

# **From Rivers to Roads: Spatial Mismatch and Inequality of Opportunity in Urban Labor Markets of a Megacity**

**EDUARDO A. HADDAD  
ANA MARIA BONOMI BARUFI**



## **From Rivers to Roads: Spatial Mismatch and Inequality of Opportunity in Urban Labor Markets of a Megacity**

Eduardo A. Haddad ([ehaddad@usp.br](mailto:ehaddad@usp.br))

Ana Maria Bonomi Barufi ([anambb@usp.br](mailto:anambb@usp.br))

**Research Group:** NEREUS – The University of São Paulo Regional and Urban Economics Lab

### **Abstract:**

The spatial mismatch between residential locations and jobs can be particularly relevant for low-skilled individuals. In this paper, we first explore such phenomenon from the perspective of the distance of the residence to the geographic concentration of jobs, and locational disadvantages of the urban form itself. Such unequal conditions in the labor market present a great challenge for the spatial sustainability of the multiple equilibria achieved simultaneously at the labor and the housing markets. We estimate the main aspects that influence wage differentials among individuals, focusing on the role of accessibility on labor market outcomes. We suggest the inclusion of geographical characteristics as instruments to deal with endogeneity problems that arise in the estimation of urban price models. Our strategy is based on the inclusion of a specific geographic/historic variable as instrument, namely the river shore access to the first school built by the Jesuits in São Paulo, the city's founding location. Geography acted as a determinant of the location of the transportation infrastructure in the region, so that the road and rail networks in the city present a strong spatial correlation with pre-urban "waterways". Nowadays, the vast majority of rivers and creeks are covered with asphalt and cement, and economic agents are practically unaware of their existence.

**Keywords:** Accessibility; inequality; urban labor market; São Paulo Metropolitan Region

**JEL Codes:** R14; R231; C26.

# From Rivers to Roads: Spatial Mismatch and Inequality of Opportunity in Urban Labor Markets of a Megacity

Eduardo A. Haddad<sup>1</sup> and Ana Maria B. Barufi<sup>2</sup>

**Abstract.** The spatial mismatch between residential locations and jobs can be particularly relevant for low-skilled individuals. In this paper, we first explore such phenomenon from the perspective of the distance of the residence to the geographic concentration of jobs, and locational disadvantages of the urban form itself. Such unequal conditions in the labor market present a great challenge for the spatial sustainability of the multiple equilibria achieved simultaneously at the labor and the housing markets. We estimate the main aspects that influence wage differentials among individuals, focusing on the role of accessibility on labor market outcomes. We suggest the inclusion of geographical characteristics as instruments to deal with endogeneity problems that arise in the estimation of urban price models. Our strategy is based on the inclusion of a specific geographic/historic variable as instrument, namely the river shore access to the first school built by the Jesuits in São Paulo, the city's founding location. Geography acted as a determinant of the location of the transportation infrastructure in the region, so that the road and rail networks in the city present a strong spatial correlation with pre-urban "waterways". Nowadays, the vast majority of rivers and creeks are covered with asphalt and cement, and economic agents are practically unaware of their existence.

**Keywords:** Accessibility; inequality; urban labor market; São Paulo Metropolitan Region

**Resumo.** O descompasso espacial entre locais residenciais e postos de trabalho pode ser particularmente relevante para as pessoas menos qualificadas. Neste artigo, exploramos esse fenômeno a partir da perspectiva da distância da residência à concentração geográfica dos postos de trabalho e das desvantagens locacionais inerentes à própria forma urbana. Tais condições desiguais no mercado de trabalho apresentam um grande desafio para a sustentabilidade espacial dos múltiplos equilíbrios alcançados simultaneamente nos mercados de trabalho e de habitação. Estimamos os principais aspectos que influenciam as diferenças salariais entre indivíduos, enfocando o papel da acessibilidade nos resultados do mercado de trabalho. Sugerimos a inclusão de características geográficas como instrumentos para lidar com problemas de endogeneidade que surgem na estimativa de modelos de preços urbanos. Nossa estratégia é baseada na inclusão de uma variável geográfica / histórica como instrumento, ou seja, o acesso à primeira escola construída pelos jesuítas em São Paulo, local de fundação da cidade caminhando pelas margens dos rios. A geografia agiu como um determinante da localização da infraestrutura de transportes na região, de modo que as redes rodoviárias e ferroviárias da cidade apresentam uma forte correlação espacial com a rede hídrica. Hoje em dia, a grande maioria dos rios e córregos é coberta por asfalto e cimento, e os agentes econômicos praticamente desconhecem a sua existência.

**Palavras-chave:** Acessibilidade; desigualdade; mercado de trabalho urbano; Região Metropolitana de São Paulo

**Área ANPEC:** Área 10 - Economia Regional e Urbana

**Classificação JEL:** R14; R23; C26

---

<sup>1</sup> Full Professor at the Department of Economics at the University of Sao Paulo, Brazil, and NEREUS – The University of Sao Paulo Regional and Urban Economics Lab

<sup>2</sup> NEREUS – The University of Sao Paulo Regional and Urban Economics Lab

## **1. Introduction**

Location decisions of firms and workers shape the spatial distribution of economic activity between and within cities. On one hand, the interaction between cities is widely investigated in the literature of regional and urban economics, which tries to assess the extent to which urban scale affects the local concentration of different skills, sectors, etc., apart from defining each city's role in the regional system. On the other hand, within-city dynamics and internal heterogeneity is the focus of urban labor economic theory, simultaneously analyzing land/housing markets with the labor market (Zenou, 2009).

Furthermore, the spatial scale of analysis shapes the choice of tools and measures to investigate specific issues. For instance, agglomeration economies are usually assessed by aggregated measures such as total population or density. However, this type of analysis does not provide a complete understanding of the intra-urban distribution of agglomeration economies effects (Melo and Graham, 2009). An alternative approach requires the estimation of market potential measures over labor market indicators, with distance decaying effects varying over space. In a way, accessibility to jobs embeds the concept of market potential and a higher chance of interactions and matching in the job market. Thus, agglomeration economies may have heterogeneous effects within cities, affecting location decisions of workers and firms.

In this paper, we look at how economic agents benefit from such interactions within a specific urban labor market area. Any equilibrium in the labor market will be directly related to the housing market. Therefore, wages and housing prices will alternatively be the interest variables, being jointly affected by, among other characteristics, the accessibility to jobs in the city.

We use the São Paulo Metropolitan Region (SPMR), the main economic and financial center of Brazil, as our case study. SPMR is the fourth largest urban agglomeration of the world, and the largest urban agglomeration in the country, with about 10% of the national population (around 20 million inhabitants), and responsible for 19% of Brazilian GDP. The city of São Paulo is the core of the metropolitan area and accounts for 5.9% of the country's population and 12% of its GDP. As we will see, the spatial mismatch between residential locations and jobs can be particularly relevant for poorer individuals, who rely more heavily on public transportation.

In what follows, we explore the relationship between commuting time, accessibility and urban prices in the context of urban markets in a megacity in the developing world. We start by lining up the theoretical framework that permeates the empirical analysis. The key point is the simultaneous determination of land use and urban prices in the context of the family of spatial general equilibrium urban models we consider. That brings challenges for dealing with simultaneity bias in the empirical estimation of urban wages and housing prices models. We address this issue by relying on an instrumental variables estimation that uses the river shore access to the city's founding location as instrument for accessibility to jobs.

## **2. Theoretical Background**

We use as the conceptual framework a spatial general equilibrium model of land use with endogenous job locations in a circular city. The model we draw on was developed in Lucas and Rossi-Hansberg (2002) – henceforth, LRH – and represents a generalization of the work by Fujita and Ogawa (1982).<sup>3</sup> The model was further extended in Rossi-Hansberg (2004) to look at efficiency properties of land allocation in the context of the same conceptual framework.<sup>4</sup> The theory determines the distribution of business and residential land together with employment and residential densities at all locations in the city, as well as urban prices (i.e. wages and land rents). Given the general equilibrium nature of the theory, one should be aware of the potential simultaneity bias in the empirical estimation of wage and housing price equations based on accessibility measures.

In this framework, productivity is higher the higher is employment in neighboring locations. Following Lucas (2001), the specification of the model considers such external effect of employment to decline exponentially with distance. This arbitrary form of production externalities will prove to be very useful for our empirical work.

Despite the admittedly very stylized nature of the model, it can be used as a conceptual framework to understand the interaction of forces that are important in determining urban structure. As such, it provides the structural background to obtain quantitative implications in more detailed empirical studies.

The take-home results from LRH that are relevant for our empirical strategy are:

1. The model simultaneously determines the location of jobs and workers, as well as wages and land rents.
2. In equilibrium, wages within the city increase with employment density.
3. In equilibrium, land rents within the city also increase with density of workers.
4. As a consequence of (1)-(3), accessibility to jobs is also simultaneously determined, for any given commuting cost function.

We will depart from reduced-form intra-urban prices equations based on accessibility to capture production externalities in the form proposed in LHR. Many studies have already considered accessibility as a key variable in models for housing prices and land rents. Duranton and Puga (2015) have surveyed a well-established urban economics literature that provides evidence that accessibility determines land and housing prices in different locations. However, a far less numerous literature provides evidence of accessibility determining intra-urban wages.

In the case of land and housing prices models, Duranton and Puga (2015, p. 471) points out that the patterns of accessibility are also affected by the location choices of firms and workers, which are determined by prices. Hence, the land use problem is in essence a hard equilibrium problem with many feedbacks. The authors also recognize

---

<sup>3</sup> Zenou (2009) provides a modern synthesis of urban labor economic theory, which consolidates research that explicitly model both the land/housing market (where both the location of workers and the price of land/housing market are endogenous) and the labor market (where both wages and unemployment are endogenous).

<sup>4</sup> The model in Rossi-Hansberg (2004) characterizes the optimal distribution of urban land. It is shown that the analysis of optimal urban structure can be used to design optimal policies to improve the efficiency of equilibrium allocations in the same setup.

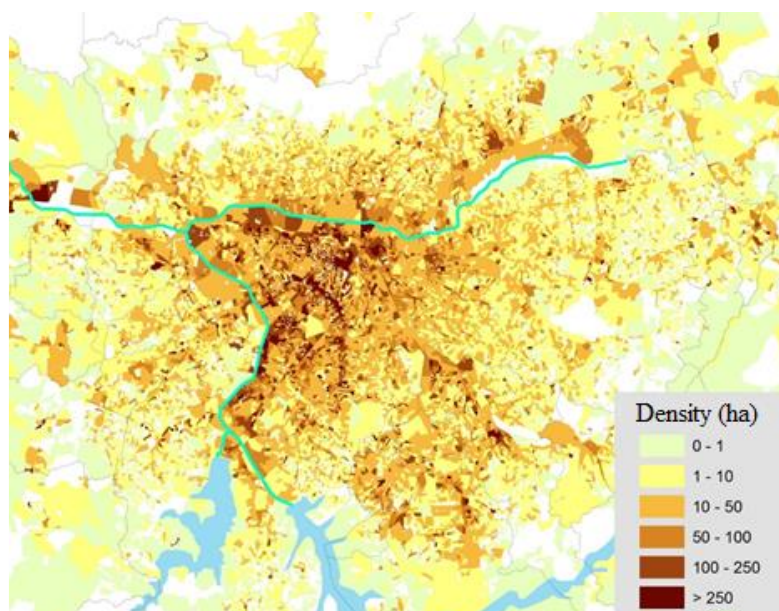
that the literature first solved it by restricting accessibility to be solely about access to jobs and by treating the location of these jobs as exogenous within a simple geography and with frictionless markets. In cities where the majority of jobs are well outside the urban core, it is also empirically problematic to equate accessibility to jobs (or even general accessibility) to the distance to the city center. In other words, better work is needed to reduce uncertainty.

Focusing on the labor market outcomes, we present the research methodology discussing the strategy to get rid of the usual identification hypothesis that restricts job locations to the CBD. Our strategy allows dealing with the simultaneity nature of business and housing locations and urban prices determination in a richer geographical setting. Before proceeding to that, we look at some stylized facts concerning the spatial structure of our study region.

### 3. The Internal Structure of the SPMR – Some Stylized Facts

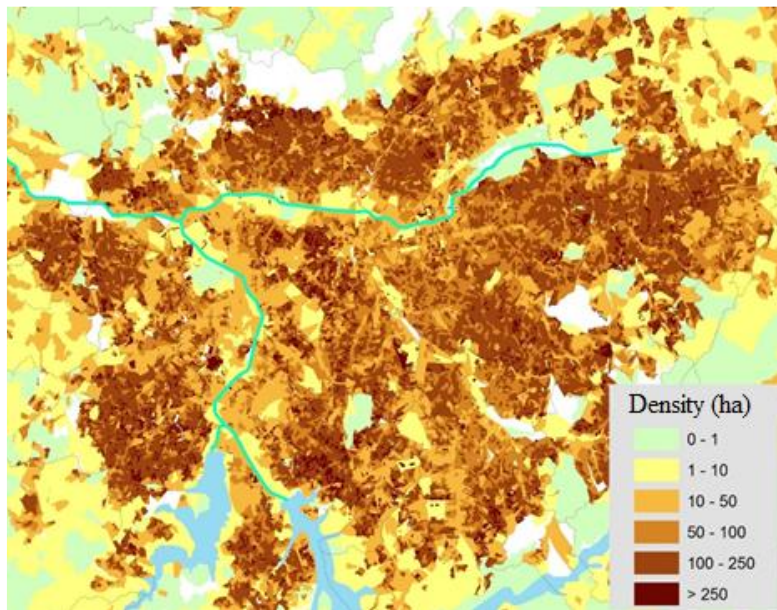
From a stylized perspective, a Muth-Mills-Alonso urban model, having as the CBD the extended center of the city of Sao Paulo (Haddad et al., 2015), may loosely approach the internal organization of the SPMR. Even though the broadly-defined CBD concentrates a great part of the jobs, a considerable level of employment decentralization is perceived in the region (Figure 1). Households are spread across the territory, mainly located in the surroundings of the center, with population density decay in the boundaries of the territory of the metropolis (Figure 2).

**Figure 1. Employment Density in SPMR, 2008**



Source: RAIS data, Ministry of Labor, IBGE (own elaboration)

**Figure 2. Population Density in SPMR, 2010**



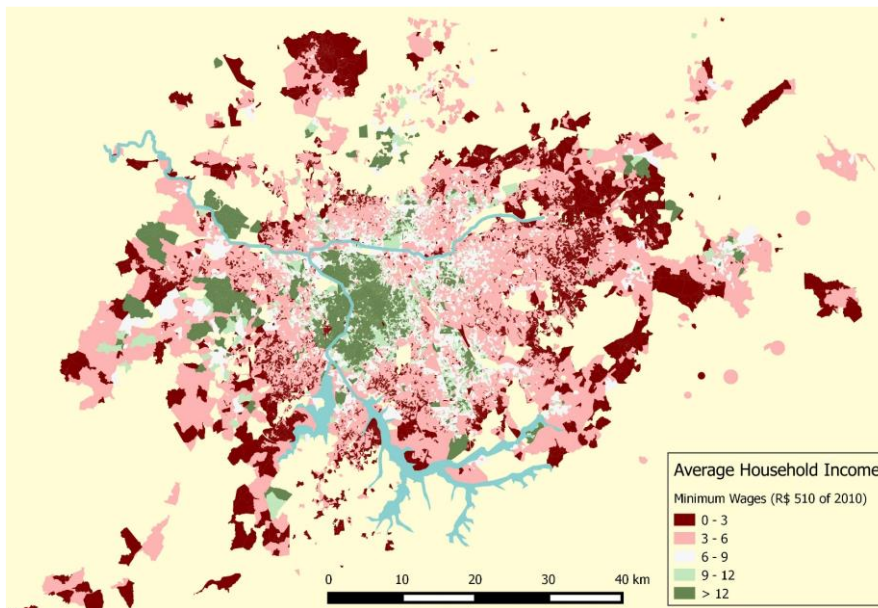
Source: Demographic Census data, IBGE, 2010 (own elaboration)

According to the 2010 population census, the city of São Paulo – the core of the SPMR – received daily an inflow of almost one million commuters, representing 15.4% of workers in the city. The great majority of commuters' flows were from peripheral regions in São Paulo and other municipalities to the metropolitan business centers in the central and western zones of São Paulo city.

Low-income residents are overwhelmingly overrepresented in the periphery (Figure 3). Considering that commuting patterns also vary across the SPMR, where the share of job commuting trips taken by public transportation tends to be higher in the peripheral areas (Figure 4), the average commuting time for the low-income population tends also to be higher, diminishing considerably their access to job opportunities.

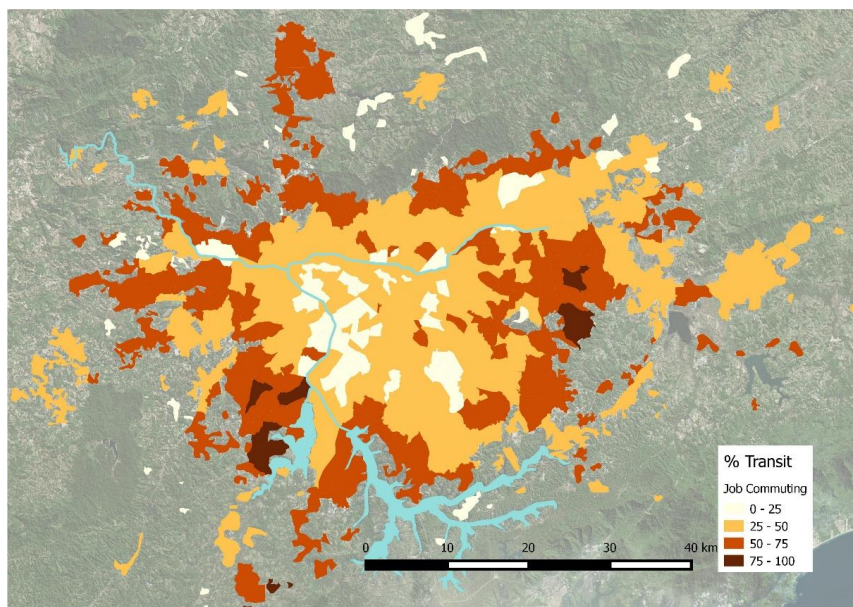
One important implication of the pattern of spatial distribution of workers and residents observed in SPMR, together with the travel demand behavior, is that it will make it harder for some workers to access jobs. In particular, the somehow diffuse job decentralization makes it more difficult for workers with no car to reach potential employers once these are located far from the focal points of accessibility associated with the public transportation infrastructure. This spatial mismatch makes it much harder for low-income workers to find employment since they traditionally reside in the less central parts of cities and depend heavily on the radial infrastructure of public transportation (Figure 5).

**Figure 3. Average Household Income in SPMR, 2010**



Source: Demographic Census, IBGE, 2010 (own elaboration)

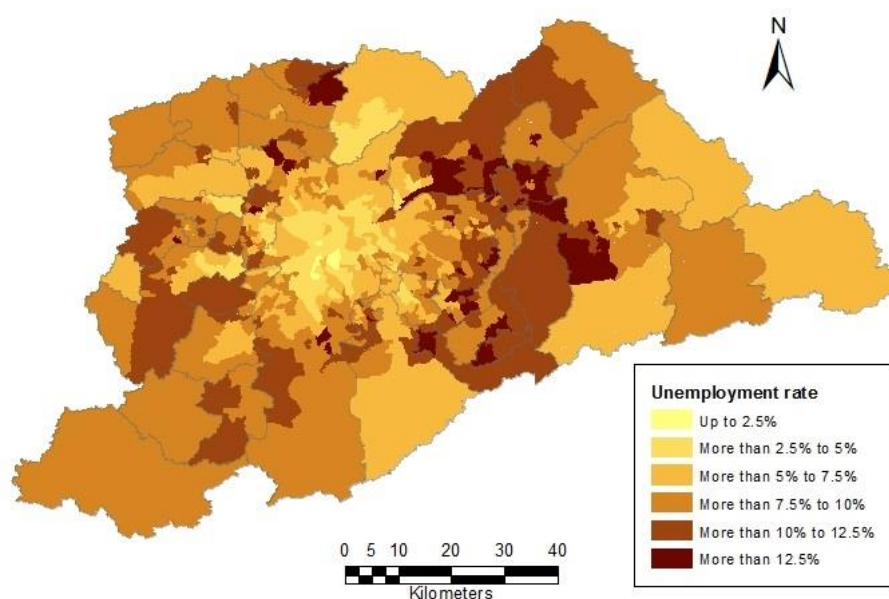
**Figure 4. Share of Commuting by Public Transportation**



Source: O-D Survey, 2007 (own elaboration)



**Figure 5. Unemployment Rate in SPMR, 2010**



Source: Demographic Census, IBGE, 2010 (own elaboration)

#### **4. Research Design**

There is a substantial empirical literature concerned with variation of urban prices within cities. The grand majority relies on the identification hypothesis that employment is concentrated in the central city. We revisit the usual empirical strategies to estimate the effects of accessibility to jobs on urban prices, attempting to address properly the problem of simultaneity bias that emerges from adopting the conceptual framework described in Section 2. Our goal is to bring empirical work closer to theory. Adopting a rich geography setting, we want to learn more about fundamentals, rather than looking at the relationship between urban prices and accessibility for its own sake.

The empirical strategy focuses on the labor market.<sup>5</sup> The model is based on the econometric estimation of a wage equation motivated by urban labor economic theory (Zenou, 2009). To estimate the model, we use micro data for a large sample of households in the SPMR and their individual household members with information for their travel behavior. The regression analysis of the wage equation considers as the dependent variable the earnings of each worker in the sample. Commuting time and an index of accessibility to jobs – discussed below – are the relevant independent variables. Long commute is expected to decrease workers' productivity as longer commuting time may induce workers to arrive late at work, or leave earlier, and

---

<sup>5</sup> A complementary empirical strategy would focus on the housing markets. The estimation of a hedonic price model (with the effects of agglomeration economies on housing prices) would be the more pragmatic approach. One would ideally study the housing market based on repeated sales prices from all transactions. However, such information is not available for Brazilian markets. As a suggestion for future studies, one way to deal with this issue would be to use the asking price of new residential developments registered by EMBRAESP. This database covers the SPMR since 1985, with geocoded micro data at the housing unit level. It is not a statistical sample, but it covers all new developments publicly announced in the formal housing market.

increase the number of absent days (Van Ommeren and Gutiérrez-I-Puigarnau, 2009); moreover workers experiencing longer commuting trips may also become less productive as they provide lower effort levels than those residing closer to jobs (Zenou, 2002). Agglomeration economies, captured by the accessibility index that resembles a market potential measure, are expected to positively influence workers' earnings. Workers are paid more in larger and denser markets because they are more productive there due to the presence of agglomeration economies (Melo and Graham, 2009).<sup>6</sup>

The main problem with the aforementioned approach is a strong theoretical concern that, since wages are determined simultaneously with accessibility to jobs, the latter will be potentially correlated with the error term, leading to bias and inconsistency in OLS estimation. In the following discussion, we provide evidence about the accessibility effects on urban prices, attempting at tackling the fundamental simultaneity problems that afflicted previous empirical work relying on the strong identification hypothesis of central city jobs. The solution for this problem requires finding a suitable instrument for our accessibility measure.

Inspired by the work by Duranton and Turner (2011), our instrument should reflect the suitability of SPMR geography to build the urban transportation infrastructure, keeping in mind that it should be uncorrelated with the residuals in our regressions.

Our instrumental variables strategy is based on the inclusion of a geographic/historic variable as the instrument for accessibility to jobs, namely the river shore distance to the first school built by the Jesuits in São Paulo, the city's founding location currently known as *Pateo do Collegio*.<sup>7</sup> The city has since expanded around that location, as can be attested by historical maps (Figure 6). Today, São Paulo's official "ground zero" lies a few blocks from there. We construct our instrument from a GIS database of the watershed within SPMR (Figure 7). The main task was to calculate the distance to the historical site from every observation unit in the database.<sup>8</sup>

---

<sup>6</sup> Agglomeration economies captured by the model accessibility index are of the "matching" type (improved matching of labor and firms, as well as higher probability of matching).

<sup>7</sup> In 1553, P. Manuel da Nobrega, after a visit to the Inhambussu plateau ("the one you see in the distance"), decided to start there a new mission for the Society of Jesus. One year later, with the intention to teach and catechize the Indians who lived in the plateau of Piratininga, the Jesuits built the first school made of wattle and daub (construction technique with mud, bamboo and straw). The location is currently known as Pateo do Collegio and it is considered the founding site of the city of São Paulo.

<sup>8</sup> To create an "Origin-Destination Cost Matrix" we used the *Network Analyst* extension from *ArcGIS*. The parameter used to build the OD Cost Matrix was route length (in kilometers). We have assumed that people could walk in both directions along the bank of the rivers.

**Figure 6. The “First” Map of São Paulo City – 1807-1810**

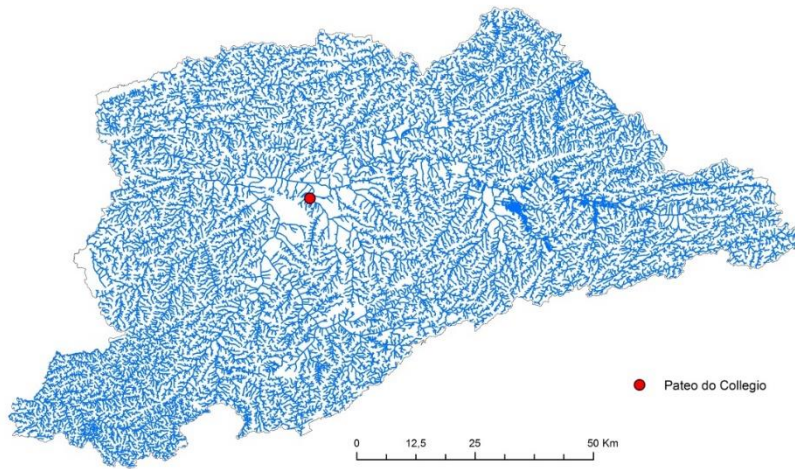


Source: São Paulo Antigo: Plantas da Cidade (1954)

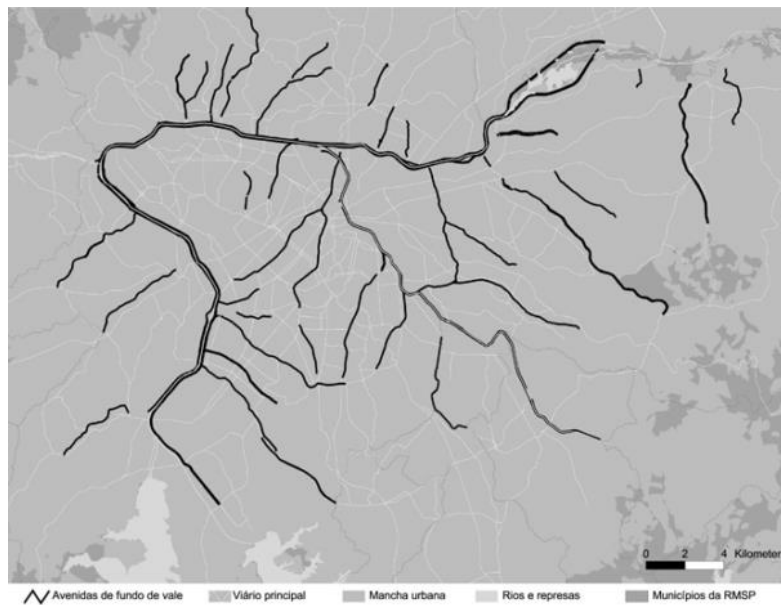
The justification for the choice of such instrument follows. Geography has acted as a determinant of the location of the transportation infrastructure in the region, so that the road network in the city presents a strong spatial correlation with pre-urban “waterways” (natural design for roads). Nowadays, the vast majority of rivers and creeks are covered with asphalt and cement, and economic agents are practically unaware of their existence. The *a priori* case for the exogeneity of our instrument seems plausible.

In a historical perspective, in the mid-nineteenth century, the city of São Paulo started a more systematic occupation of the floodplains of the main watercourses of the city. The reason was associated with the implementation of regional and urban infrastructure (railways and roads). In 1929, the *Plano de Avenidas* elaborated by Prestes Maia further reinforced the occupancy of the floodplains of rivers, by focusing in the development of large avenues in the thalwegs. This conception of the use of floodplains as the preferential space for circulation prevailed in the following development plans for the city, so that thalwegs avenues represent nowadays the main arteries of São Paulo (Travassos and Grostein, 2003; Meyer et al., 2004) (Figure 8).

**Figure 7. Watershed in SPMR**



**Figure 8. Urban Rivers Corridors in the City of São Paulo**



Source: Meyer, Grostein and Biderman (2004)

#### **4.1. Measuring Accessibility to Jobs<sup>9</sup>**

The previous discussion highlighted the importance of instrumenting for accessibility to jobs in the empirical analysis. The accessibility index to be used in this study is based on Hansen's formulation (Hansen, 1959), which defines accessibility as the potential of opportunities for interaction. It weighs the travel time of each location to

---

<sup>9</sup> This section draws on Vieira and Haddad (2015)

the various job locations in SPMR, in a rich geographical setting. This formulation is convenient to introduce the form of production externalities specified in LRH.

According to Du and Mulley (2006), Hansen gravity-based accessibility measures can be considered the most robust approach to measure accessibility to a certain service, such as employment, since the main advantage of this indicator is the ability to combine the effects of transport and land use. The limitations are the following: the calibration of the impedance function and the segregation of the effects of separation and attractiveness. Nevertheless, indicators derived from Hansen's model have been widely used for this type of measurement in the literature (Raia, 2000). We show its basic formulation below:

$$A_i = \sum_{j=1}^n \frac{w_j}{d_{ij}(t)} \quad (1)$$

According to this formula, the accessibility  $A$  for each region  $i$  is given by the sum of opportunities  $w$  available at each region  $j$ , divided by the impedance  $d_{ij}(t)$  to go from  $i$  to  $j$ .

We define the opportunities  $w_j$  as the total employment  $E$  in each region  $j$ .

$$w_j = E_j \quad (2)$$

The values of  $E_j$  can be extracted from the Origin-Destination (OD) Survey of 2007, conducted by the São Paulo Metropolitan Company, in which 30,000 households were interviewed to collect socioeconomic and commuting data on all residents. Thus, regions  $i$  and  $j$  are defined according to the 460 zones of the OD survey.

We define the impedance function  $d_{ij}(t)$  as an exponential function of travel time between each pair of zones.

$$d_{ij}(t) = e^{\alpha t} \quad (3)$$

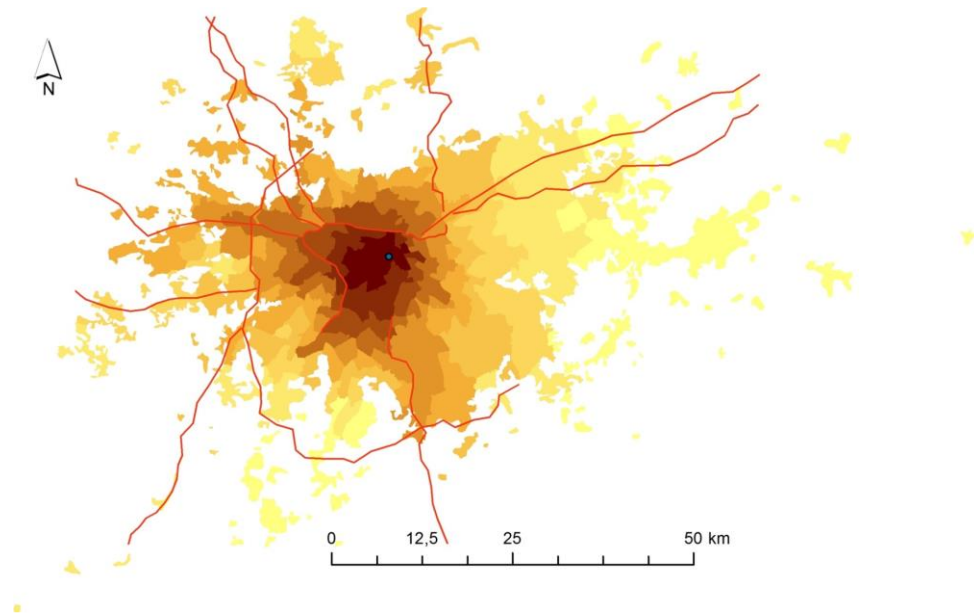
where  $\alpha$  is a parameter to be calibrated.

Figures 9 and 10 show the spatial distribution of two accessibility indices calculated for private vehicle mode and public transportation. Both maps are colored according to a division of the 460 zones into 10 groups with 46 zones in each group; the darker the colors the higher the accessibility of the group.

While accessibility by private vehicle has a spatial distribution in the shape of concentric rings, with higher accessibility the closer the zone is to the center, accessibility by public transportation follows a different pattern. In this case, we can observe a relationship between the spatial distribution of the index and the urban railway network of the city in an approximate radial shape.

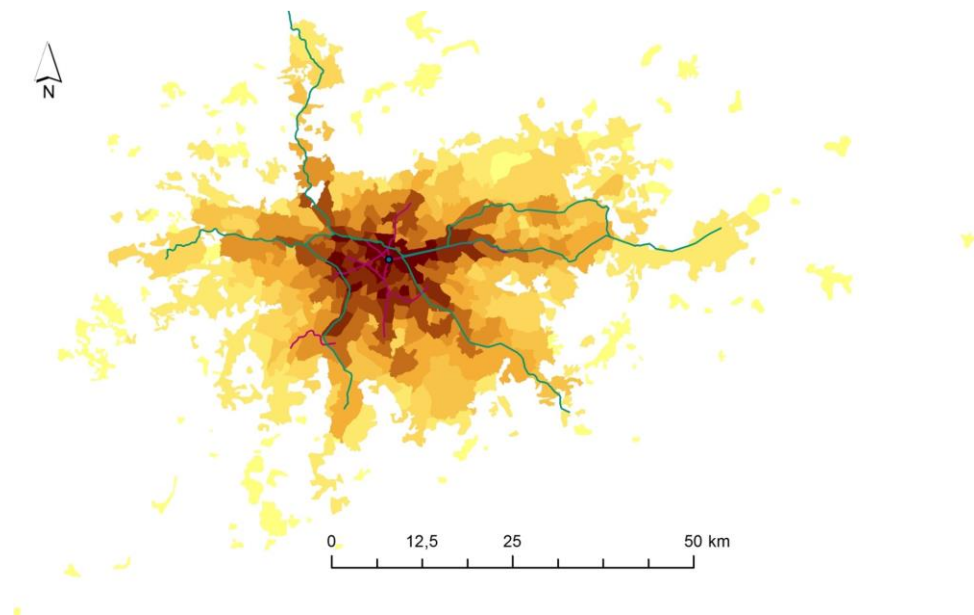
Vieira and Haddad (2015) also show an interesting pattern of spatial mismatch relating accessibility and income. They find strong spatial correlation between accessibility and income, i.e., there are significant clusters of high income and high accessibility in the central part of São Paulo, and low income and low accessibility clusters in the periphery.

**Figure 9. Spatial Distribution of Accessibility by Private Vehicle Mode**



Source: Vieira and Haddad (2015)

**Figure 10. Spatial Distribution of Accessibility by Public Transportation Mode**



Source: Vieira and Haddad (2015)

## 4.2. Econometric Approach

We estimate the empirical models related to wages determination. We consider a mincerian wage equation in the following way:

$$\ln w_i = \alpha_0 + X_i\beta + R_{r(i)}\gamma_1 + WP_{wp(i)}\gamma_2 + \theta A_{r(i)} + \varepsilon_i$$

(4)

Where the individual wage ( $w_i$ ) is explained by a set of individual characteristics ( $X_i$ ), attributes of the region of residence ( $R_{r(i)}$ ) and the region of work ( $WP_{wp(i)}$ ), the accessibility index ( $A_{r(i)}$ ), and the error term  $\varepsilon_i$ . There are potential endogeneity issues related to simultaneity of  $w_i$  and  $A_{r(i)}$ , as discussed above.<sup>10</sup>

In addition to a vector of individual characteristics (years of education, age and gender), other controls in the model include place of work and sectoral dummies. While the place of work dummies attempt to capture the urban structure and the peaks and valleys of the wage surface in a multi-centric urban configuration (Fujita and Ogawa, 1982; Lucas and Rossi-Hansberg, 2002), the sectoral dummies take into account the industry-mix and differences in wages (and productivity) across sectors within the metropolitan area.

## 5. Data and Descriptive Analysis

The theoretical approach outlined in Section 2 is based on the simultaneous analysis of the labor and the housing markets within an urban area. The analysis that follows covers the SPMR using as the main database the micro data of the aforementioned OD survey.

### 5.1. OD Survey and the Wage Equation

Origin-destination (OD) surveys aim at gathering information on trips done by the population of a given metropolitan area in a typical business day. They provide the basis for urban and transportation planning, feeding simulation models to forecast trips and evaluate future projects of the transportation network.

On a decennial basis, the São Paulo Metropolitan Company carries out the OD survey in São Paulo. The most recent available data refers to 2007, with 169,625 observations obtained from around 30,000 households about trips made in that year in the SPMR. The metropolitan region comprises 39 municipalities (including São Paulo city). These municipalities are subdivided in 460 survey zones, the basic spatial units of analysis (the municipality of São Paulo itself is composed by 320 zones). The survey uses a stratified sampling technique, with error margins below 5% and confidence intervals of 95%.

---

<sup>10</sup> Other sources of endogeneity may arise from workers' behavior. On the one hand, workers are more prone to commute long distances in exchange for higher wages. On the other hand, individuals with higher incomes may locate further away from their jobs to have access to amenities usually displaced in the outskirts of the city (Melo and Graham, 2009).

The OD survey provides detailed information at the individual level (income, work characteristics, house and durable goods ownership, household size, among others). In addition, it has trip information on time, latitude and longitude for departure and arrival, transportation mode, reason of the trip, etc. This survey is the main data source to investigate the relationship between wages and commuting time.

The analysis considers the information on employed individuals with positive wages and their trips from home to work. An important distinction is made among people commuting by public transportation and with private vehicles. In this context, the main descriptive statistics that are relevant for this analysis are presented in Table 1.

**Table 1. Descriptive Statistics of the Main Variables for the OD Survey, 2007**

	Private vehicle		Public transportation		Total	
	mean	s.d.	mean	s.d.	mean	s.d.
Wage (individual income)	R\$ 2,036	R\$ 2,035	R\$ 982	R\$ 909	R\$ 1,316	R\$ 1,455
Commuting time (min.)	39	29	76	40	64	41
Distance to Ground Zero (km)	34.6	21.5	39.8	21.1	38.1	21.4
Distance to Pateo do Collegio by rivers (km)	32.8	21.8	38.1	21.4	36.4	21.7
Accessibility to jobs through transportation	4,334,939	1,139,496	3,551,412	862,184		
	<b>Individuals</b>	<b>%</b>	<b>Individuals</b>	<b>%</b>	<b>Individuals</b>	<b>%</b>
Weighted sample	1,041,992	31.7%	2,247,343	68.3%	3,289,335	100.0%

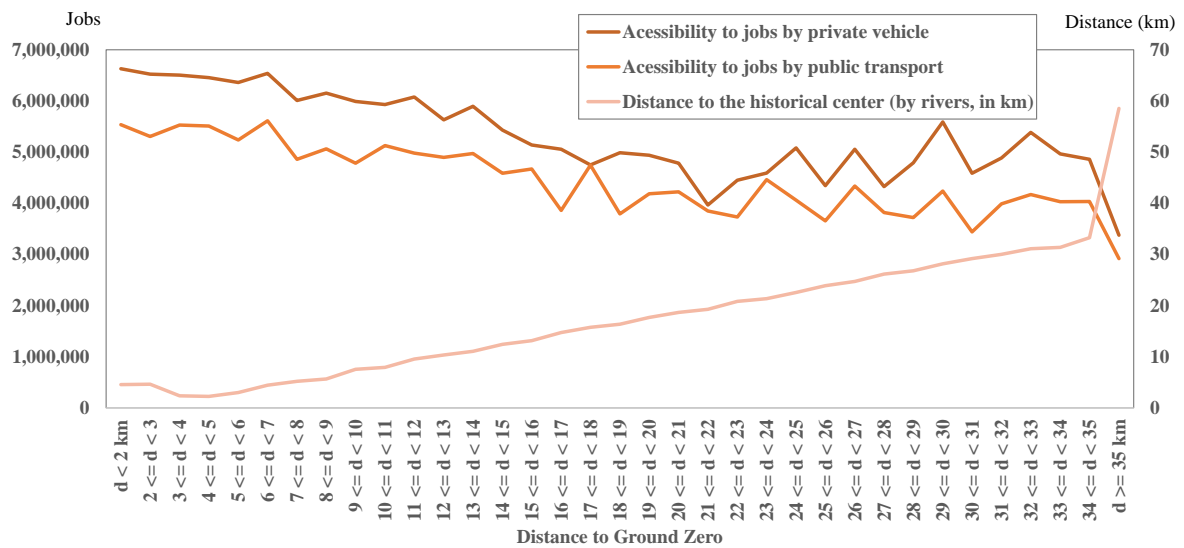
\*Sampling weights applied

Source: OD Survey

We can highlight a few important aspects of Table 1. First, average wage for private vehicle users is more than twice the average wage of public transportation users. Second, commuting time is much higher for public transportation users, despite the location of their residences being on average just a few kilometers further away from downtown than the average location of private vehicle users. Third, private vehicle users have a higher accessibility to jobs from their zone of residence than public transportation users. Finally, as shown in Figure 11, there is a strong negative correlation between accessibility to jobs and the distance to the historical center by the rivers.



**Figure 11. Accessibility to Jobs and Distance to *Pateo do Collegio* by Rivers**



Source: Calculation by the authors

The next section discusses the main results from different estimations. The focus is on the relevance of our instrument, distance by rivers to the historical center, as an exogenous measure of the centrality that shaped urban infrastructure and urban markets.

## 6. Results

The empirical strategy discussed in Section 4 requires the estimation of different models to assess the relevance of accessibility to jobs to explain wages. However, accessibility is potentially endogenous, implying the need to use an exogenous instrument to avoid bias in this analysis.

Following a mincerian equation estimation strategy, Table 3 presents the main results for users of private vehicles and public transportation as transportation modes for daily trips from home to work. The first two models consider commuting time as an explanatory factor for individual wages. It is noticeable that this element has a negative influence only on wages of individuals commuting by public transportation. In fact, dependency upon public transportation in SPMR may limit the amount of effective time spent at work, as discussed on Section 4, resulting in a lower wage. Longer commuting time by mass transit may be also associated to peripheral residences, which may also reduce both the efficiency of the search for jobs and its intensity due to lower accessibility.

The benchmark models usually considered in the literature (Models 3 and 4 from Table 3) include, apart from individual controls, the logarithm of the Euclidean distance to “ground zero” from the zone of residence. In these cases, individuals commuting by private vehicles would perceive a higher negative effect of distance to the CBD on their earnings than individuals using public transportation would. When accessibility to jobs is considered in a richer geography setting (Models 5 and 6),

private vehicle users perceive a higher effect on wages than public transportation users. Taken together, these results are consistent with the fact that cars allow more flexibility to individuals, who can follow a broader range of paths in space, while such flexibility lessens for SPMR public transportation users, since the public transport infrastructure is closer to a radial structure. Therefore, proximity to the center or accessibility to jobs will be translated more easily in real opportunities in the labor market in the case of private vehicle commuters.

Finally, Models 7 and 8 aim at dealing with the endogeneity issue raised earlier. Accessibility to jobs is instrumented by the distance to the historical center (*Pateo do Collegio*), following a path formed by rivers. The main conclusion we get from these results is that the effects measured by accessibility seem to be underestimated in Models 5 and 6.

**Table 3. Estimation Results for the Wage Equation<sup>11</sup>**

(Dependent variable – logarithm of wage)

	Model 1 Private vehicles OLS	Model 2 Public transp. OLS	Model 3 Private vehicles OLS	Model 4 Public transp. OLS	Model 5 Private vehicles OLS	Model 6 Public transp. OLS	Model 7 Private vehicles IV	Model 8 Public transp. IV
Ln(commuting time)	0.007 (0.016)	-0.058*** (0.015)						
Ln(Euclidean distance to Ground Zero)			-0.114*** (0.019)	-0.092*** (0.011)				
Ln(accessibility to jobs by car)					0.260*** (0.068)		0.466*** (0.073)	
Ln(accessibility to jobs by public transportation)						0.183*** (0.025)		0.418*** (0.050)
Instrument								
Ln(distance to Pateo do Collegio by rivers)							Yes	Yes
Observations	5,730	8,564	5,730	8,564	5,730	8,564	5,730	8,564
R-squared	0.526	0.397	0.530	0.401	0.530	0.401	0.527	0.391

Robust t-statistics in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Obs.: Additional controls: age, age squared, gender, role in the household, schooling level, type of job, sector of activity, ln(time of departure from home), workplace zone dummies, constant term. Sampling weights are applied to the estimation.

Source: Calculation by the authors

One of the aims of this paper is to understand the effects of agglomeration economies on urban markets at the SPMR. The strategy was to estimate a wage equation to measure the effects of agglomeration economies on workers' productivity. We could complement this analysis by analyzing the other relevant urban price, namely the price of land. This could be done, for instance, through the estimation of a hedonic price model (to measure the effects of agglomeration economies on housing prices). However, there is potential endogeneity from the inclusion of our index of accessibility to jobs in these models, requiring the adoption of an instrumental variables strategy. To do so, we consider, in the exercise of the wage equation, an exogenous geographical measure, of which economic agents are fully unaware – the distance from the zone of residence to the historical center by “invisible” rivers.

The orthogonality condition between an instrument and the dependent variable given all other controls requires that wages are affected by this geographical variable only through the relationship it has with the accessibility index (and consequently by its relation with the transportation system of the city). Because of that, it is possible to measure how accessibility affects wages through agglomeration effects that arise in the labor market.

The results indicate that the IV models imply that the accessibility coefficient is higher than their OLS counterparts are. This suggests that the OLS models are underestimating agglomeration economies effects on workers' productivity. This may be because wages and accessibility to jobs are simultaneously determined (see Section 2). Therefore, a share of the true effect of accessibility on wages is being lost because a third determinant may be capturing part of that relationship. Given the exogeneity of

<sup>11</sup> In the appendix, Table A.1 provides similar results whenever the commuting time variable is included in specifications such as Models 3 to 8 of Table 3. Therefore, the main results from Table 3 seem to be robust to this slight change.

the instrument, it can overcome this problem and generate a higher estimated coefficient for the instrumented accessibility variable.

Therefore, significant biases are generated by estimating directly the relationship between wages and accessibility.<sup>12</sup> This research shows that an instrument based in geography with historical roots is able to control for at least part of this issue.

## 7. Final Remarks

In 1959 Hansen asked the question: “How does accessibility shape land use?” In the last several decades, a great deal of research has focused on the measurement of the impact of accessibility on urban markets, including analyses that have attempted to identify the main channels through which accessibility to jobs may affect urban prices.

A great part of the empirical studies has adopted the identification hypothesis of concentration of jobs in the central city to get rid of the uneasy endogeneity problem that could have potentially emerged. While appeal to dual outcomes could be made, it seems that larger metropolitan areas in the developing world rarely comply with the pure monocentric paradigm. A considerable level of employment decentralization is a common characteristic of megacities like São Paulo. The very size of SPMR provides opportunities that could not be realized by similar workers located within other metropolises elsewhere. The case of SPMR has been further complicated by an urban transport infrastructure that until recently was poorly regulated and biased towards investments in roads to the exclusion of other modes.

There are significant spatial inequalities in a large metropolitan region such as SPMR. Depending on the transportation mode, workers may see harmed their productivity or their opportunities to find a better matching in the job market. As for the housing market, individuals who can afford a higher price dwelling will be better located in the city, closer to their jobs.

Our empirical analysis, despite being motivated by a theoretical general equilibrium framework, was based on partial equilibrium modelling strategies. A complete structural model comprising the simultaneous determination of the housing and the labor markets can provide further insights on these relationships. Nonetheless, the inclusion of an exogenous instrument allowed us to control for at least a share of the bias generated in this context. Rivers, nowadays hidden by roads, seem to have helped shaping urban markets in SPMR.

## References

Du, H.; Mulley, C. Relationship between transport accessibility and land value: local model approach with geographically weighted regression. **Transportation Research**, No. 1977, pp. 197-205. 2006.

---

<sup>12</sup> Similar argument is valid for the case of land and housing prices empirical models.

Duranton, G.; Puga, D. Micro-foundations of urban agglomeration economies. In: Henderson, J.V., Thisse, J.F. (Eds.) **Handbook of Regional and Urban Economics**. Vol. 4. Amsterdam: North-Holland, p. 2063-2117, 2004.

Duranton, G.; Puga, D. Chapter 8 – Urban land use. In: Duranton, G.; Henderson, J.V.; Strange, W.C. **Handbook of Regional and Urban Economics**, Vol. 5. Amsterdam: North-Holland, p. 467-560, 2015.

Duranton, G.; Turner, M. A. Urban growth and Transportation. **Review of Economic Studies**. Vol. 79, pp. 1407-1440, 2011.

Fujita, M.; Ogawa, H. Multiple equilibria and structural transition of non-monocentric urban configurations. **Regional Science and Urban Economics**. Vol. 12, Issue 2, pp. 161-196, 1982.

Haddad, E.A.; Hewings, G.J.D.; Porsse, A.A.; Van Leeuwen, E.S.; Vieira, R.S. The underground economy: tracking the higher-order economic impacts of the São Paulo subway system. **Transportation Research Part A**, Vol. 73, pp. 18-30, 2015.

Hansen, W.G. How accessibility shapes land use. **Journal of the American Institute of Planners**, Vol. 25, Issue 2, pp. 73-76, 1959.

Lucas, R.E.; Rossi-Hansberg, E. On the internal structure of cities. **Econometrica**, Vol. 70, No. 4, pp. 1445-1476, 2002.

Melo, P.C.; Graham, D.J. Agglomeration economies and labour productivity: evidence from longitudinal worker data for GB's travel-to-work areas. **SERC Discussion Paper**, No. 31, October 2009.

Raia Jr., A. **Acessibilidade e mobilidade na estimativa de um índice de potencial de viagens utilizando redes neurais artificiais e sistemas de informação geográfica**. Tese (Doutorado em Engenharia Civil) – Escola de Engenharia de São Carlos – USP. São Carlos. 2000.

Rosenthal, S.S., Strange, W.C. Chapter 49 – Evidence on the nature and sources of agglomeration economies. In HENDERSON, J.V., THISSE, J.F. (Eds.) **Handbook of Regional and Urban Economics**. Volume 4. New York: Elsevier, p. 2119-2171, 2004.

United Nations, Department of Economic and Social Affairs, Population Division. **World Urbanization Prospects: The 2014 revision**, (ST/ESA/SER.A/366), 2015.

Van Ommeren, J.; Gutiérrez-I-Puigarnau, E. Are workers with a long commute less productive? An empirical analysis of absenteeism. **Tinbergen Institute Discussion Paper**, TI 2009-014/3, 2009.

Vieira, R., Haddad, E. A. An Accessibility Index for the Metropolitan Region of São Paulo In: K. Kourtit, P. Nijkamp and R. R. Stough (editors), **The Rise of the City: Spatial Dynamics in the Urban Century**, Cheltenham, UK: Edward Elgar Publishing, pp. 242-258, 2015.

Zenou, Y. How do firms redline workers? **Journal of Urban Economics**, Vo. 52, pp. 391-408, 2002.

Zenou, Y. **Urban Labor Economics**. Cambridge University Press, New York, 2009.

## Appendix

**Table A.1. Estimation Results for the Wage Equation**

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>	<b>Model 5</b>	<b>Model 6</b>
	<b>Private</b>	<b>Public</b>	<b>Private</b>	<b>Public</b>	<b>Private</b>	<b>Public</b>
	<b>vehicles</b>	<b>transp.</b>	<b>vehicles</b>	<b>transp.</b>	<b>vehicles</b>	<b>transp.</b>
	OLS	OLS	OLS	OLS	IV	IV
Ln(commuting time)	0.019 (0.016)	-0.032** (0.015)	0.024 (0.016)	-0.040*** (0.015)	0.038** (0.016)	-0.014 (0.016)
Ln(Euclidean distance to Ground Zero)	-0.118*** (0.019)	-0.083*** (0.012)				
Ln(accessibility to jobs by car)			0.277*** (0.069)		0.497*** (0.076)	
Ln(accessibility to jobs by public transportation)				0.165*** (0.026)		0.399*** (0.056)
Instrument						
Ln(distance to Pateo do Collegio by rivers)					Yes	Yes
Observations	5,730	8,564	5,730	8,564	5,730	8,564
R-squared	0.531	0.402	0.530	0.402	0.527	0.393

Robust t-statistics in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Obs.: Additional controls: age, age squared, gender, role in the household, schooling level, type of job, sector of activity, ln(time of departure from home), workplace zone dummies, constant term. Sampling weights are applied to the estimation.

Source: Calculation by the authors