (ΔR_{ct}^1 below) and another that captures the changes resulting from variations in national level wage premia of sectors (ΔR_{ct}^2 below):

$$\begin{aligned}\Delta R_{ct} &= \sum_i \eta_{cit+1}\varpi_{it+1} - \sum_i \eta_{cit}\varpi_{it} \\ &= \sum_i \eta_{cit+1}\varpi_{it+1} - \sum_i \eta_{cit+1}\varpi_{it} + \sum_i \eta_{cit+1}\varpi_{it} - \sum_i \eta_{cit}\varpi_{it} \\ &= \sum_i (\eta_{cit+1} - \eta_{cit})\varpi_{it} + \sum_i \eta_{cit+1}(\varpi_{it+1} - \varpi_{it}) = \Delta \mathrm{R}_{\mathrm{ct}}^1 + \Delta \mathrm{R}_{\mathrm{ct}}^2,\end{aligned} \tag{3.11}$$

Each decomposition works like a manual for constructing instruments; IV1 based on ΔR_{ct}^1 and IV2 based on ΔR_{ct}^2:

$$IV1 = \sum_i (\hat{\eta}_{cit+1} - \eta_{cit})\varpi_{it}, \tag{3.12}$$

$$IV2 = \sum_i \hat{\eta}_{cit+1}(\varpi_{it+1} - \varpi_{it}), \tag{3.13}$$

where $\eta_{cit} = \frac{e_{cit}}{\sum_i e_{cit}}$, $\hat{\eta}_{cit+1} = \frac{\hat{e}_{cit+1}}{\sum_i \hat{e}_{cit+1}}$, ϖ_{it} is the wage premium in industry i at the national level, and $\hat{e}_{cit+1} = e_{cit}(1 + g_{ic})$ with g_{ic} being the fitted values from the following regression:

$$\Delta ln(e_{cit}) = \theta_0 + \theta_{1i} d_i + \sum_p \theta_{1icp} \left(d_i \times dist_{cp} \right) + err., \tag{3.14}$$

where e_{cit} is industry *i*'s employment in city *c*, in year *t*, $dist_{cp}$ is the physical distance of city *c* from the a major international commercial port in Brazil[33], $err.$ is the error term, and *ln(.)* is the natural logarithm function. The fitted values from (3.14) would generate city-industry growth rates that depend on the geographic distance of cities from the major commercial ports in Brazil, which can be denoted g_{ic} and used as shown above to generate IV1 and IV2.

The equality between OLS and IV estimates will be indicative of two important points: that OLS is consistently estimating the coefficients in (3.1) and that the stronger condition required for consistency of OLS is valid in the data (i.e., in a city absolute advantages are independent of comparative advantages during all periods). Of course these results are to the extent that the assumptions required for validity of the instrumental approach attend in the sample.

It remains to address the concerns about the omitted variable bias given the existing alternative hypothesis in the literature regarding wage determination in cities. To make sure that the OLS estimates are robust at the presence of alternative explanations for differences in wages across cities such as those related to city size, education levels (Moretti, 2004; Acemoglu and Angrist, 1999), and diversity of employment in a city (Glaeser, Kallal, Scheinkman, and Shleifer, 1992), additional variables representing these alternative hypotheses will be added to equation (3.1).

Before moving to the estimation results in the next sections, it is important to mention that similar to ΔR_{ct}, ΔER_{cit} is also likely to be endogenous as explained before. To deal with the endogeneity of this variable an approach similar to Blanchard and Katz (1992) is used. Specifically, $\eta_{cit-1} g_i$ is used as an instrument, were g_i is the national growth rate of employment in industry i.[34] It can easily be shown that under the same weaker assumption required for the consistency of other instruments, this instrument is also valid.

[33] Physical Distance from 15 major commercial seaports is considered here.

[34] Blanchard and Katz (1992) use $\sum_i \eta_{cit-1} g_i$ as an instrument for ΔER_{ct}.

4 Data

The data used here are extracted from the tenth and eleventh Brazilian General Censuses for years 1991 and 2000, originally produced by the Brazilian Institute of Geography and statistics (IBGE)[35] and preserved and harmonized by Minnesota Population Center (2008). The sample is narrowed down to employed males and females aged 16 to 65, who are wage or salary workers in an identified industry with identified levels of education and positive monthly income. The sample is composed of 1,902,065 and 2,365,104 individual observations, respectively for years 1991 and 2000.

It is necessary in this study to define geographic limits of the local labour markets that are fairly consistent over time and are broad enough so that inhabitants do not commute beyond the boundaries to work. At the same time, since the relationship of interest is identified based on variations across these geographic units, having a large number of them in the sample is favourable. The ideal would have been to have a large number of consistent metropolitan areas in the sample. In the absence of such a privilege, minimum comparable areas had to be defined. The smallest geographic identifier in the publically available Brazilian censuses is the municipality. The number of municipalities in these data grew from 4,500 in 1991 to 5,280 in 2000. Some sort of aggregation of municipalities is needed to define geographically numerous and economically independent regions. This has been carried out by Potter, Schmertmann, and Cavenaghi (2002) by aggregating municipalities that are consistent over time and are large enough to be used for proper economic analysis of this sort.[36] They aggregate the municipalities into 502 comparable geographic regions across Brazilian censuses.[37] Amaral, Hamermesh, Potter, and Rios-Neto (2007) use these geographic divisions for studying the impact of demographic changes on structure of wages in Brazil. The same geographic identifiers (435 of

[35] Instituto Brasileiro de Geografia e Estatística (or IBGE in short).

[36] If some of these geographic divisions are not large enough so that some workers who live in one division commute to another division for work, then this translates into an error in the measurement of industrial composition in each geographic division and will adversely impact the consistency of estimation results. A proper instrumental variable (IV) approach is then required to deal with this issue. Since geographic distance forms the basis of the IV approach, two adjacent geographic divisions that should have been considered as one will be assigned very similar measures of distance, as if they were in fact considered as one geographic division. This approach, therefore, should compensate for the likely inconsistency generated by the geographic divisions.

[37] They based their aggregation on census micro-regions, which are officially designated sets of contiguous municipalities defined according to economic homogeneity, and commercial and transportation links. Like municipalities, micro-regions also changed between censuses. They eliminate these inconsistencies by aggregating municipalities to obtain minimum comparable areas (MCAs), representing the smallest areas uniquely identifiable from the 1960, 1970, 1980, 1991, and 2000 municipality codes. They then aggregate these MCAs into clusters that closely approximate the 1991 census microregions.

them that could be merged with the census data publically available), dubbed cities here after, are used in this study.[38] In this way, 22% of the sample in 1991 and 25% of the sample in 2000 that contain observations that do not belong to the list of cities used here are dropped.

The wage variable associated with each individual reports the person's total monthly income from their labour (from wages) in the previous month. This variable is adjusted for inflation to constant year 2000 values. Industries considered are 15 major divisions of all economic activities in Brazil, namely Agriculture, fishing, and forestry, Mining, Manufacturing, Electricity, gas and water, Construction, Wholesale and retail trade, Hotels and restaurants, Transportation and communications, Financial services and insurance, Public administration and defence, Real estate and business services, Education, Health and social work, Other services, and Private household services. A higher number of industries would have been preferred but due to impossibility of matching the definition of industries across years in the census only the use of these 15 major categories was feasible. Using aggregated industries is likely to make capturing the wage complementarity more difficult since with detailed industry definitions there would have been higher degree of variation in wages within an industry and across cities and in the measure of industrial composition.

5 Estimation Results

This section describes the estimation results. First, the baseline results are reported and then likely issues that may be associated with it (selection bias and endogeneity) are dealt with. The robustness checks will from the last step.

5.1 Baseline Estimation Results

Table (5.2) reports the estimation results of equation (3.1). Three different specification of this equation are estimated and reported here: one with city-industry employment rates as a control variable under OLS (1), one with city employment rates under OLS (2), and under OLS (3) with the same specification as under OLS (1), that is controlling for city-industry employment rates, and corrected for sample selection bias as explained in section 3.

The results under the first two columns in table (5.2) indicate positive and statistically highly significant estimates of the coefficient of ΔR_c. Controlling for changes in city employment rate

[38] Courtesy of Joseph E. Potter.

or city-industry employment rates do not significantly affect the size and significance of the coefficient of the change in the measure of industrial composition[39]. This result is also robust to correcting for sample selection bias. Graph (5.1) depicts a scatter diagram of the controlled[40] variation in the change in city-industry wage versus controlled variation in the change in the measure of industrial composition. The OLS results do not seem to be driven by outliers.

If the OLS estimate of the relationship between industrial composition of employment and wages within cities is consistent, which relies on a set of assumptions reviewed in previous sections and in detail in the appendix, the positive and significant coefficient of ΔR_c indicates very important and interesting points. First, it supports the existence of the G.E. link between average wage and all industrial wages in cities in Brazil during the 1990s through which composition of industrial employment in cities has a causal impact on local industrial wages. Second, the magnitude of the estimate indicates that a restructuring of industrial composition in a city favouring high-wage industries that raises the average wage in this city by only 1%[41] through shift-share accounting[42] is estimated to increase all industry wages in the city by at about twice as much (1.6%) in the long-run only through the G.E impacts. In other words, the total increase in average wage from such a change in industrial composition is 2.6%, with 62% of it being only due to the G.E. wage effects of shifts in industrial composition. It is in a way an estimate of average *long-run elasticity of industrial wages with respect to industrial-composition* (due to the G.E. impacts), which according to the results reported in Table (5.2) is about 1.6. Third, the magnitude of the coefficient indicates that the wage spillover associated with a shift in industrial composition is at least one and half times larger than the conventional accounting measure of the associated wage impact of such a shift.

As explained in section 2 under equation (2.1), the conventional accounting approach measures the average wage effect of a shift in industrial composition by keeping wages fixed and allowing only industrial employment shares to change. In other words, it ignores the linkage

[39] In fact, dropping the change in employment rate from specification, does not affect the coefficient on ΔR_c. Testing the equality of the coefficient on ΔR_c when ΔER_{ci} is controlled for and when it is dropped from the specification fails to reject the null at any conventional level of significance (p-value = 0.38). In other words, ΔER_{ci} is orthogonal to ΔR_c.

[40] 'Controlled' here means that the effect of changes in sectoral demand for workers (change in the sectoral employment rates) are taken out of the variation in the variable of interest. For the case of changes in wages, the residuals in the regression of changes in city-sector wages on changes in sectoral employment rates give the controlled version of the change in city-sector wages.

[41] This is about 0.002 units increase in the measure of industrial composition in a typical city in Brazil. The sample mean of the measure of industrial composition over 1991-2000 is 0.25 with a standard deviation of 0.046.

[42] That is, keeping the wages constant.

between local average wage and local industrial wages and the fact that wages could change with a change in average wage. When the wage spillover mechanism is not ignored and equation (2.1) depicts the underlying structural relationship, $[1/(1-\gamma_{c2})] \times A_{ct} = [1+\gamma_{c2}/(1-\gamma_{c2})] \times A_{ct}$ is the total wage impact of a pure[43] shift in industrial composition that changes the average wage in city c by A_{ct} units. It should be noticed that A_{ct} is the conventional accounting measure of the average wage impact of such a shift. To see this more clearly, the average wage effect conventionally associated with a shift in industrial composition in a city can be written as $A_{ct} = \sum_j (\Delta\eta_{cjt+1}) w_{jt}$, where the effect of the change in industrial composition on wages are neglected. If the shift in local industrial composition of employment does not affect industry wages, then A_{ct} measures the total effect of change in industrial composition on average wages. However, if wages are instead determined according to equation (2.1), then the total wage impact of a shift in industrial composition becomes A_{ct} plus the general equilibrium impact of such a shift on average wages, or $A_{ct} + \frac{\gamma_{c2}}{(1-\gamma_{c2})}\Delta R_{ct}$. Given the decomposition of the change in the measure of industrial composition, the total effect can be written as[44]:

$$
\begin{aligned}
A_{ct} + \tfrac{\gamma_{c2}}{(1-\gamma_{c2})}\Delta R_{ct} &= A_{ct} + \tfrac{\gamma_{c2}}{(1-\gamma_{c2})}\left(\sum_j (\Delta\eta_{cjt+1})\varpi_{jt} + \sum_i \eta_{cjt+1}(\Delta\varpi_{jt+1})\right) \\
&= A_{ct} + \tfrac{\gamma_{c2}}{(1-\gamma_{c2})}\cdot\left(A_{ct} + \sum_i \eta_{cjt+1}(\Delta\varpi_{jt+1})\right) \qquad (5.1) \\
&= \left(1 + \frac{\gamma_{c2}}{1-\gamma_{c2}}\right) A_{ct} + \frac{\gamma_{c2}}{(1-\gamma_{c2})} \sum_i \eta_{cjt+1}(\Delta\varpi_{jt+1})
\end{aligned}
$$

Thus, in the case where industry wage premia are constant over time, the total effect becomes $\left(1+\frac{\gamma_{c2}}{1-\gamma_{c2}}\right)A_{ct}$, in which case $\frac{\gamma_{c2}}{1-\gamma_{c2}}$ averaged across cities, that is the coefficient of ΔR_c in estimating equation (3.1)[45], measures how big the wage effect of a shift in industrial composition is, in a general equilibrium setting, relative to its conventional measures.

[43] Pure in the sense that total employment in the city does not change, or $\sum_j \Delta\eta_{cjt+1} = 0$.

[44] Notice that since $\sum_j \Delta\eta_{cjt+1} = 0$, the conventional measure can be written as $A_{ct} = \sum_j (\Delta\eta_{cjt+1})(w_{jt} - w_{1t}) = \sum_j (\Delta\eta_{cjt+1})\varpi_{jt}$

[45] Notice that $(\gamma_2/1-\gamma_2)$ is the average of $\gamma_{c2}/(1-\gamma_{c2})$ across all cities.

Hence, an estimate of $\beta^{OLS} = 1.6$ measures the local wage impact of a shift in local industrial composition to be at least one and a half times larger than the conventional accounting measures. With an estimate of this magnitude, a pure change in local industrial composition[46] that brings about one unit direct impact on average wages in a city (the accounting measure of the impact of changes in industrial composition ignoring the spillover effect and the consequent wage changes) will generate waves of general equilibrium effects on city-industry wages so that by the time the new steady-state equilibrium establishes, average wages increase almost 2.6 folds ($\cong 1 + 1.6$) in total, letting the general equilibrium impact to be responsible for 62% of this change. [47]

5.2 Correcting for Sample Selection Bias

The first pit-hole that should be dealt with is the selection bias. In order to address the issue of selection bias, as discussed in section 3.1, following the approach of Dahl (2002) the probabilities of migration need to be calculated. To do so, first the sample is divided into "movers" and "stayers". Movers are individuals who are now living in a city that is not located in their state of birth. Stayers are individuals who are now living in a city that is part of their state of birth. For movers, mover cells are defined based on some of the attributes of the individuals within this group; three age categories, three education categories, and two gender categories. In total, these categories generate 18 cells for the movers. For the stayers, as well as these groups, additional categories based on two marital status categories are added; single or otherwise. In this way, a total of 36 cells are generated for the stayers. The higher number of cells for the stayers is in accordance with their higher share in the sample. P_{kbc} is defined as the fraction of individuals born in state *b* that are in the same mover cell as person *k* and have moved to city *c*. In a similar way, P_{kbb} is defined as the fraction of individuals born in state *b* that are in the same stayer cell as person *k* and have stayed in the same state.

[46] Pure in the sense that overall employment does not change.

[47] The sample mean of the measure of industrial composition (R_{ct}) has increased from 0.237 in 1991 to 0.264 in 2000 and its standard deviation has decreased from 0.048 to 0.039. This is while the average change in this measure (average ΔR_{ct}) over this period is 0.027 with a minimum of -0.034 and a maximum of 0.089 and a standard deviation of 0.018. Of this mean change in the measure of industrial composition in the sample, in average 0.005 units is associated with shifts in employment share (average ΔR_{ct} calculated keeping wage premia constant, or ΔR_c^1) and 0.023 units is associated with changes in wage premia (average ΔR_{ct} calculated keeping employment shares constant, or ΔR_c^2). One standard deviation improvement in the measure of local industrial composition, R_c, brings about an average 8.3% increase in local wages.

After preparing the migration probabilities, the selection-bias-correcting version of equations used for generating national industry wage premia and city-industry wages, equations (3.8) and (3.9), are employed and corrected estimates of national industry wage premia and corrected city-industry wage levels are estimated. The estimation results of the equations (3.8) and (3.9), which are not presented here, indicate that the correction terms are highly significant, which makes it likely that the previous result under OLS (1) and OLS (2) may actually suffer from selection bias.

Proceeding with the sample-selection-corrected estimates of wage premia and city-industry wages, the third column in Table (5.2) under OLS (3) reports the results of OLS estimation of equation (3.1) after correcting for sample selection bias. The results are very similar to those reported in columns one to three, when no correction for self selection in the sample had been implemented. Thus, it can be concluded that the OLS estimates are not contaminated by the existing self selection in the sample.

5.3 Dealing with Endogeneity

The next is to address the endogeneity of the regressors. To start with constructing instrumental variable, a pre first-stage regression is estimated in which growth rate of employment in each city-industry is estimated[48] based on the logarithm of distance between the city and 15 major international commercial seaports in Brazil.[49] This regression is used to predict the city-industry growth rates of employment that are induced by the distributive impacts of trade opening during this decade across different cities and industries as perceived through their distance from the major international commercial ports[50]. Overall, the IV estimates in this section indicate that the OLS estimates do not suffer from endogeneity.

[48] Approximated by the change in natural logarithm of city-industry employment over the decade 1991-2000.

[49] These major ports are: Port Santos (Santos - SP), Port Vitória (Vitória - ES), Port Paranaguá (Paranaguá - PR), Port Itaguaí (Itaguaí - RJ), Port Rio Grande (Rio Gradne - RS), Port Rio de Janeiro (Rio de Janeiro - RJ), Port Itajaí (Itajaí - SC), Port Itaqui (São Luís - MA), Port São Sebastião (São Sebastião - SP), Port São Francisco do Sul (São Francisco do Sul - SC), Port Aratu (Candeias - BA), Port Manaus (Manaus - AM), Port Suape (Ipojuca - PE), Port Pecém (São Gonçalo do Amarante - CE), Port Ilhéus (Ilhéus - BA). See Figure 1.1 for a map of the ports.

[50] Distance from Argentina-Brazil border was independently used to predict the city-industry growth rates of employment as induced by the distributive impacts of MERCOSOUR. However, instruments based on this measure of distance alone proved to have a poor correlation with the change in the measure of industrial composition across all cities in the first stage. This correlation was essentially too low to allow any reliable use of this instrument. This could be taken as an indication that although as a result of the formation and enactment of MERCOSUR Argentina became an important trade partner to Brazil, the size of trade between the two countries was not large enough to source major changes in industrial composition in all cities across Brazil.

The last three columns in table (5.2) report the results of IV estimations using the instruments explained in section three and presented here again:

$$IV1 = \sum_i (\hat{\eta}_{cit+1} - \eta_{cit})\varpi_{it}$$

$$IV2 = \sum_i \hat{\eta}_{cit+1}(\varpi_{it+1} - \varpi_{it})$$

$$IV_{ER} = \eta_{cit-1} g_i$$

Columns IV (4), IV (5), and IV (6) report the results of applying the distance based instrumental variable strategy to estimating equation (3.1) after correcting for sample selection bias, respectively for using only *IV1* and *IV$_{ER}$* under column IV (4), only *IV2* and *IV$_{ER}$* under column IV (5), and *IV1*, *IV2*, and *IV$_{ER}$* together under column IV (6). The IV results should be compared to the results under OLS (3), where appropriate self-selection correction is implemented and the city-industry employment rate is controlled for. The associated first-stage results are reported in table (5.3).

Column IV (4) in table (5.2) reports the result of estimating equation (3.1) using *IV1* and *IV$_{ER}$* as instrumental variables, for which the results of the first-stage estimations is reported under column IV (4) in table (5.3). The joint redundancy F-test and p-value in the first stage regression of ΔR_c are 98.9 and 0.00 respectively, and those associated with the first stage of ΔER_{ci} are 107 and 0.00 respectively. The results of the second stage are indicative of a highly significant and positive relationship between city-industry wages and the measure of industrial composition, which is of a magnitude that is not significantly different from the OLS estimates[51]. Under the assumptions required for validity of instruments, it seems to be the case that the stronger condition required for the consistency of OLS is satisfied. In other words, the OLS estimates of the relationship between the change in city-industry wages and the change in the measure of industrial composition is consistent and does not suffer from endogeneity. What is interesting is that even though it seems to be the case that the change in the city-industry employment rate is endogenous, as the test for exogeneity of this variable rejects the null (see footnote 51), it does

[51] The test for equality of the OLS and IV estimates is formally carried out in this section by testing for exogeneity of the variable(s) of interest, separately or jointly, through comparing the distance of the OLS and IV estimates from each other via the endogtest(.) option of the ivreg2 command in STATA. The null hypothesis is that the specified endogenous regressor can actually be treated as exogenous. For more information see the help for ivreg2 in STATA. The P-value for the case of the estimates reported under column IV (4) is 0.66 for ΔR_c, 0.001 for ΔER_c, and is 0.004 for ΔR_c and ΔER_c jointly.

not affect the OLS estimates of the relationship between wages and the measure of industrial composition. As it was discussed before (see footnote 39), the change in the measure of industrial composition seems to be orthogonal to the change in employment rate, as a result of which endogeneity of the change in employment rate does not affect the OLS estimates of the coefficient on change in the measure of industrial composition.

Column IV (5) in table (5.2) reports the result of estimating equation (3.1) using *IV2* and IV_{ER} as instrumental variables, for which the results of the first-stage estimations is reported under column IV (5) in table (5.3). The results of the first and second stage are very similar to the previous case. Column IV (6) in table (5.2) reports the result of estimating equation (3.1) using *IV1*, *IV2*, and IV_{ER} as instruments, for which the results of the first-stage estimations is reported under column IV (6) in table (5.3). The results of the first and second stage are similar to the previous cases here too. The test for over-identification fails to reject the null hypothesis (p-value = 0.87), which indicates that the instruments are correctly excluded from the estimating equation.[52]

5.4 Decomposition of Impacts

The estimation results of equation (3.1) using the decomposed change in the measure of industrial composition ($\Delta R_c = \Delta R_c^1 + \Delta R_c^2 \equiv \sum_j (\Delta\eta_{cjt+1})\varpi_{jt} + \sum_i \eta_{cjt+1}(\Delta\varpi_{jt+1})$) are reported in table (5.4). ΔR_c^1 ($\equiv \sum_j (\Delta\eta_{cjt+1})\varpi_{jt}$) captures the change in the measure of local industrial compositions that are based only on shifts in employment shares (changes in industrial compositions), keeping the structure of the wage premia constant. ΔR_c^2 ($\equiv \sum_i \eta_{cjt+1}(\Delta\varpi_{jt+1})$) depicts the change in local industrial composition that are only based on changes in wage premia, keeping the structure of industrial employment constant. The aim is to make sure of the relevancy of what is claimed here to be the impact of shifts in industrial composition on wages. The mechanism and the intuition described here as being the source of the wage spillover effect is the resulting improvement in the outside options of the bargaining, unemployed workers in all sectors within a city as industrial composition shifts in favour of higher paying industries. In principle, the same impact could be observed without any shifts in industrial composition; a change in the structure of industry wage premia that relatively improves the wage premium of

[52] This is a test of the joint null hypothesis that the excluded instruments are valid instruments, i.e., uncorrelated with the error term and correctly excluded from the estimated equation. A rejection casts doubt on the validity of the instruments.

high-share industries in the absence of any changes in the shares of industries can be considered as having similar outside-option-improving effects.

While from a theoretical perspective it does not matter whether the increase in the average wage is initiated by shifts in industry shares or changes in industry wages, from an empirical perspective, and especially from a policy making perspective, it matters. In other words it should matter for the policy maker if there is a difference between the G.E. impacts of an increase in the average wage that is initiated by restructuring of industrial employment and that of one that is initiated by changes in industry wage premia only. If the source of the G.E. wage effect cannot be the shifts in shares of industries, then policy changes that impact the shares will not have any general equilibrium impacts on wages that are channelled through the shifts in shares.

To this end, table (5.4) reports the results of the decomposition of the wage impact of changes in industrial composition. In this table under OLS (1), the results of estimating a decomposed version of equation (3.1) that controls for city-industry changes in employment rate are shown. Under OLS (2), similar results but only controlling for city changes in employment rates is reported. And under OLS (3), the self selection-corrected version of the results under OLS (1) is shown. In all these results, it is only the shift-share changes in the measure of industrial composition that has statistically significant impact on local city-industry wages. This result does not change by controlling for city changes in employment rates (instead of city-industry employment rates) or by correcting for the self selection in the sample. While equality of the coefficients on ΔR_c^1 and ΔR_c^2 cannot be rejected, the latter is not statistically significant. This is while a joint test for zero wage impact for both parts of the decomposition strongly rejects the null. This result indicates that the relevant variation in the measure of industrial composition in the sample, as far as the G.E. wage effects are concerned, comes from the share-based shifts in the measure and what is captured as the coefficient of ΔR_c is the wage impact of a shift in industrial composition (as in ΔR_c^1) .

Column IV (4) in table (5.4) reports the IV estimation results of equation (3.1) in a decomposed fashion while controlling for the change in city-industry employment rate, using *IV1*, *IV2*, and IV_{ER} as instruments for ΔR_c^1, ΔR_c^2, and ΔER_{ci}. Table (5.5) under column IV (5) reports the results of the associated first-stage regressions. The first stage results are fairly satisfying and the results of the second stage suggest that, similar to the OLS results, only the shares-based changes in the measure of industrial composition significantly and positively impacts local city-industry wages.[53] The test for equality of the coefficients on the two parts of the decomposition fails to reject the null and this is while the test for both coefficients being zero jointly is strongly rejected.[54]

5.5 Robustness

It remains to make sure that the estimates of the impact of shifts in industrial composition on city-industry wages are robust to introduction of alternative explanations for differences in wage growth across cities. Specifically, these alternative explanations include diversity of employment in a city (Glaeser, Kallal, Scheinkman, and Shleifer, 1992) and education levels (Moretti, 2004; Acemoglu and Angrist, 1999). Additional variables representing these alternatives are added to equation (3.1) to ensure of the robustness of previous estimates. The results are shown in table (5.6), and in table (5.8) for the decomposed changes in the measure of industrial composition. The coefficient of the measure of industrial composition is fairly stable and remains highly significant after the introduction of the new variables one at a time or altogether at once. Among the new controls only the diversity of employment in a city is highly significant, which positively affects city-industry wages.

Glaeser et al. (1992) examine predictions of various theories of growth externalities (knowledge spillovers) within and between industries at city level in the U.S. during 1956 and 1987. They try to verify whether it is the geographic specialization or competition of geographically proximate industries that promote innovation spillovers and growth in those industries and cities. One measure of city growth they use is growth in wages. By testing empirically in which cities industries grow faster, as a function of geographic specialization and competition, they find that specialization has no effect on wage growth and diversity in a city helps the wage growth of the industries. Here, following Beaudry et al. (2009), a measure of "diversification" of employment in each city at the start of the decade measured by one minus the Herfindahl index, or one minus sum of squared-industry-shares in the city, is introduced. The results are reported under columns OLS (1) and IV (1) in table (5.6), and table (5.8) for the decomposed changes in measure of industrial composition. The change in the measure of industrial composition in table (5.6) and the share-based change in the measure of industrial

[53] The P-value for endogeneity test (that the specified endogenous regressors can actually be treated as exogenous) of ΔR_c^1 is 0.68, of ΔR_c^2 is 0.04, and of ΔER_c is 0.00.

[54] The P-value for the null hypothesis of having equal IV estimates for the coefficients of ΔR_c^1 and ΔR_c^2 is 0.87, and for them to be jointly zero is 0.006.

composition in table (5.8) are both robust to introducing this alternative explanation for growth in wages at the city level.

Notice that with inclusion of the measure of diversification in employment, both the OLS and IV estimates of the coefficient on ΔR_c in table (5.6) and the coefficient on ΔR_c^1 in table (5.8) become larger. The measure of industry diversification is significant and positive both in OLS and after using instruments, which confirms the finding of Glaeser et al. (1992) for the case of Brazilian cities. It is indicating that cities with a more diversified composition of industrial employment in 1991 – that is a composition of employment in which industries have relatively similar shares employment – experienced higher growth rates in industry wages during 1991-2000 due to more competition between different industries. Notice that an increase in the measure of industrial composition (ΔR_c or ΔR_c^1) is the result of a move away from a diversified composition of employment toward a polar composition in favour of high-paying industries. In other words, ΔR_c or ΔR_c^1 are negatively correlated with the measure of diversification in 1991 (see Graphs (1.5), (1.6), and (1.7)). As a result, given the positive impact of both measures on growth in industry wages and given the negative correlation between the two measures, not controlling for the measure of diversification dampens the coefficient on change in the measure of industrial composition. In other words, the estimates of the G.E. wage impacts of shifts in industrial composition presented in tables (5.6) and (5.8) are more precise than the estimates presented in previous tables.

Moretti (2004) examines wages in U.S. cities in the 1980s and finds that cities with greater increase in the proportion of workers with a BA or higher education have higher wage gains. Acemoglue and Angrist (1999) find weaker results for the impact of education using average years of education in a state. Because here it is already controlled for the level of education in estimating the industry wage premia and therefore, the measure of industrial composition does not reflect cities with higher wages due to having higher levels education. However, it will be controlled for both measures of education discussed in the two studies mentioned above; one measure is the change in the proportion of workers with a BA or higher education and the other is using average years of schooling as an alternative measure of the education level of a city. The results are shown under columns OLS (2) and IV (2) in tables (5.6) and (5.8), where initial levels of workers with a Bachelor or higher and initial average year of schooling among workers are used as instruments in addition to the instruments used so far. The change in the measure of industrial composition in table (5.6) and the share-based change in the measure of industrial composition in table (5.8) are both robust to introduction of these variables. Neither of the new variables is significant after instrumenting, which is similar to the results in Acemoglue and Angrist (1999).

Finally, the last two columns in tables (5.6) and (5.8) under columns OLS (4) and IV (4) report the results of the robustness tests, introducing all the alternative explanations discussed above at once. The wage impact of change in local industrial compositions remains intact.

6 Conclusion

Conventionally, it is presumed that restructurings of industrial composition do not have important impacts on average wages other than the direct shift-share impact through changes in industries' employment shares (see Bound and Johnson, 1992). Average wage in a city is the local-employment-share weighted sum of local industry wages. Keeping local wages constant, an increase in the share of a high-paying industry will change the average wage figure only by the *wage premium* paid in that industry multiplied by the change in its share. As a result, in the event of a 5% increase in the share of an industry that pays 20% premium relative to average in other industries, the shift-share accounting predicts that it results in only 1% increase in the average wage. This is because an increase in an industry's share of local employment is a decrease in other industries' shares. As a result, the net direct impact on average wage from a shift in industrial composition is deemed to be modest. However, in a general equilibrium search and bargaining model of a labour market, the average wage can play the role of an outside option for the bargaining unemployed workers in all industries (see Beaudry, Green, and Sand, 2009), so that even a modest increase in average wage results in wages in all sectors to increase, which will bring about a further rise in average wage that will impact all industrial wages again through the same bargaining mechanism, and so on and so forth. This cycle will continue until it eventually dies off and new steady-state equilibrium takes shape. As a result of these cycles, the G.E. wage impacts of a shift in industrial composition could turn out to be large even though the initial direct impact on average wage may be small.

Contrary to the conventional presumption, this paper brings substantial evidence indicating that, aside from the wage impact of changes in within-industry labour demand, a shift in industrial composition has sizable between-industry impacts on wages through the general equilibrium (G.E.) feedbacks from the average wage onto *all* industrial wages. Industrial composition in a city is measured here as local-employment-share weighted sum of national industrial wage premia. Using Brazilian census data for years 1991 and 2000, this paper empirically identifies the G.E. wage impact of exogenous shifts in industrial composition of cities in Brazil during the 1990s that were brought about by substantial international trade liberalization in this country during the decade.

A restructuring of industrial composition in a city favouring high-wage industries that raises the average wage in this city by only 1%[55] through shift-share accounting[56] is estimated to increase all industry wages in the city by at least twice as much – between 2 to 4 percent – in the long-run only through the G.E impacts. In other words, the total increase in average wage from such a change in industrial composition is at least 3% – between 3 to 5 percent – with 67% of it being only due to the G.E. wage effects of shifts in industrial composition. The G.E. impacts are interpreted as spillovers from high-wage industries to other industries within a city; i.e., a shift in local industrial composition favouring higher paying industries improves the chances of getting hired in those industries for the wage-bargaining unemployed workers in the city, which results in higher wages being paid within each skill group by all industries in that locality.

It should be emphasized that the G.E. wage impacts discussed here are different from the wage effects of changes in within-industry demand for labour. A shift in industrial composition is made up of changes in within-industry labour demands. However, the within-industry changes in labour demand are expected to only affect the within-industry wages. What is identified here is the between-industry spillover of wages. An industrial composition shifts in favour of high-wage industries increases the average wage at first only through the shift-share effect. Such an increase in the average wage is then transmitted to all other industries through the G.E. mechanism in which the average wage plays the role of a reservation wage. In other words, through this G.E. mechanism, the high wages paid by the high-wage industries spill over onto all other industries. As a result, what is identified here is the between-industry spillover of wages and in terms of the empirics these are the estimated wage impacts associated with changes in industrial composition while keeping the city-industry or city level labour demands constant.[57]

The estimates of the relationship of interest are shown to be robust to correcting for the likely endogeneities discussed above and the sample selection bias that is caused by the migration of workers across cities within Brazil. Although it is verified that the sample suffers from selection

[55] This is about 0.002 units increase in the measure of industrial composition in a typical city in Brazil. The sample mean of the measure of industrial composition over 1991-2000 is 0.25 with a standard deviation of 0.046.
[56] That is, keeping the wages constant.
[57] In the empirics, this is done by controlling for city-industry or city employment rates.

bias, after correcting for it according to the approach in Dahl (2002), the estimates of the effects of changes in industrial compositions on wages remain significant and do not significantly change in magnitude. The findings are also shown to be robust in significance and size to the introduction of other alternative explanations for differences in wage changes across cities such as those related to diversity of employment in a city (Glaeser, Kallal, Scheinkman, and Shleifer, 1992), and levels of education (Moretti, 2004; Acemoglu and Angrist, 1999). It is shown here that diversification in composition of employment is the only alternative explanatory mechanism that is statistically significant (and positive) both in OLS and after using instruments, which confirms the finding of Glaeser et al. (1992) for the case of Brazilian cities.

Controlling for the measure of diversification in composition of employment increases the magnitude of the OLS and IV estimates of the G.E. wage effects associated with shifts in industrial composition. The diversification of employment in each city is measured here as the start of the decade, in 1991, by one minus the Herfindahl index, or one minus sum of squared-industry-shares in the city. The results indicate that cities with a more diversified composition of industrial employment in 1991 – that is a composition of employment in which industries have relatively similar shares in total employment – experienced higher growth rates in industry wages during 1991-2000 due to more competition between different industries. Since an increase in the measure of industrial composition (ΔR_c or ΔR_c^1) is a move away from a diversified composition of employment toward a polar composition in favour of high-paying industries (see graphs (1.5), (1.6), and (1.7)), not controlling for the measure of diversification dampens the coefficient on change in the measure of industrial composition. In other words, the estimates of the G.E. wage impacts of shifts in industrial composition that controls for the effect of diversification are (presented as the coefficient on ΔR_c in table (5.6) and on ΔR_c^1 in table (5.8)) are more precise estimates of the relationship of interest here.

The finding of this paper is important in two respects. Firstly, the sizable G.E. wage impact associated with shifts in industrial composition suggests that ex-ante evaluation of trade or industrial policies should not be carried out based on partial equilibrium assumptions and analysis (such as shift-share accounting here). Instead a proper evaluation of the impacts of such policy changes should consider general equilibrium interactions such as the one found here. Secondly, the finding of this paper highlights the role that changes in national trade or industrial policies play in creating (geographic) winners and losers depending on how the impact of such policy changes are transmitted sub-nationally to different regions. A change in the policies may induce different patterns of shift in industrial composition in different regions – favouring high-paying industries in some and low-paying industries in others. Given the sizable G.E. wage impacts found here, an un-even pattern of shifts in composition of employment could significantly contribute to the worsening of regional wage disparities and formation of wide spatial wage gaps. Both of these aspects are especially important in developing countries given that they are relatively more prone to major policy changes and sub-nationally egalitarian spatial economic development is of major importance to their overall developmental progress. Realizing that in most developing countries already the distribution of economic development across localities is non-uniform, such distribution of national policy impacts could especially be worrying if the losers-to-be regions are the already less-developed ones.

In the case of Brazil, the pattern of shifts in local industrial compositions in fact helped reduce the regional wage gap between the poor and rich areas in this country, probably due to the fact that Brazil benefits from geographically widespread major international commercial seaports (see Figure 1.1). Drawing on the finding of economic geography models of trade[58] (see Hanson, 2005; Redding and Venables, 2004; Head and Meyer, 2006; Knaap, 2006; Hering and Poncet, 2008; Mion and Naticchioni, 2005; Combes et al., 2008; Lederman et al., 2004; Da Mata et al., 2005; Fally et al., 2010) the distance of cities from these ports is effectively their distance from trade partners of Brazil and should partly determine how trade liberalization in Brazil during the 1990s were distributed across different cities and regions. Thus, even though the Northeast region was still the poorest in Brazil at the end of the 1990s[59] (see Figure 1.2) given the high share of agricultural employment in this region (see Graph 1.1), probably due to the existence of major seaports in this region and the fact that the northeast tip of Brazil on the side of the Atlantic ocean is the closest to the major trade partner of Brazil in North America, Europe, and Asia (see Figures 1.1, 1.2, and 1.3), majority of cities in this region benefited from shifts in their industrial compositions that favoured high-paying industries. As a result, not only the average wages increased in majority of these cities due to the favourable shifts in industrial composition, but because of adverse shifts in the composition of employment in cities in the richest region of Brazil in Southeast, the spatial wage gap between these two densely populated extremes in fact

---

[58] That distance as a trade barrier determines the size of trade between two economies.

[59] A World Bank report calls the Northeast region the region with the "... most remaining income poverty ..." in Brazil (World Bank, 2001, p. 1). In a study on the evolution of the regional GDP's 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, was 7.2 times that of Piauí, the poorest North Eastern state (Lall et al., 2004).

shrank during the 1990s (see Graphs 1.1, 1.2, 1.3, and 1.4). Of course, as is clear from Graph (1.5), there were also exceptions to this rule in both regions among cities.

Comparing the case of Brazil with Mexico is helpful here. Mardukhi (2009) studies the same G.E. wage impacts discussed here for the case of Mexican cities during the 1990s, the decade in which Mexico became a member of NAFTA and substantially increased its trade relations with the US. He estimates the G.E. wage impacts of shifts in industrial composition in Mexican cities to be almost twice as much as what is estimated here for Brazilian cities. He further finds that the pattern of shifts in compositions was inversely related to distance from Mexico-US border so that in South of Mexico, far away from the US border, the composition of employment shifted in favour of the low-paying industries while in the northern regions employment shifted toward the high-paying industries. Given that most southern cities in Mexico were already among the poorest in Mexico, he finds that the geographic losers of the trade liberalization of the 1990s in Mexico – as far as the impacts on industrial compositions and consequently wages are concerned – were regions that already were suffering from lower levels of economic development. This regressive spatial pattern of shifts in industrial compositions created a north-south disparity in growth rates of wages and helped increase the spatial wage gap in Mexico. In Brazil, however, due to a progressive pattern of shifts in compositions of employment across different regions, the wage impacts of trade policy-induced shifts in industrial composition were also progressively distributed spatially among regions and helped reduce the spatial wage gap between the poorest and richest regions in the country during 1991-2000.

References

Acemoglu, D., and J. Angrist (1999). “How Large are the Social Returns to Education? Evidence from Compulsory Schooling Laws”. *NBER Working Paper No. 7444* .

Amaral, E. F. L., D. S. Hamermesh, J. E. Potter, and E. L. G. Rios-Neto. (2007). “Demographic Change and the Structure of Wages: A Demand-Theoretic Analysis for Brazil”. NBER Working Paper 13533.

Beaudry, P., D. Green, and B. Sand. (2009). "Does Industrial Composition Matter? An empirical evaluation based on Search and barganing theory". University of British Columbia.

Blanchard, O. J., and L. F. Katz. (1992). "Regional Evolutions". *Brookings Papers and Economic Activity,* (1), pp. 1-75.

Bound, J., and G. Johnson. (1992). "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations". *American Economic Review* , pp. 371 - 392.

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

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

Dahl, G. B. (2002). "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets". *Econometrica , 70* (6), pp. 2367-2420.

Da Mata, D., U. Deichmann, J. V. Henderson, S. V. Lall, and H. G. Wang. (2007). “Determinants of City Growth in Brazil”. Journal of Urban Economics, 62(2007), p. 252-272.

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. (2010). "Economic geography and wages in Brazil: Evidence from Micro-data". *Journal of Development Economics , 91*, p. 155-168.

Glaeser, E., H. Kallal, J. Scheinkman, and A. Shleifer. (1992). “Growth in Cities”. *Journal of Political Economy , 100* (6), pp.1126-1152.

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

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

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), p. 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.

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

Manski, C. F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem". *Review of Economic Studies , 60* (3), pp. 531-42.

Mardukhi, J. (2009). "Impacts of Shifts in Industrial Composition on Wages in Developing Countries: The Case of Mexican Cities and NAFTA". University of British Columbia.

Ministry of External Relations. (2008). "An Ocean of Opportunities". *Brazil Now*, November.

Minnesota Population Center. (2008). Integrated Public Use Microdata Series [IPUMS] — International: Version 4.0. Minneapolis, University of Minnesota.

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

Moffitt, R. A. (2001). "Policy Interventions, Low-Level Equilibria And Social Interactions". In S. N. Durlauf, & H. P. Young (Eds.), *Social Dynamics* (pp. 45-82). Washington, D.C.: Brookings Institution Press.

Moretti, E. (2004). "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data". *Journal of Econometrics , 121* (1-2), pp. 175-212.

Mossi, M., P. Aroca, I. Fernandez, and C. Azzoni. (2003). "Growth Dynamics and Space in Brazil". *International Regional Science Review , 26* (3), p. 393-418.

Potter, Joseph E., Carl P. Schmertmann, and Suzana M. Cavenaghi. (2002). "Fertility and Development: Evidence from Brazil". *Demography , 39* (4), pp. 739-761.

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

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

World Bank. (2001). Attacking Brazil's Poverty: A Poverty Report with a Focus on Urban Poverty Reduction Policies. *Report No. 20475-BR* .

Table (5.2) – OLS and IV Estimation Results of Equation (3.1)

Dependent Variable: $\Delta ln(w_{ci})$	OLS (1)	OLS (2)	OLS (3)
$\Delta \boldsymbol{R_c}$	1.62^{***} (0.44)	1.58^{***} (0.43)	1.59^{***} (0.43)
$\Delta \boldsymbol{ER_{ci}}$	0.16 (0.17)	–	0.16 (0.16)
$\Delta \boldsymbol{ER_c}$	–	0.21 (0.16)	–
Industry Fixed Effects $(\boldsymbol{d_i})$	Yes	Yes	Yes
Corrected for Sample Selection Bias	No	No	Yes
Obs.	6,213	6,213	6,213
R^2	0.08	0.08	0.08
Over-id Test	–	–	–

(.): Robust, city-clustered standard deviation. $***$, $**$, $*$: Respectively, significance at 1%, 5%, and 10% levels of significance. **OLS (1):** OLS estimation of changes in city-industry wages on changes in city measure of industrial composition and changes in city-industry employment rate, controlling for industry fixed effects (equation (3.1)). **OLS (2):** Same specification as OLS (1) but controlling for changes in city employment rate instead of changes in city-industry employment rate. **OLS (3):** Same specification as OLS (1) and correcting for sample selection bias.

Table (5.2) – Continued

Dependent Variable: $\Delta ln(w_{ci})$	IV (4)	IV (5)	IV (6)
$\Delta \boldsymbol{R_c}$	1.80^{**} (0.83)	2.01^{**} (0.92)	1.86^{***} (0.62)
$\Delta \boldsymbol{ER_{ci}}$	0.91^{***} (0.28)	0.88^{***} (0.29)	0.91^{***} (0.27)
$\Delta \boldsymbol{ER_c}$	–	–	–
Industry Fixed Effects $(\boldsymbol{d_i})$	Yes	Yes	Yes
Corrected for Sample Selection Bias	Yes	Yes	Yes
Obs.	6,213	6,213	6,213
R^2	–	–	–
Over-id Test	–	–	P-value = 0.87

(.): Robust, city-clustered standard deviation. $***$, $**$, $*$: Respectively, significance at 1%, 5%, and 10% levels of significance. **IV (4):** IV estimation associated with OLS (3) using IV1 and IV_{ER}. **IV (5):** IV estimation associated with OLS (3) using IV2 and IV_{ER}. **IV (6):** IV estimation associated with OLS (3) using IV1, IV2, and IV_{ER}. $IV1 = \sum_i(\hat{\eta}_{cit+1} - \eta_{cit})\varpi_{it}$ and $IV2 = \sum_i \hat{\eta}_{cit+1}(\varpi_{it+1} - \varpi_{it})$.

Table (5.3) – First Stage Results Associated with Specifications in Table (5.2)

1st Stage Associated with:	IV (4)		IV (5)	
Endogenous Variable:	ΔR_c	ΔER_{ci}	ΔR_c	ΔER_{ci}
IV1	1.16^{***} (0.16)	0.05 (0.04)	–	–
IV2	–	–	1.48^{***} (0.19)	0.17^{***} (0.05)
$\boldsymbol{IV_{ER}}$	-0.01^{***} (0.00)	-0.13^{***} (0.01)	-0.02^{***} (0.00)	-0.13^{***} (0.01)
$\boldsymbol{d_i}$	Yes	Yes	Yes	Yes
R^2	0.45	0.56	0.21	0.56
Partial R^2 for Excl. Instr. ♣	0.44	0.29	0.20	0.30
Joint Redundancy Test for Excl. Instr. ♦	F(2,434) = 98.9 P-value = 0.00	F(2,434) = 107 P-value = 0.00	F(2,434) = 73.1 P-value = 0.00	F(2,434) = 110 P-value = 0.00

(.): Robust, city-clustered standard deviation. ***, **, *: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. $IV1 = \sum_i (\hat{\eta}_{cit+1} - \eta_{cit})\varpi_{it}$ and $IV2 = \sum_i \hat{\eta}_{cit+1}(\varpi_{it+1} - \varpi_{it})$.

Table (5.3) – Continued

1st Stage Associated with:	IV (6)	
Endogenous Variable:	ΔR_c	ΔER_{ci}
IV1	1.10^{***} (0.10)	0.04 (0.04)
IV2	1.31^{***} (0.11)	0.17^{***} (0.05)
$\boldsymbol{IV_{ER}}$	-0.01^{***} (0.00)	-0.13^{***} (0.01)
$\boldsymbol{d_i}$	Yes	Yes
R^2	0.59	0.56
Partial R^2 for Excl. Instr. ♣	0.59	0.30
Joint Redundancy Test for Excl. Instr. ♦	F(3,434) = 237 P-value = 0.00	F(3,434) = 76.6 P-value = 0.00

(.): Robust, city-clustered standard deviation. ***, **, *: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. $IV1 = \sum_i (\hat{\eta}_{cit+1} - \eta_{cit})\varpi_{it}$ and $IV2 = \sum_i \hat{\eta}_{cit+1}(\varpi_{it+1} - \varpi it$.

Table (5.4) – Decomposing $\Delta R_{ct} = \Delta R^1_{ct} + \Delta R^2_{ct}$

Dependent Variable: $\Delta ln(w_{ci})$	**OLS (1)**	**OLS (2)**	**OLS (3)**
$\Delta \boldsymbol{R}^1_{ct} = \sum_i \Delta \boldsymbol{\eta}_{cit+1} \boldsymbol{v}_{it}$	1.71*** (0.50)	1.66*** (0.50)	1.67*** (0.49)
$\Delta \boldsymbol{R}^2_{ct} = \sum_i \boldsymbol{\eta}_{cit+1} \Delta \boldsymbol{v}_{it+1}$	0.81 (1.70)	0.78 (1.71)	0.88 (1.70)
$\Delta \boldsymbol{ER}_{ci}$	0.16 (0.16)	–	0.16 (0.16)
$\Delta \boldsymbol{ER}_{c}$	–	0.21 (0.16)	–
Industry Fixed Effects (d$_i$)	Yes	Yes	Yes
Corrected for Sample Selection Bias	No	No	Yes
Obs.	6213	6213	6213
R^2	0.08	0.08	0.08
Test if: Coef ΔR^1_{ct} = Coef ΔR^2_{ct}	P-val. = 0.63	P-val. = 0.65	P-val. = 0.68
Test if: Coef ΔR^1_{ct} = Coef ΔR^2_{ct} = 0	P-val. = 0.001	P-val. = 0.001	P-val. = 0.001

(.): Robust, city-clustered standard deviation. *****, **, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. **OLS (1):** OLS estimation of changes in city-industry wages on the decomposition of changes in city measure of industrial composition and changes in city-industry employment rate, controlling for industry fixed effects (equation (3.1)). **OLS (2):** Same specification as OLS (1) but controlling for changes in city employment rate instead of changes in city-industry employment rate. **OLS (3):** Same specification as OLS (1) and correcting for sample selection bias.

Table (5.4) – Continued

Dependent Variable: $\Delta ln(w_{ci})$	**IV (4)**
$\Delta \boldsymbol{R}^1_{ct} = \sum_i \Delta \boldsymbol{\eta}_{cit+1} \boldsymbol{v}_{it}$	1.79** (0.87)
$\Delta \boldsymbol{R}^2_{ct} = \sum_i \boldsymbol{\eta}_{cit+1} \Delta \boldsymbol{v}_{it+1}$	2.20 (1.98)
$\Delta \boldsymbol{ER}_{ci}$	0.90*** (0.27)
$\Delta \boldsymbol{ER}_{c}$	–
Industry Fixed Effects (d$_i$)	Yes
Corrected for Sample Selection Bias	Yes
Obs.	6213
R^2	–
Test if: Coef ΔR^1_{ct} = Coef ΔR^2_{ct}	P-val. = 0.87
Test if: Coef ΔR^1_{ct} = Coef ΔR^2_{ct} = 0	P-val. = 0.006

(.): Robust, city-clustered standard deviation. *****, **, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. **IV (4):** IV estimation associated with OLS (3) using IV1, IV2, and IV_{ER}. $IV1 = \sum_i (\hat{\eta}_{cit+1} - \eta_{cit}) \varpi_{it}$ and $IV2 = \sum_i \hat{\eta}_{cit+1} (\varpi_{it+1} - \varpi_{it})$.

Table (5.5) – First Stage Results Associated with Specifications in Table (5.4)

1st Stage Associated with:	IV (4)		
Endogenous Variable:	ΔR_c^1	ΔR_c^2	ΔER_{ci}
IV1	1.11*** (0.10)	-0.01 (0.02)	0.04 (0.04)
IV2	0.52*** (0.11)	0.79*** (0.02)	0.17*** (0.05)
$\boldsymbol{IV_{ER}}$	-0.01*** (0.00)	-0.00 (0.00)	-0.13*** (0.01)
Industry Fixed Effects (d_i)	Yes	Yes	Yes
R^2	0.51	0.82	0.56
Partial R^2 for Excl. Instr. ♣	0.51	0.82	0.30
Joint Redundancy Test for Excl. Instr. ♦	F(3,434) = 129 P-value = 0.00	F(3,434) = 597 P-value = 0.00	F(3,434) = 76.6 P-value = 0.00

(.): Robust, city-clustered standard deviation. *****, **, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. $IV1 = \sum_i (\hat{\eta}_{cit+1} - \eta_{cit})\varpi_{it}$ and $IV2 = \sum_i \hat{\eta}_{cit+1}(\varpi_{it+1} - \varpi_{it})$.

Table (5.6) – Robustness Checks Associated with Table (5.2)○

Dependent Variable: $\Delta ln(w_{ci})$	OLS (1)	IV (1)	OLS (2)	IV (2)
$\boldsymbol{\Delta R_c}$	2.32*** (0.44)	3.24*** (0.96)	1.41*** (0.40)	1.93** (0.93)
$\boldsymbol{\Delta ER_{ci}}$	0.14 (0.16)	0.67** (0.28)	0.12 (0.16)	0.92*** (0.28)
1 – Herfindahl	0.26*** (0.08)	0.34*** (0.10)	–	–
ΔBA +	–	–	-1.17e-06 (1.75e-06)	5.15e-07 (2.76e-06)
$\boldsymbol{\Delta Ave.yrs.schl.}$	–	–	–	–
Industry Fixed Effects (d_i)	Yes	Yes	Yes	Yes
Corrected for Sample Selection Bias	Yes	Yes	Yes	Yes
Instrumented for the new variable	–	No	–	Yes
Obs.	6213	6213	6213	6213
R^2	0.09	–	0.08	–

○: Associated with the results under OLS (3) in table (5.2), which were corrected for sample selection bias. **(.)**: Robust city-clustered standard deviation. *****, **, *:** Respectively, significance at 1%, 5%, and 10% levels of significance. ***1 – Herfindahl***: A measure of diversification of employment in each city at the beginning of the decade studied. **ΔBA +**: The change in the number of people with a university degree in city. $\boldsymbol{\Delta Ave.yrs.schl.}$: Change in the average years of schooling in a city.

Table (5.6) – Continued

Dependent Variable: $\Delta ln(w_{ci})$	**OLS (3)**	**IV (3)**	**OLS (4)**	**IV (4)**
$\boldsymbol{\Delta R_c}$	1.91*** (0.40)	2.44** (1.06)	2.24*** (0.40)	3.44*** (1.11)
$\boldsymbol{\Delta ER_{ci}}$	0.04 (0.16)	0.87*** (0.29)	-0.02 (0.16)	0.68** (0.29)
1 – Herfindahl	–	–	0.22*** (0.08)	0.34*** (0.11)
ΔBA +	–	–	-1.76e-06 (1.30e-06)	6.62e-07 (2.71e-06)
$\boldsymbol{\Delta Ave.yrs.schl.}$	-0.06*** (0.02)	-0.07 (0.04)	-0.06*** (0.02)	0.00 (0.04)
Industry Fixed Effects $(\boldsymbol{d_i})$	Yes	Yes	Yes	Yes
Corrected for Sample Selection Bias	Yes	Yes	Yes	Yes
Instrumented for the new variable	–	Yes	–	Yes
Obs.	6213	6213	6213	6213
R^2	0.10	–	0.20	–

○: Associated with the results under OLS (3) in table (5.2), which were corrected for sample selection bias. **(.)**: Robust city-clustered standard deviation. *****, **, *:** Respectively, significance at 1%, 5%, and 10% levels of significance. ***1 – Herfindahl***: A measure of diversification of employment in each city at the beginning of the decade studied. **ΔBA +**: The change in the number of people with a university degree in city. $\boldsymbol{\Delta Ave.yrs.schl.}$: Change in the average years of schooling in a city.

Table (5.7) – First Stage Results Associated with Specifications in Table (5.6)

1st Stage Associated with:	**IV (1)**	
Endogenous Variable:	ΔR_c	ΔER_{ci}
IV1	0.97*** (0.14)	0.02 (0.05)
$\boldsymbol{IV_{ER}}$	-0.01*** (0.00)	-0.13*** (0.01)
1 – Herfindahl	-0.04*** (0.01)	-0.00 (0.00)
Industry Fixed Effects $(\boldsymbol{d_i})$	Yes	Yes
R^2	0.57	0.56
Partial R^2 for Excl. Instr. ♣	0.32	0.29
Joint Redundancy Test for Excl. Instr. ♦	F(2,434) = 65.2 P-value = 0.00	F(2,434) = 109 P-value = 0.00

(.): Robust, city-clustered standard deviation. *****, **, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. ***1 – Herfindahl***: A measure of diversification of employment in each city at the beginning of the decade studied.

Table (5.7) – Continued

1st Stage Associated with:	IV (2)		
Endogenous Variable:	ΔR_c	ΔER_{ci}	$\Delta \mathbf{BA}+$
IV1	1.11*** (0.15)	0.02 (0.04)	-67555.6** (25991.5)
IV_{ER}	-0.01*** (0.00)	-0.13*** (0.01)	1536.5*** (481.9)
BA+ (1991)	-4.78e-08** (1.93e-08)	-2.22e-08*** (8.23e-09)	0.25*** (0.04)
Industry Fixed Effects (d_i)	Yes	Yes	Yes
R^2	0.45	0.56	0.90
Partial R^2 for Excl. Instr. ♣	0.44	0.30	0.90
Joint Redundancy Test for Excl. Instr. ♦	F(3,434) = 136 P-value = 0.00	F(3,434) = 81.7 P-value = 0.00	F(3,434) = 40.3 P-value = 0.00

(.): Robust, city-clustered standard deviation. *****, ****, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. $\Delta \mathbf{BA}+$: The change in the number of people with a university degree in city.

Table (5.7) – Continued

1st Stage Associated with:	IV (3)		
Dependent Variable:	ΔR_c	ΔER_{ci}	$\Delta \boldsymbol{Ave.yrs.schl.}$
IV1	0.93*** (0.14)	0.03 (0.05)	-4.28 (2.98)
IV_{ER}	-0.01*** (0.00)	-0.13*** (0.01)	0.08* (0.04)
Ave. Yrs. Schl. (1991)	-0.00*** (0.00)	-0.00 (0.00)	-0.17*** (0.02)
Industry Fixed Effects (d_i)	Yes	Yes	Yes
R^2	0.48	0.56	0.27
Partial R^2 for Excl. Instr. ♣	0.47	0.29	0.27
Joint Redundancy Test for Excl. Instr. ♦	F(3,434) = 86.0 P-value = 0.00	F(3,434) = 78.1 P-value = 0.00	F(3,434) = 50.7 P-value = 0.00

(.): Robust, city-clustered standard deviation. *****, ****, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. $\Delta \boldsymbol{Ave.yrs.schl.}$: Change in the average years of schooling in a city.

Table (5.7) – Continued

1st Stage Associated with:	IV (4)			
Dependent Variable:	ΔR_c	ΔER_{ci}	$\Delta \mathbf{BA}+$	$\Delta \boldsymbol{Ave.yrs.schl.}$
IV1	0.87*** (0.13)	0.01 (0.05)	-36207.7 (21120.1)	-3.22 (2.81)
$\boldsymbol{IV_{ER}}$	-0.01*** (0.00)	-0.13*** (0.01)	1171.6*** (339.3)	0.12*** (0.04)
1 – Herfindahl	-0.03*** (0.01)	-0.01 (0.00)	-3082.5** (1312.0)	0.64** (0.29)
BA+ (1991)	-3.01e-08* (1.72e-08)	--2.44e-08*** (8.42e-09)	0.24*** (0.04)	1.32e-07 (1.30e-06)
Ave. Yrs. Schl. (1991)	-0.00*** (0.00)	0.00 (0.00)	605.6*** (205.1)	-0.20*** (0.02)
Industry Fixed Effects $(\boldsymbol{d_i})$	Yes	Yes	Yes	Yes
R^2	0.49	0.56	0.91	0.28
Partial R^2 for Excl. Instr. ♣	0.33	0.30	0.90	0.24
Joint Redundancy Test for Excl. Instr. ♦	F(4,434) = 49.4 P-value = 0.00	F(4,434) = 59.3 P-value = 0.00	F(4,434) = 33.3 P-value = 0.00	F(4,434) = 31.6 P-value = 0.00

(.): Robust, city-clustered standard deviation. *****, **, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. ***1 – Herfindahl***: A measure of diversification of employment in each city at the beginning of the decade studied. $\Delta \mathbf{BA}+$: The change in the number of people with a university degree in city. $\Delta \boldsymbol{Ave.yrs.schl.}$: Change in the average years of schooling in a city.

Table (5.8) – Robustness Checks Associated with Table (5.4)○

Dependent Variable: $\Delta ln(w_{ci})$	OLS (1)	IV (1)	OLS (2)	IV (2)
$\Delta \boldsymbol{R_c^1}$	2.42*** (0.51)	3.23*** (0.90)	1.46*** (0.47)	1.93** (0.99)
$\Delta \boldsymbol{R_c^2}$	1.43 (1.61)	2.15 (1.78)	0.97 (1.73)	2.09 (2.02)
$\Delta \boldsymbol{ER_{ci}}$	0.15 (0.16)	0.70*** (0.27)	0.13 (0.16)	0.92*** (0.27)
1 – Herfindahl	0.26*** (0.08)	0.33*** (0.09)	–	–
$\Delta \boldsymbol{BA} +$	–	–	-1.13e-08 (1.78e-06)	5.09e-07 (2.79e-06)
$\Delta \boldsymbol{Ave.yrs.schl.}$	–	–	–	–
Industry Fixed Effects $(\boldsymbol{d_i})$	Yes	Yes	Yes	Yes
Corrected for Sample Selection Bias	Yes	Yes	Yes	Yes
Instrumented for the new variable	–	No	–	Yes
Obs.	6213	6213	6213	6213
R^2	0.09	–	0.08	–

○: Associated with the results under OLS (3) in table (5.4), which were corrected for sample selection bias. **(.)**: Robust city-clustered standard deviation. *****, **, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. ***1 – Herfindahl***: A measure of diversification of employment in each city at the beginning of the decade studied. $\Delta \mathbf{BA} +$: The change in the number of people with a university degree in city. $\Delta \boldsymbol{Ave.yrs.schl.}$: Change in the average years of schooling in a city.

Table (5.8) – Continued

Dependent Variable: $\Delta ln(w_{ci})$	OLS (3)	IV (3)	OLS (4)	IV (4)
$\boldsymbol{\Delta R_c^1}$	2.07*** (0.46)	2.43** (0.97)	2.38*** (0.49)	3.41*** (1.02)
$\boldsymbol{\Delta R_c^2}$	0.47 (1.67)	0.32 (2.18)	1.11 (1.62)	2.09 (1.88)
$\boldsymbol{\Delta ER_{ci}}$	0.04 (0.16)	0.92*** (0.04)	-0.02 (0.16)	0.72*** (0.27)
1 – Herfindahl	–	–	0.22*** (0.08)	0.33*** (0.09)
$\boldsymbol{\Delta BA+}$	–	–	-1.67e-06 (1.33e-06)	7.15e-07 (2.79e-06)
$\boldsymbol{\Delta Ave.yrs.schl.}$	-0.06 (0.02)	-0.06 (0.04)	-0.06*** (0.02)	0.00 (0.04)
Industry Fixed Effects $(\boldsymbol{d_i})$	Yes	Yes	Yes	Yes
Corrected for Sample Selection Bias	Yes	Yes	Yes	Yes
Instrumented for the new variable	–	Yes	–	Yes
Obs.	6213	6213	6213	6213
R^2	0.10	–	0.11	–

○: Associated with the results under OLS (3) in table (5.4), which were corrected for sample selection bias. **(.)**: Robust city-clustered standard deviation. *****, **, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. ***1 – Herfindahl***: A measure of diversification of employment in each city at the beginning of the decade studied. $\mathbf{\Delta BA+}$: The change in the number of people with a university degree in city. $\boldsymbol{\Delta Ave.yrs.schl.}$: Change in the average years of schooling in a city.

Table (5.9) – First Stage Results Associated with Specifications in Table (5.8)

1st Stage Associated with:	**IV (1)**		
Dependent Variable:	ΔR_c^1	ΔR_c^2	ΔER_{ci}
IV1	0.98*** (0.10)	-0.02 (0.02)	0.02 (0.04)
IV2	0.47*** (0.11)	0.78*** (0.02)	0.16*** (0.05)
$\boldsymbol{IV_{ER}}$	-0.01*** (0.00)	-0.00 (0.00)	-0.13*** (0.01)
1 – Herfindahl	-0.03*** (0.01)	-0.00 (0.00)	-0.00 (0.00)
Industry Fixed Effects $(\boldsymbol{d_i})$	Yes	Yes	Yes
R^2	0.53	0.82	0.56
Partial R^2 for Excl. Instr. ♣	0.39	0.81	0.30
Joint Redundancy Test for Excl. Instr. ♦	F(3,434) = 95.3 P-value = 0.00	F(3,434) = 557 P-value = 0.00	F(3,434) = 72.8 P-value = 0.00

(.): Robust, city-clustered standard deviation. *****, **, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. ***1 – Herfindahl***: A measure of diversification of employment in each city at the beginning of the decade studied.

Table (5.9) – Continued

1st Stage Associated with:	IV (2)			
Dependent Variable:	ΔR_c^1	ΔR_c^2	ΔER_{ci}	Δ**BA** $+$
IV1	1.06*** (0.10)	-0.01 (0.02)	0.02 (0.04)	-66468.2** (26348.6)
IV2	0.53*** (0.11)	0.79*** (0.02)	0.17*** (0.05)	-22752.6 (70349.6)
IV_{ER}	-0.01*** (0.00)	-0.00 (0.00)	-0.13*** (0.01)	1442.0** (624.8)
BA+ (1991)	-5.62e-08*** (1.74e-08)	2.58e-09 (3.96e-09)	-2.30e-08*** (6.43e-09)	0.25*** (0.04)
Industry Fixed Effects (d_i)	Yes	Yes	Yes	Yes
R^2	0.52	0.82	0.56	0.90
Partial R^2 for Excl. Instr. ♣	0.52	0.82	0.30	0.90
Joint Redundancy Test for Excl. Instr. ♦	F(4,434) = 108 P-value = 0.00	F(4,434) = 470 P-value = 0.00	F(4,434) = 77.1 Ps-value = 0.00	F(4,434) = 41.7 P-value = 0.00

(.): Robust, city-clustered standard deviation. *****, **, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. Δ**BA** $+$: The change in the number of people with a university degree in city.

Table (5.9) – Continued

1st Stage Associated with:	IV (3)			
Dependent Variable:	ΔR_c^1	ΔR_c^2	ΔER_{ci}	$\Delta \boldsymbol{Ave.yrs.schl.}$
IV1	0.93*** (0.09)	-0.04* (0.02)	0.03 (0.04)	-3.06 (2.34)
IV2	0.51*** (0.10)	0.78*** (0.02)	0.17*** (0.05)	-19.3*** (3.63)
IV_{ER}	-0.01*** (0.00)	-0.00 (0.00)	-0.13*** (0.01)	-0.00 (0.04)
Ave. Yrs. Schl. (1991)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00 (0.00)	-0.18*** (0.02)
Industry Fixed Effects (d_i)	Yes	Yes	Yes	Yes
R^2	0.54	0.83	0.56	0.32
Partial R^2 for Excl. Instr. ♣	0.53	0.82	0.30	0.31
Joint Redundancy Test for Excl. Instr. ♦	F(4,434) = 137 P-value = 0.00	F(4,434) = 468 P-value = 0.00	F(4,434) = 61.5 P-value = 0.00	F(4,434) = 45.8 P-value = 0.00

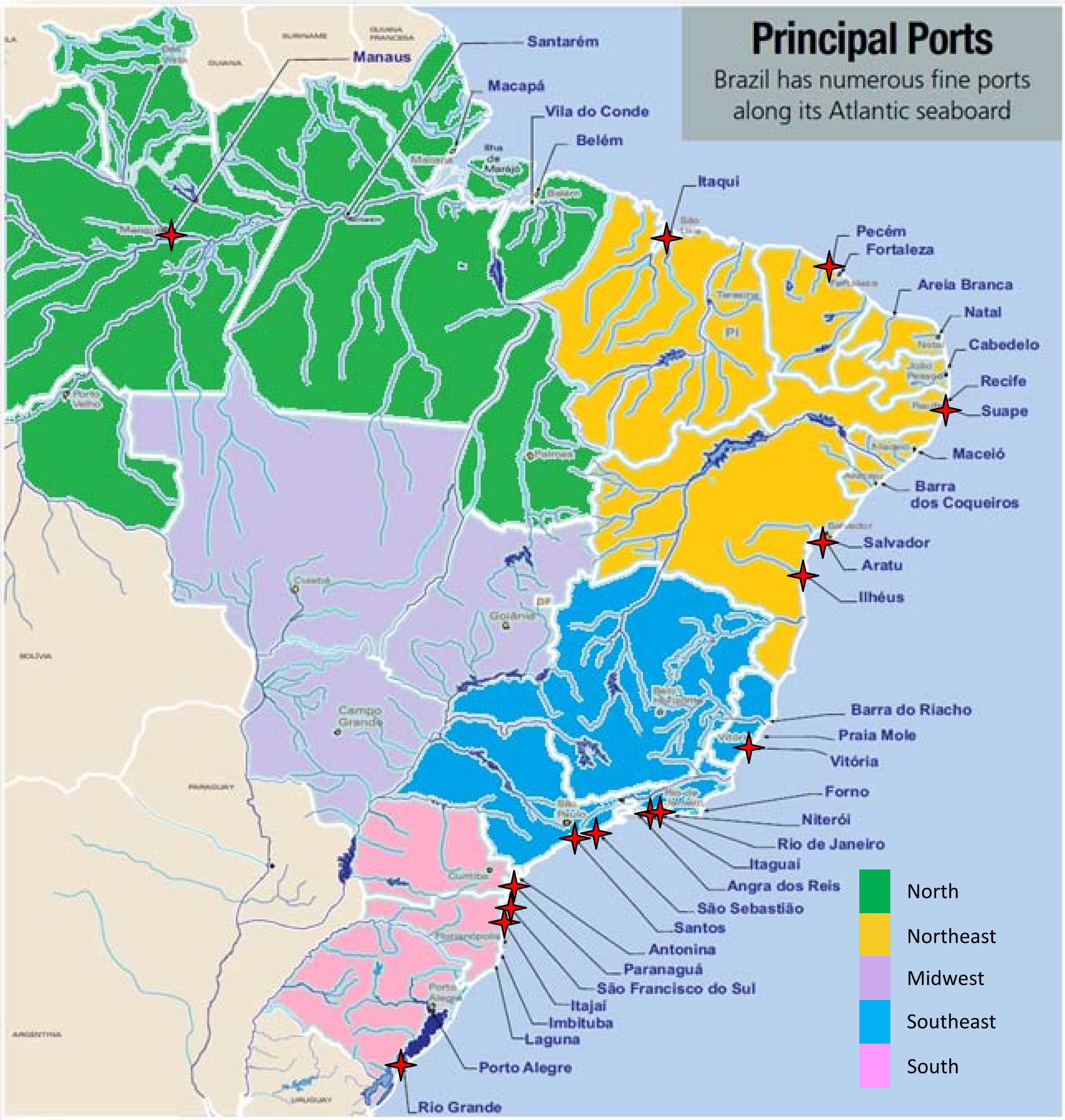
(.): Robust, city-clustered standard deviation. *****, **, ***: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. $\Delta \boldsymbol{Ave.yrs.schl.}$: Change in the average years of schooling in a city.

Table (5.9) – Continued

1st Stage Associated with:	IV (4)				
Dependent Variable:	ΔR_c^1	ΔR_c^2	ΔER_{ci}	Δ**BA** +	Δ***Ave. yrs. schl.***
IV1	0.88*** (0.08)	-0.04* (0.02)	0.01 (0.04)	-35620.3* (19763.2)	-2.79 (2.24)
IV2	0.48*** (0.10)	0.78*** (0.02)	0.16*** (0.05)	-25515.3 (67227.0)	-18.4*** (3.66)
IV_{ER}	-0.01*** (0.00)	0.00 (0.00)	-0.13*** (0.01)	1054.2** (497.2)	0.03 (0.04)
1 – Herfindahl	-0.02*** (0.01)	0.00 (0.00)	-0.00 (0.00)	-3353.1*** (1280.5)	0.49* (0.27)
BA+ (1991)	-4.14e-08*** (1.08e-08)	7.43e-09** (3.03e-09)	-2.49e-08*** (6.81e-09)	0.24*** (0.04)	1.87e-07 (1.01e-06)
Ave. Yrs. Schl. (1991)	-0.00* (0.00)	-0.00*** (0.00)	-0.00 (0.00)	610.4*** (198.7)	-0.20*** (0.02)
Industry Fixed Effects (d_i)	Yes	Yes	Yes	Yes	Yes
R^2	0.55	0.83	0.56	0.91	0.32
Partial R^2 for Excl. Instr. ♣	0.41	0.82	0.30	0.90	0.28
Joint Redundancy Test for Excl. Instr. ♦	F(5,434) = 88.0 P-value = 0.00	F(5,434) = 368 P-value = 0.00	F(5,434) = 51.7 P-value = 0.00	F(5,434) = 36.2 P-value = 0.00	F(5,434) = 30.0 P-value = 0.00

(.): Robust, city-clustered standard deviation. ***, **, *: Respectively, significance at 1%, 5%, and 10% levels of significance. ♣: Squared-partial correlation between excluded instruments and the endogenous variable in the associated first stage regression. ♦: The test is robust to clustering and heteroskedasticity. ***1 – Herfindahl***: A measure of diversification of employment in each city at the beginning of the decade studied. Δ**BA** +: The change in the number of people with a university degree in city. Δ***Ave. yrs. schl.***: Change in the average years of schooling in a city.

Figure (1.1) – Map of Major Commercial Seaports in Brazil (✦ indicates the top 15 major seaports used as reference in this paper[1])



Source: "An Ocean of Opportunities," Brazil Now, Ministry of External Relations, Brazil, November 2008.

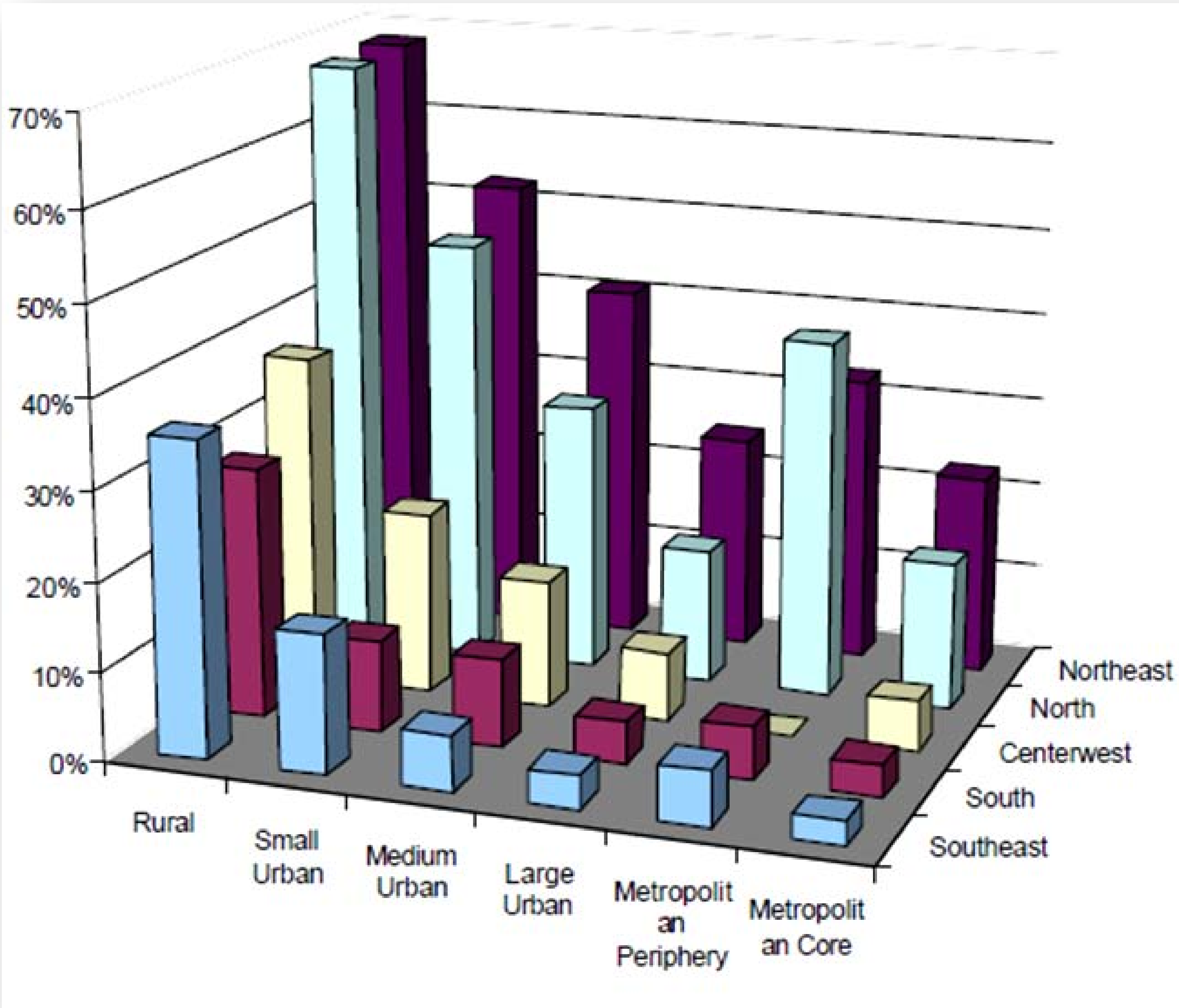
[1] These major ports are: Port Santos (Santos - SP), Port Vitória (Vitória - ES), Port Paranaguá (Paranaguá - PR), Port Itaguaí (Itaguaí - RJ), Port Rio Grande (Rio Gradne - RS), Port Rio de Janeiro (Rio de Janeiro - RJ), Port Itajaí (Itajaí - SC), Port Itaqui (São Luís - MA), Port São Sebastião (São Sebastião - SP), Port São Francisco do Sul (São Francisco do Sul - SC), Port Aratu (Candeias - BA), Port Manaus (Manaus - AM), Port Suape (Ipojuca - PE), Port Pecém (São Gonçalo do Amarante - CE), Port Ilhéus (Ilhéus - BA).

Figure (1.2) – Map of Regions in Brazil

Figure (1.3) – Brazil and the Rest of the World

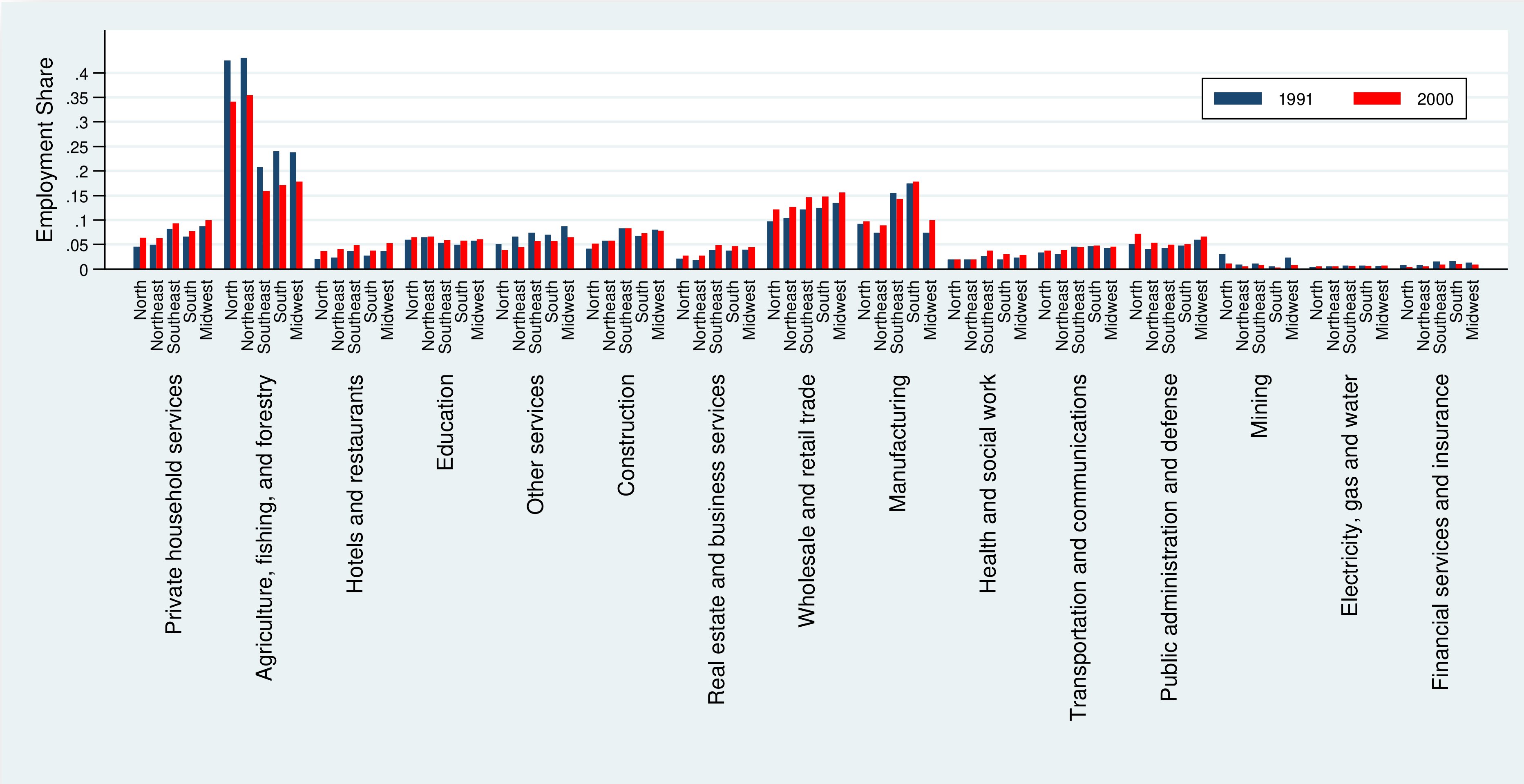
Source: Google Maps

Figure (1.4) – Extreme Poverty Rate by Region in Brazil

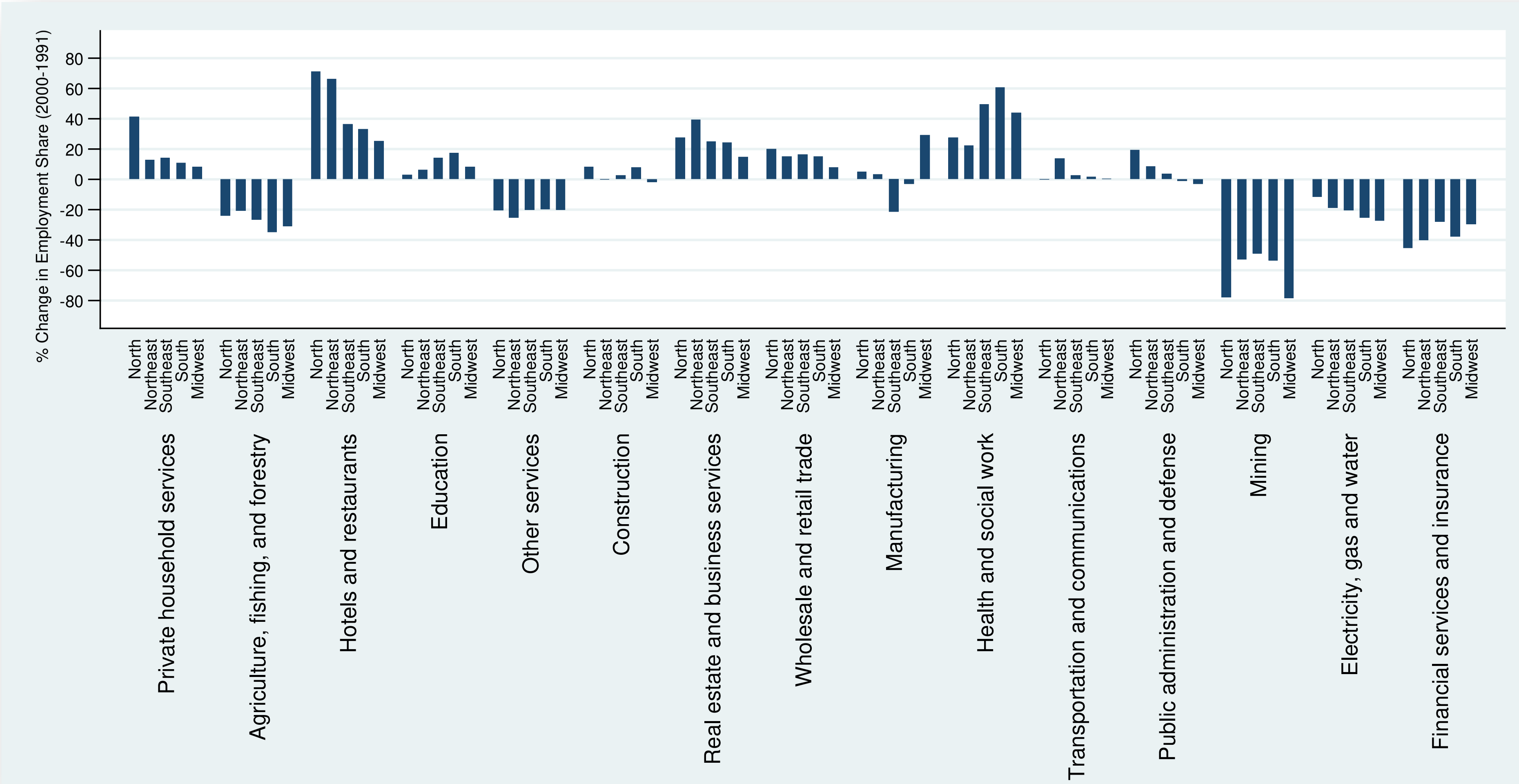


Source: "Attacking Brazil's Poverty: A Poverty Report with a Focus on Urban Poverty Reduction Policies," World Bank Report No. 20475-BR, Vol. I, March 31, 2001.

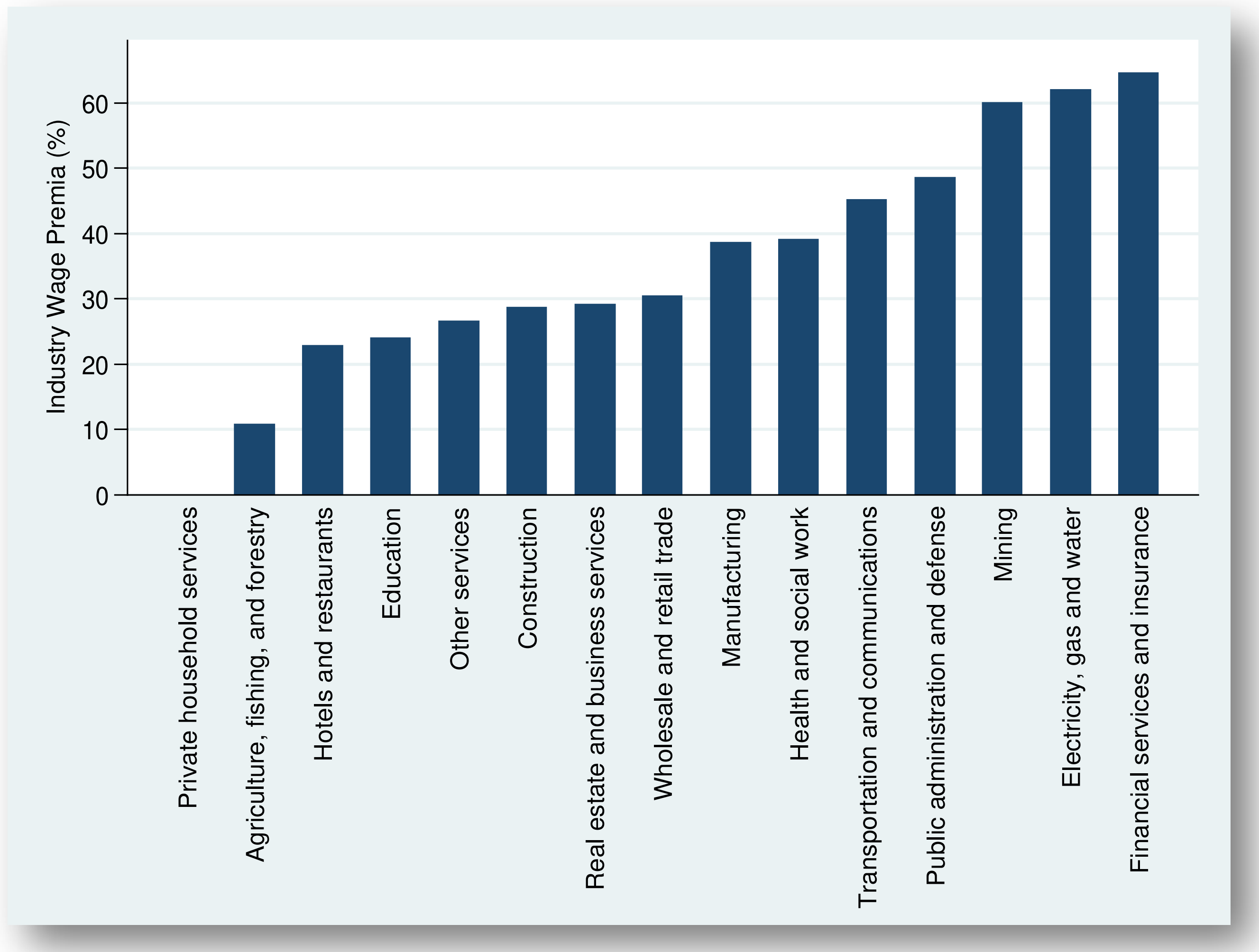
Graph (1.1) – Employment Share of Industries by Regions in Brazil (Industries sorted by average national wage premia increasing left to right)



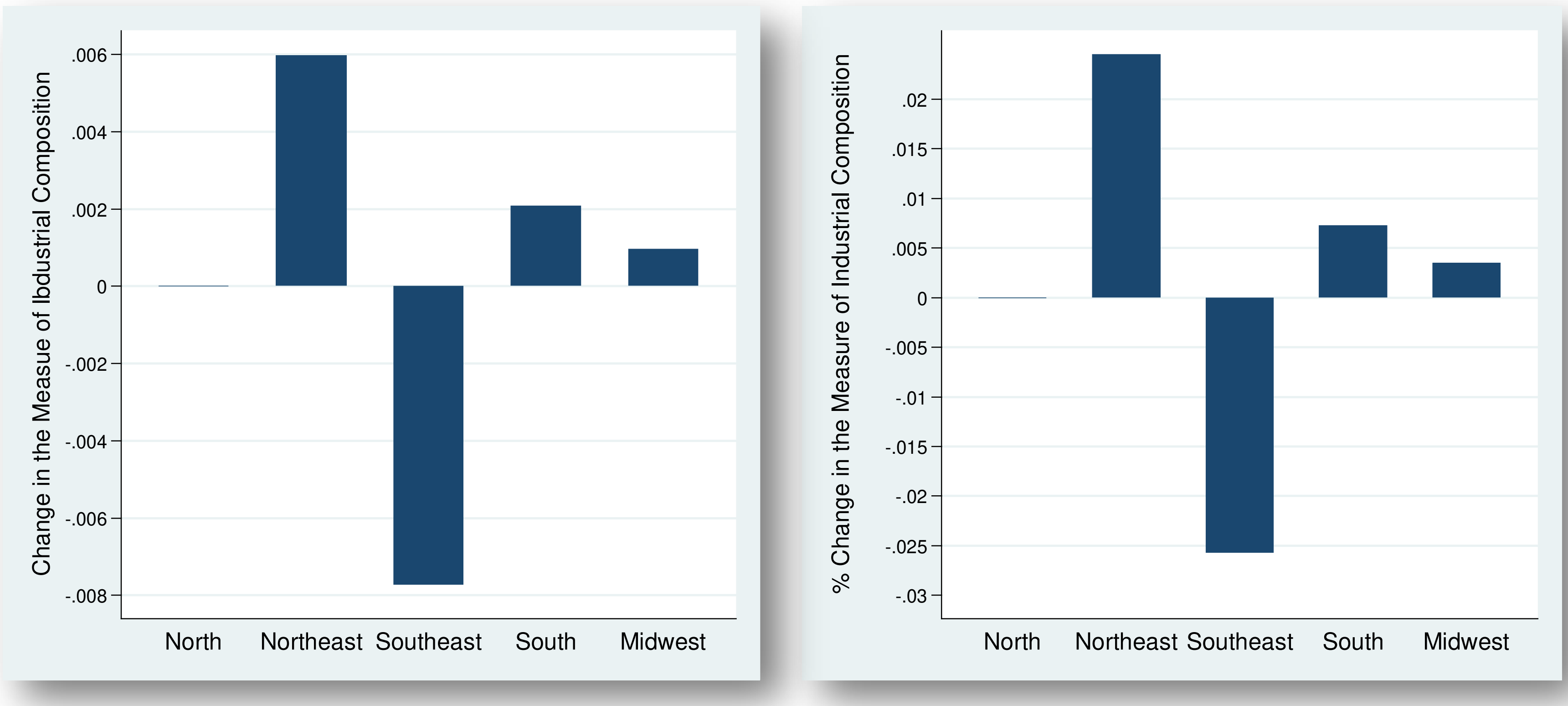
Graph (1.2) – Percentage Change in Employment Share of Industries by Region, 1991-2000 (Industries sorted by average national wage premia increasing left to right)



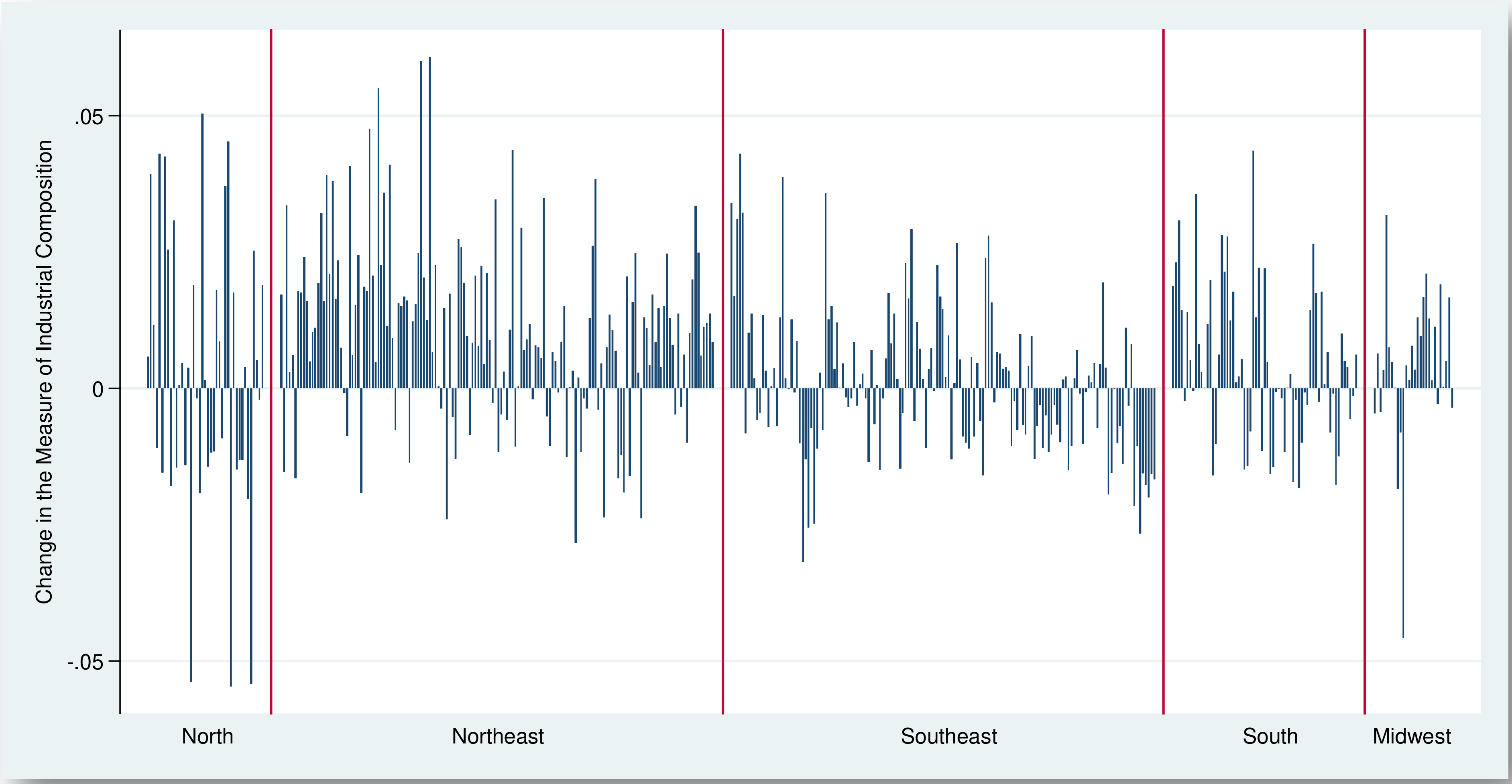
Graph (1.3) – Average Industry Wage Premia, 1991-2000 (Percentage difference from the base industry)



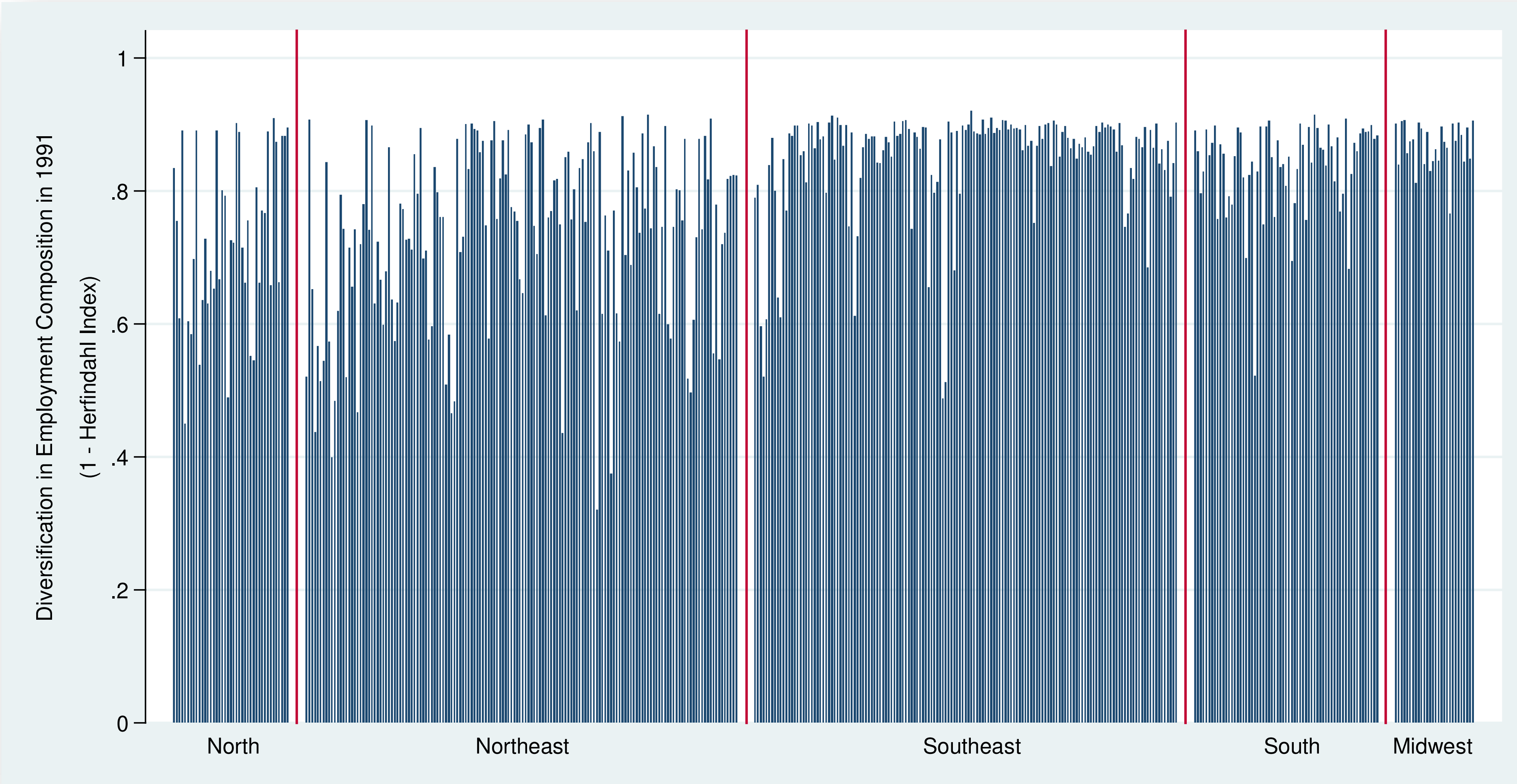
Graph (1.4) – Change in the Measure of Industrial Composition by Region



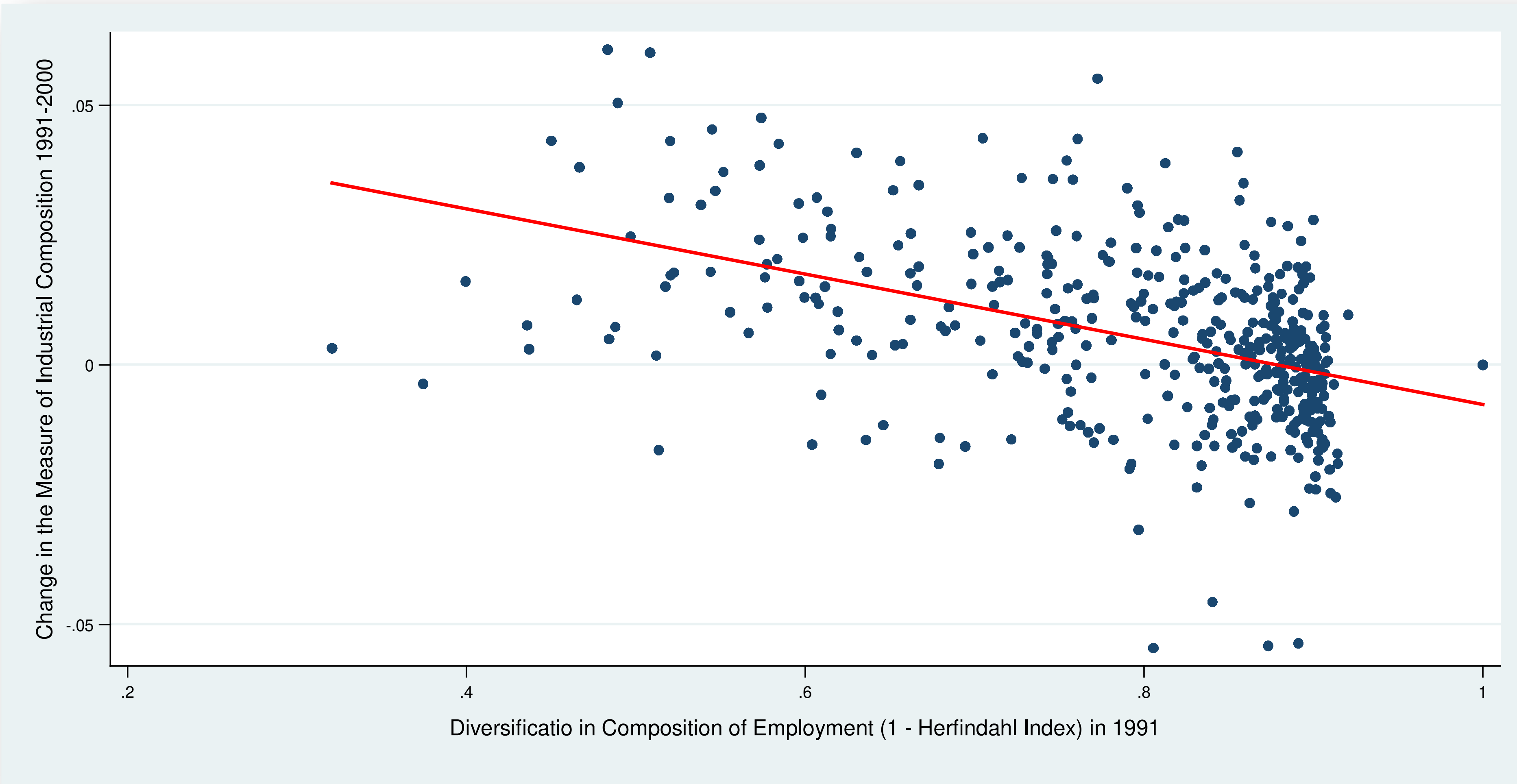
Graph (1.5) – Change in the Measure of Industrial Compositions by City and Region, 1991-2000



Graph (1.6) – Diversification in the Composition of Employment in 1991 by City and Region



Graph (1.7) – Diversification in the Composition of Employment in 1991 vs. Change in the Measure of Industrial Composition 1991-2000



Graph (5.1) - Controlled Variation in Changes in City-Sector Wages vs. Controlled Variation in Changes in the Measure of Local Industrial Compositions