Urban sprawl and spatial segregation in São Paulo Metropolitan Region

Frederico Roman Ramos*
Ciro Biderman*
*Centro de Estudos de Política e Economia do Setor Público - Fundação Getúlio Vargas - SP

ABSTRACT
We propose an econometric approach to investigate the causal relationships between urban sprawl and socio-spatial segregation. The problem of this analysis is that a household that decides to settle in remote areas may want to isolate itself. So, it is not clear if sprawl is causing segregation of the poor or if there are other non-observed characteristics driven the results. Using an instrument from topographical characteristics we show that urban sprawl do cause segregation. It is interesting to note that it causes the isolation of the poor but does not cause isolation of the rich. The analyses were applied to the São Paulo Metropolitan Region (SPMR) for the year 2000 using data from different sources treated in a geo-referenced computational environment. The segregation indices showed a significant association to the urban form configuration indicating that sprawled areas reinforce segregation. This result shed some lights on the traditional housing policy locating social housing projects in the far periphery in order to save in land costs. According to our results, this policy decision would reinforce segregation.

Key words : urban sprawl, spatial segregation, housing policy

RESUMO
Propomos uma abordagem econométrica para investigar as relações causais entre dispersão urbana e segregação socioespacial. O problema nesta análise reside no fato de que o morador que se estabelece em áreas remotas pode estar decidindo pelo isolamento. Assim, não fica claro se a dispersão urbana causa a segregação ou se há outras características não-observáveis influindo neste resultado. Aplicando um instrumento derivado das característica de relevo demonstramos que a dispersão urbana causa segregação. Os resultados demonstram que a dispersão causa o isolamento dos pobres mas não o isolamento dos ricos. As análises foram aplicadas para a Região Metropolitana de São Paulo (RMSP) para o ano de 200 usando dados de diferentes fontes tratados em ambiente computacional georreferenciado. Os índices de segregação apresentaram associação significativa com a configuração da forma urbana indicando que áreas com urbanização mais dispersa reforçam a segregação. Os resultados podem trazer luz ao debate sobre as tradicionais políticas de habitação social que desenvolvem projetos em áreas periféricas distantes como modo de economizar recursos com os custos da terra. Nosso resultados apontam que estas políticas reforçam a segregação.

Palavras chave: dispersão urbana, segregação socioespacial, políticas habitacionais

ANPEC: Área 10 - Economia Regional e Urbana
JEL classification: R14, R21, R23

1 - Urban sprawl and spatial segregation: the urban structure and its constituents.

In the last decade urban land prices have increased substantially in Brazil. This movement has deepened the dispute for urban locations and as a consequence promoted changes within the urban spatial structures of the cities in various regions of Brazil. The understanding of how these movements processed passes through the ability to formulate hypotheses derived from a conceptual framework to be further developed by Brazilian scholars. The canonic models of urban economics assume conditions that are often absent in cities in developing countries. Among the most problematic are those where all the property owners observe the urban rules and regulations, where transactions are realized under formal arrangements, where legal authorization is required for changes in land use and the property taxes are paid. Furthermore, these models assume that the urban land is fully served by infrastructure at the time they are occupied (Smolka and Biderman, 2012).

Even in developed markets exists a high level of complexity in urban situations that complicates their representation and analytical operation. For Richardson (1978) this is one of the reasons why economics for decades neglected this field of study. For the author, the conventional economical analyses can hardly reveal the complexity of the big city riddled with externalities and "excrescences" that make the neoclassical marginalism approach unfeasible. According Krugman (1998) the late introduction of the spatial dimension into the mainstream of economics is a consequence of the late evolution of models capable to deal with scenarios of imperfect competition. Until the late 1980s, most models were not able to analyze a market with economies of scale and increasing returns. In this perspective, the urban spatial structure should be analyzed as a result of the action of centrifugal and centripetal forces or the attraction and repulsion that ultimately define the dispute for locations of firms and households through the interurban space (Krugman, 1998; Fujita *et al*., 1999).

The monocentric model prevails for decades as the most influential representation of the spatial urban structure within the literature on urban economics. Initially formulated by Alonso (1964) as a direct derivation of the agricultural land use model of Von Thünen, the monocentric model is the most simple and operational stylized fact of the spatial urban structure. One of the critiques for this model, the existence of a single business center, can be relaxed without relevant consequences to the relationship between land rent and transport costs (Fujita and Ogawa, 1982). The analytical framework provided by this model clarify the relationship of several aspects of the urban spatial structure as social segregation, urban expansion and sprawl, land price valuation and variations in population densities.

The way urban areas expand is one of the central points of the urban economics debate and continues to guide the research agenda. Already in the mid-sixties appeared the first studies dealing with the issue of urban sprawl, which then began to awaken specific interest of the academic community especially among North American researchers (Harvey and Clark, 1965; Clawson, 1962; Whyte, 1958). At that point in time, the negative economic and aesthetic consequences derived from this pattern of urbanization were already pointed out. The spatial configuration of the expanded and sprawled urban structure was then associated with the result of a number of interfering factors able to change the basic parameters of the classical model. Among these factors are the economic and technologic impact related to the dissemination of the automobile (Hall, 2002; Bawm-Snow, 2007), the availability of areas to be converted into urban land, speculation and market failures (Clawson, 1962; Bahl, 1968; Archer, 1973; Mills and Hamilton, 1993).

In addition to the processes of expansion and dispersion of the urban area, the social segregation is another aspect often inserted in the center of the analyses of the spatial urban structure. Muth (1975) argues that the spatially concentrated poverty in cities is the main urban problem to be faced. For this author, many other urban problems derive directly from the inability of the government to establish appropriate prices for the use of scarce resources within urban areas. Within this perspective, the main aspect that could explain the uneven distribution between rich and poor in urban areas would be related to the differences between the elasticity of income and costs of urban transport and the elasticity of income and costs of housing for each population group according to their income level. This approach first tested by Leroy and Sonstelie (1983) establishes the relationship between the processes of suburbanization and gentrification of central areas and the costs associated to collective or individual transport modes. Glaeser *et. al.* (2000) also addresses the question of the unequal distribution of the poor in the city as a central point in urban economics. In this research, the observed patterns of segregation results from the costs associated with individual motorized trips in opposition to public transit and differences between the costs of opportunity for the rich and the poor. In Brazil, Abramo (2007) suggests that socio-spatial segregation is a result of intentional behavior of families whose residential location decision arises from the decision to establish neighborhood relationship with families of the same socioeconomic profile.

In 1969, Thomas Schelling developed his classical research to describe the processes of segregation through the use of models obtained from what he called "individual discriminatory behaviors". Schelling's work pioneered in proposing dynamic models of spatial segregation. The American tradition of focusing on the racial aspect in studies on segregation gained strength in the 1950s with the civil right movements against segregationist laws (Massey and Denton, 1987). In this period some studies developed metrics of socio-spatial segregation as a strategy to represent the phenomenon, where the most popular for this kind of analyses is the Dissimilarity Index proposed by Duncan and Duncan (1955). Since then, the issue of segregation continued on the research agenda incorporating other issues related not only to the ethnical and racial question but also to the extent of social and economical segregation. To capture these dimensions of segregation, other indexes have been proposed based on the recognition of the limitations of the dissimilarity index to represent attributes of socio-economic nature that tends to present a continuous variation such as income and educational level. The neighborhood sorting index NSI proposed by Jargowsky (1996) is based on this perspective. In Latin America most studies on segregation were more concerned with socioeconomic segregation rather than race segregation (Feitosa et al. 2007; Marques e Torres 2004; Sabatini e Salcedo 2007; Torres et al. 2002; Villaça 1998). Villaça (1998) identifies that the most significant elements structuring the urban space are the locations of industrial zones and segregated neighborhoods of population in the highest income level (Villaça, 1998: 140).

In the various disciplines engaged in urban research, both high levels of socio-spatial segregation as urban sprawl are understood as negative aspects with perverse consequences in costs of infrastructure deployment, the provision of services, utilities and levels of social inequality. Although there exist a considerable amount of research on the scattered pattern of urban growth and socio-spatial segregation, the economic perspective dominating the discussion faces the challenge of establishing empirical and operational representations of these concepts. Some research has been dedicating efforts in proposing different metrics. The research of Galster et al. (2001) develops a conceptual and operational definition of eight dimensions of urban sprawl. Others propose the construction of urban dispersion metrics based on the use of satellite imagery as a source of primary data as in Burchfield et al (2006) and Angel et al (2011). The latter derives from the analytic tradition of landscape ecology proposing the use of satellite imagery to calculate metrics into five dimensions necessary for understanding the processes of urban sprawl: extent of urbanized land, density, fragmentation, centrality and compactness. For each of these dimensions a number of metrics and indicators are proposed with the specific purpose to operate these dimensions in a way that analytical strategies could be established, able to set a comparative table of 120 large urban areas around the world.

The objective of this study is to investigate the causal relationships that may exist between urban sprawl and socio-spatial segregation. The intuition is that the origin of both processes is related to the same forces of attraction and repulsion that define the spatial arrangements of the urban structure and the distribution of different population groups according to their socioeconomic characteristics. We recognize that Ordinary Least Square (OLS) techniques applied to data where there are no guarantees of randomness as produced by controlled experiments, can lead to false interpretations of causal relationship. However, the argument goes, using an instrumental variables, it is possible to assume conditional independence in empirical experiment and achieve the desired causal interpretation (Angrist and Pischke, 2009).

In the next section we discuss the empirical strategy behind this paper arguing that the instrument proposed respect the exclusion restriction because it is constructed from a topographical base. The following section present the results both for an OLS estimation as a two stage least square (2SLS) estimation showing the difference between the results as an evidence that the OLS estimation is biased. The final section concludes the paper.

2 - Empirical Strategy

We are interested in the impact that urban sprawl has on the socio-spatial segregation in different parts of a city. Thus, we assume that at a point in time, a city occupies a certain region A and that region is subdivided into n sub-regions A_i and $i=1,2,...,n$. Every sub-region A_i can be characterized by its degree of urban sprawl, S_i. This measure of sprawl represents the proportion of urban land use in A_i and its surroundings in relation to its total area. For each sub region A_i, we know its level of segregation, Y_i, calculated in relation to the overall composition of the population in social groups considering the characteristics of the population composition around A_i. The subdivision of the population into groups is determined by their socioeconomic characteristics such as income level and education attainment. Moreover, there is a measure R_i calculated for each A_i displaying its geomorphologic characteristics. This measure represents the configuration of the physical terrain and is related to natural aptitude for the occupation by urban land use.

We are aiming to estimate the conditional expected value of segregation in A_i given a certain level of sprawl. Formally, we want to know $\mathrm{E}[Y_i \mid S_i = s]$. In this case we are attempting to understand if the level of urban sprawl (our treatment) causes socio-spatial segregation in A_i. Since we are working with empirical data and not with experimental data, we seek to ensure the conditional independence assumption. In this case, selection bias is highly expected since the decision of the household to locate in remote areas might be connected with the decision to isolate itself. That is the reason why it is so difficult to establish causality from urban sprawl to segregation. In this paper we explore the fact that the geographical physical configuration of the sub-regions A_i is expected to be exogenous to segregation, so it satisfies the exclusion restriction. On the other hand, surface ruggedness (a topographical measure explained bellow) is associated with urban sprawl guaranteeing the first stage condition. Consequently, surface ruggedness is a good candidate for an instrument and we can actually check the casual relation of interest.

Using the method 2SLS version of instrumental variables, it is possible to estimate the effect of the treatment (S_i) on the observed socio-spatial segregation indexes. So R_i is an instrument that acts on s_i. This is the first-stage. Secondly, we assume the exclusion restriction, or in other words, the only reason for Y_i and R_i to correlate is the first-stage. Thus the instrument somehow recovers the desired randomity. Formally:

$$\hat{s}_i = X_i'\hat{b}_1 + \hat{b}_2 R_i \qquad \text{(1}^{\text{st}}\text{ stage)}$$

$$Y_i = \alpha' X_i + \rho \hat{s}_i + [\eta_i + \rho(s_i - \hat{s}_i)] \qquad \text{(2}^{\text{nd}}\text{ stage)}$$

Where,

$\hat{s}_i$ is the estimated treatment variable (S_i) in the 1st stage;
$\hat{b}_1$ e $\hat{b}_2$ are estimated coefficients for the covariates and I.V. at the 1st stage;
R_i is the value of the instrumental variable that indicates the geomorphologic configuration of the terrain;
Y_i is the level of social-spatial segregation at area unit *i*;
ρ is the expected effect of the treatment;
X_i is a covariate related with the geometric conformation of the area unit;
η_i is the compound error term.

The analyses were applied to São Paulo Metropolitan Region (SPMR) for the year 2000 as a way to verify the feasibility of applying the methodology suggested. The choice of the SPMR as a case study is justified by its importance in the Brazilian urban hierarchy, population size and its level of urban complexity. However, the way we used and treat the data allow the study to be replicated to other metropolitan contexts in Brazil.

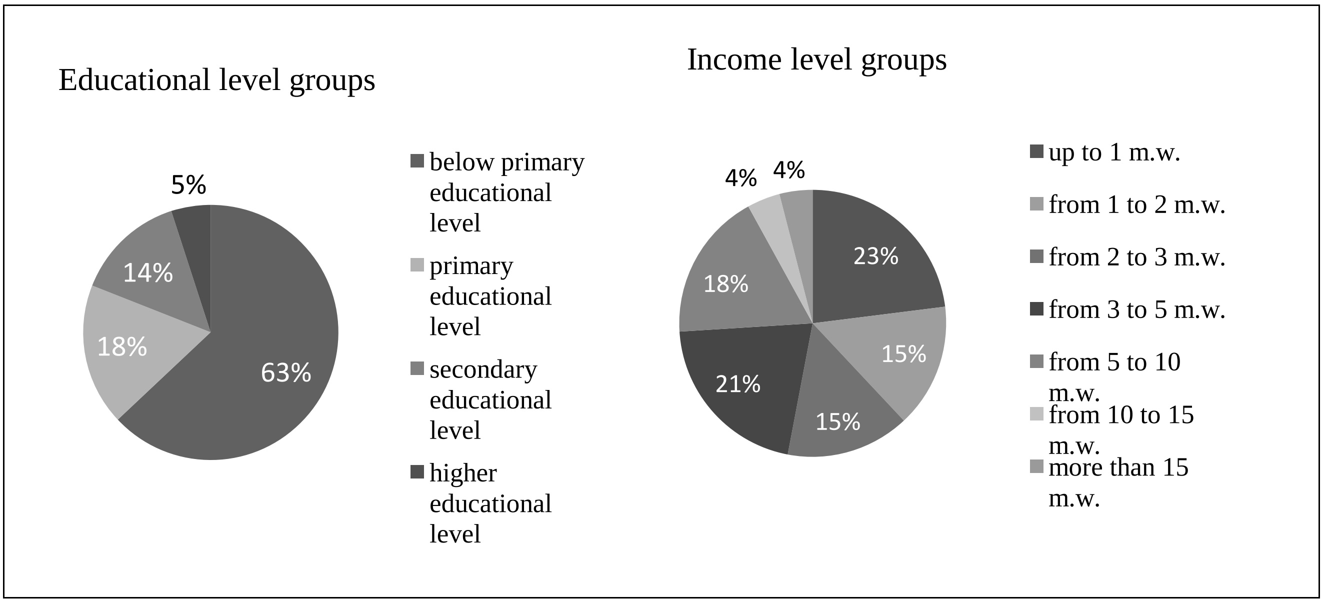


Figure 1 - Population groups based on relevant levels of education attainment and income level of heads of the household of the census of 2000

The data used for the analyses come from different sources and were treated in a geo-referenced computational environment. The indexes of social-spatial segregation were calculated from the census of 2000 (IBGE[1]) in its most disaggregated level possible, the census block. We used two measures of social-spatial segregation as proposed by Feitosa et al.(2007), the local dissimilarity index, $\breve{d}_i(m)$, and the local isolation index, $\breve{q}_i(m)$. The isolation index is quite relevant for this analysis since it allows us to distinguish between the impact of sprawl on the isolation of the poor and on the isolation of the rich. Most segregation indices are computed for all groups.

[1] Instituto Brasileiro de Estatística e Geografia

We defined population groups based on relevant levels of education attainment and income level of heads of the household for the 2000 census. Four population groups were defined in relation to educational level (Figure 1): *e0* – head of household (HH) with incomplete primary education; *e1* – HH with primary education; *e2* – HH with high school; *e3* – HH with higher education. Seven other groups were defined in relation to income level: *r0* - HH with income below one minimum wage, *r1* - HH with income between one and two minimum wages, *r2* - HH with income between 2 and 3 minimum wages, *r3* - HH with income between 3 and 5 minimum wages, *r4* - HH with income between 5 and 10 minimum wages, *r5* - HH with income between 10 and 15 minimum wages and *r6* - HH with income above 15 minimum wages. These groups were used to calculate the local segregation index. The calculated indexes according to income are indicated with the suffix *_r* and the calculations of the educational level with the suffix *_e*.

Initially the local population intensities, $\check{L}_i = \sum_{i=1}^{I} d(N_i)$, and the local intensities of the population groups, $\check{L}_{mi} = \sum_{i=1}^{I} d(N_{mi})$, were calculated for every sub-region A_i (corresponding to the census blocks of the MRSP). The population intensity is a sum of individuals (N_i) in A_i and its surroundings defined by a decreasing function based on the distance *d(.)* with its origin at the central points of every A_i . In the same way, it is possible to calculate the intensity of local population groups for every group of individuals *m*, and N_{mi} according to the number of individuals of group *m* in A_i and its surroundings defined by the same distance function *d(.)*. It is now possible to calculate the relative proportion, $\check{\tau}_{mi}$, of the intensity of population group *m* in every sub-region *i* in relation to the intensity of the total population by the ratio $\check{\tau}_{mi} = \frac{\check{L}_{mi}}{\check{L}_i}$. The local dissimilarity $\check{d}_i(m)$ and isolation $\check{q}_i(m)$ indices can be calculated for every sub-region A_i applying the following equations:

$$\check{d}_i(m) = \sum_{m=1}^{M} \frac{N_i}{2NI} |\check{\tau}_{mi} - \tau_m|$$

where $I = \sum_{m=1}^{M} (\tau_m)(1 - \tau_m)$ and

$$\check{q}_i(m) = \frac{N_{mi}}{N_m} \left(\frac{\check{L}_{mj}}{\check{L}_j} \right)$$

In the case of this research, the distance *d(.)* kernel function was applied as a Gaussian with a radius of a 1000 meters range. The local segregation indices are calculated in order to capture the local contribution to the global segregation index. Algebraically, from the sum of every local contribution it is possible to reset the global segregation indices (Feitosa, *et al.*, 2007). For this reason an adequacy of the local indices scales is necessary for the use in econometric applications. In the case of SPMR, for the year 2000 we have a total of 21,744 census blocks covering urban and rural areas of the regions. For this reason the values of local contributions calculated for every census tract were multiplied by 10^5 to ensure greater intelligibility in the multivariate statistical analyses. The isolation indexes were calculated for the two extremes groups defined by both criteria. For the level of education, the isolation of the group with incomplete primary education and of the group with higher education was calculated. In the case of income criterion, the isolation of the group with an income up to one minimum wage and on the other hand to the group with more than 15 times the minimum wage was calculated.

The data relating to the urban sprawl indices were generated through techniques of digital processing of satellite imagery *LANDSAT5* obtained for the year of analyses. The data used to calculate the surface ruggedness indexes were obtained by the *USGS – U.S. Geological Survey* that provides digital models of the surface generated by the *Shuttle Radar Topography Mission (SRTM)*. All the calculations necessary for the assembly of the databank were processed by geographical information systems *ArcView 9.3, TerraView 4.1.0* and *SRING 5.1.5*.

A measure of urban sprawl was calculated with the use of satellite image *TM-LANDSAT 5* (June 2000) already classified in three classes of landcover: water, urban (impervious surface) and non-urban. The image was obtained directly from the data bank formed to elaborate a comparative research on urban sprawl in 120 cities financed by the *Lincoln Institute of Land Policy*[2]. The classified image was imported to *SPRING 5.1.7* where it was converted into a regular grid assigning the value **null** to the pixels of category water, value **zero** to the non-urban pixels and, value **one** to the pixels identified as urban. Once converted to a regular grid, a routine of neighborhood analyses was applied. Every pixel, which in the image refers to an area of an equilateral square with side of 30 meters, received the value of the sum class pixels, U_p, considering a radius covering 1000 meters. Thus, pixels with a high value of U_p indicate insertion in highly urbanized areas, while low values of U_p indicate areas with a disperse urbanization, in other words, with higher level of urban sprawl. The conversion of the data on a regular grid to the zonal support of the census blocks occurred from an operation recovering the average value of the pixels that are contained in A_i. So the sprawl index s_i can be formulized as:

$$s_i(d) = 1 - \frac{1}{N_{pi}} \sum_{i=1}^{I} U_{pi}$$

Where, N_{pi} is the number of pixels in the census tract A_i and its surroundings

Finally, the standard deviation of the values of terrain elevation was calculated for every census tract. As the spatial resolution of the data of SRTM (90 meters) is less than presented by the landcover images (30 meters), a re-sampling process by deterministic interpolation was applied to guarantee the same spatial resolution between the two information layers. The use of standard deviation is an adaptation to the surface ruggedness index proposed by Riley, De Gloria e Elliot (1999) and can be written as a measure of dispersal between the values of elevation at a certain point in this way $\sqrt{\sum_{i=r-1}^{r+1} \sum_{i=c-1}^{c+1} (e_{i,j} - e_{r,c})^2}$. Where $e_{i,j}$ and $e_{r,c}$ are values of elevation of a pixel on the line *i* and column *j* of a grid and *r* and *c* are its neighbors.

3 - Results

The econometric analyses were realized in *Stata11* using the algorithm *ivregress2sls* to estimate the *2SLS* model. The first step was the definition of the subset to be analyzed determined by the distance from the limit of the urban fringe detected by the satellite image. This decision aroused from the fact that the urban – rural limit is where new urban development occurs, in other words, it is in this region that we can expect areas with higher mixture of urban and non-urban use. In other words, the analysis makes sense just around the urban fringe. We define four subsets to use in the analyses. The first consists of 5,176 census blocks located at 500 meters or less from the observed limit of the urban fringe both towards the inner side of the urban area (inbound) as towards the outer side (outbound), including in this way census blocks predominantly occupied by non-urban landcover. These blocks often present large territory extension and low population density. The other three subsets were defined restricting the selection only to inbound census blocks. At a distance of 1 km of the limit there were 6,184 census blocks; the second subset contains 13,494

[2] http://www.lincolninst.edu/pubs/2094_Planet-of-Cities

census blocks at a distance of 5 km and; the largest analyzed subset contains 18,773 census blocks at a distance of 10 km of the urban fringe.

We started the analyses running an OLS regression of the local segregation indexes on the average urban sprawl index of the census block for the four selected subsets. The results (Table 1) pointed to a significant correlation between segregation and sprawl with coefficients significantly different from zero in almost all the regressions. The exceptions were found in the analyses carried out in the subset of the urban fringe relative to the dissimilarity index calculated according to income (*d_r_5*), to the isolation calculated for the group with the lowest level of education (*q_e0_5*) and, to the subset of census blocks within 1 km inbound to the urban fringe relative to the isolation of the group with the highest income (*q_r6_5*).

The analyses OLS allows to establish a first reading of the behavior of the depending variable relative to the selected segregation indexes. This first analytical approach points to a relationship where the direction of the signals of the found coefficients in accordance with the initial expectations where areas with higher average in the sprawl index or, in other words, areas with more scattered urbanization, tend to be areas with higher local dissimilarity indices. This is true both for rich as for poor households as one can see in the regression results using the isolation index for poor (*q_r0_5* and *q_e0_5*) or for the rich (*q_r6_5* and *q_e3_5*).

Table 1 – Urban Sprawl and Socio-Spatial Segregation for Different Distances to the Urban Fringe. Results from OLS regression

Range of 0.5 Km distance from the Urban Fringe (inbound and outbound)						
Dependent Variable:	d_r_5	q_r0_5	q_r6_5	d_e_5	q_e0_5	q_e3_5
Urban Sprawl	0.001	0.004	-0.002	0.001	0.001	-0.002
	-1.74	(4.88)**	(2.10)*	(3.00)**	-0.76	(4.02)**
R^2	0	0	0	0	0	0
N	5,176	5,176	5,176	5,176	5,176	5,176
Range of 1Km distance from the Urban Fringe (inbound)						
Dependent Variable:	d_r_5	q_r0_5	q_r6_5	d_e_5	q_e0_5	q_e3_5
Urban Sprawl	0.004	0.007	0.000	0.005	0.011	-0.002
	(10.18)**	(7.86)**	-0.25	(11.18)**	(6.94)**	(2.55)*
R^2	0.02	0.01	0	0.02	0.01	0
N	6,184	6,184	6,184	6,184	6,184	6,184
Range of 5Km distance from the Urban Fringe (inbound)						
Dependent Variable:	d_r_5	q_r0_5	q_r6_5	d_e_5	q_e0_5	q_e3_5
Urban Sprawl	0.006	0.012	-0.003	0.008	0.023	-0.005
	(19.50)**	(18.37)**	(2.56)*	(21.24)**	(19.82)**	(6.33)**
R^2	0.04	0.04	0	0.04	0.04	0
N	13,494	13,494	13,494	13,494	13,494	13,494
Range of 10Km distance from the Urban Fringe (inbound)						
Dependent Variable:	d_r_5	q_r0_5	q_r6_5	d_e_5	q_e0_5	q_e3_5
Urban Sprawl	0.005	0.016	-0.015	0.006	0.031	-0.017
	(15.14)**	(26.24)**	(12.61)**	(16.17)**	(28.31)**	(17.96)**
R^2	0.01	0.06	0.01	0.01	0.06	0.01
N	18,773	18,773	18,773	18,773	18,773	18,773

* $p<0.05$; ** $p<0.01$

As we discussed, we are very suspicious with this result. There might be a selection bias where households deciding to live in remote areas prefer to be isolated (and that is the reason why they choose to live there in the first place). To solve this problem we proceeded to applying regression in two stages (*2SLS*) using the surface ruggedness as an instrument. The first point of interest in this approach is the relationship of the instrument with the instrumented variable (scattered development). One way to assess the "strength" of an IV is through the evaluation of the outcomes of the first-stage observing the value of the F statistics. According Angrist and Pischkle (2009, p.208) we can say that 2SLS estimates are "biased towards OLS estimates" as the F-statistic gets small, or in other words there isn`t much of the first-stage. In our case, we seem to be in a safe condition when analyzing the performance of the F and R^2 obtained in the first-stage for our subsets (table 2).

Table 2 – First-stage results for the *2SLS* regressions

	Fringe	*1km*	*5km*	*10km*
# observations	5176	6184	13494	18773
F	468.48	889.09	2055.08	2611.0
$Prob > F$	0.000	0.000	0.000	0.000
R^2	0.2398	0.1297	0.165	0.1722

The results obtained from the 2SLS regression (table 3) confirmed the initial expectation of a significant causal association between urban sprawl and segregation indices. We confirm the causality of scattered development and segregation. In other words, living in remote areas cause (more) segregation. Using the instrument this result is confirmed for any specification. Furthermore the magnitude increases four fold. When we move to the isolation index the results change qualitatively: isolation of the poor is still reinforced by sprawl. However, isolation of the rich is not reinforced by sprawl. Actually, in some specifications, living in remote areas may reduce isolation of the rich.

This result is probably reflecting the fact that the poor move to remote areas with no infrastructure, no public services or even private services. In other words, poor households live in areas lacking the "city". None wants to move in to this area unless you cannot afford to live in any other region. So, the causal relationship between sprawl and isolation of the poor have a consequence for traditional housing policies that locate large social housing projects in the far periphery in order to save in land prices and build more units. This policy may house a lot of poor households but it may be reinforcing their economic condition helping to trap those families in the vicious cycle of poverty.

Although, to the best of our knowledge, no other study has shown this result in such a robust way, most of the literature concerned with housing policies point out the risk of settling poor households in remote areas. More surprising is the fact that there is probably no causality between rich isolation and sprawl. If there is any causal relationship is the other way around: sprawl would reduce isolation of the rich. This result dispute the common wisdom that gated communities cause more segregation. If the bias not captured by the OLS regression is due to the fact that households that decide to live in remote areas do want to be isolated we can explain why the literature confound the impact of gated communities. Rich households do move to remote areas to isolate themselves. However, this decision will bring the "city" to those areas and attract more middle and poor households to this neighborhood.

Table 3 – Urban Sprawl and Socio-Spatial Segregation for Different Distances to the Urban Fringe. Results from 2SLS regression

Range of 0.5 Km inbound and outbound

Dependent Variable:	d_r_5	q_r0_5	q_r6_5	d_e_5	q_e0_5	q_e3_5
Urban Sprawl	0.006	0.016	0.005	0.007	0.023	0.001
	(5.53)**	(6.45)**	-1.64	(5.97)**	(5.74)**	-0.44
N	5,176	5,176	5,176	5,176	5,176	5,176

Range of 1Km inbound

Dependent Variable:	d_r_5	q_r0_5	q_r6_5	d_e_5	q_e0_5	q_e3_5
Urban Sprawl	0.020	0.043	0.005	0.025	0.095	-0.002
	(12.96)**	(12.19)**	-1.64	(13.96)**	(14.39)**	-0.87
N	6,184	6,184	6,184	6,184	6,184	6,184

Range of 5Km inbound

Dependent Variable:	d_r_5	q_r0_5	q_r6_5	d_e_5	q_e0_5	q_e3_5
Urban Sprawl	0.019	0.041	-0.003	0.024	0.097	-0.009
	(19.79)**	(19.02)**	-1.23	(21.12)**	(23.09)**	(4.45)**
N	13,494	13,494	13,494	13,494	13,494	13,494

Range of 10Km inbound

Dependent Variable:	d_r_5	q_r0_5	q_r6_5	d_e_5	q_e0_5	q_e3_5
Urban Sprawl	0.012	0.048	-0.038	0.014	0.113	-0.047
	(14.04)**	(24.73)**	(10.75)**	(13.28)**	(29.22)**	(15.32)**
N	18,773	18,773	18,773	18,773	18,773	18,773

* $p<0.05$; ** $p<0.01$

Finally, we introduced covariates in the model for the urban fringe and 1Km subset in order to evaluate the stability of the model. Two covariates related with the geometric configuration of the census block were used: the area (ln_area) and the perimeter (ln_perimeter). The definition of the census block geometry is based on operational criteria for the application of the census questionnaires. In this sense, the definition of the boundaries of the tracts is not necessarily related with neighborhoods or other sociological territorial category. Duo to this aspect, we could expect some interference in the metrics of sprawl and segregation, especially when we include in the analysis blocks with the predominance of non-urban uses with low population densities. As the metrics are calculated according to a bandwidth from the central point of each block, in our case 1km, it is expected that the geometrical configuration of the areal unit have some effects on the indices. We found different results for each subset (table 4). For the fringe subset which includes rural census blocks located externally to the urban spot, the introduction of the covariates has an impact in the model annulling the statistical significance of the coefficients of interest. For the subset that includes only internal blocks 1Km distant from the urban spot limit, the inclusion of covariates did not have interference in the statistical significance of the coefficients of interest, only some marginal effects in its magnitude.

Table 4 – Results obtained from regression with *2SLS* including covariates in the model

Range of 0.5 Km inbound and outbound

	d_r_5	d_r_5	d_r_5	p_r_5	p_r_5	p_r_5	q_r0_5	q_r0_5	q_r0_5
Urban Sprawl	0.006	-0.003	-0.004	0.002	0.001	0.001	0.016	0.017	0.014
	(5.53)**	-0.88	-1.15	(6.02)**	-0.66	-0.73	(6.45)**	-1.87	-1.61
ln_area		0.097	0.043		0.017	0.018		-0.01	-0.276
		(3.30)**	-0.88		-1.86	-1.06		-0.13	(2.22)*
ln_perimeter			0.116			-0.002			0.571
			(2.11)*			-0.12			(4.45)**
N	5,176	5,176	5,176	5,176	5,176	5,176	5,176	5,176	5,176

Range of 1Km inbound

	d_r_5	d_r_5	d_r_5	p_r_5	p_r_5	p_r_5	q_r0_5	q_r0_5	q_r0_5
Urban Sprawl	0.020	0.013	0.012	0.004	0.002	0.002	0.043	0.044	0.04
	(12.96)**	(5.93)**	(5.79)**	(10.24)**	(4.44)**	(4.78)**	(12.19)**	(8.36)**	(8.02)**
ln_area		0.084	0.022		0.019	0.025		-0.017	-0.32
		(6.37)**	-0.83		(4.95)**	(2.85)**		-0.56	(5.14)**
ln_perimeter			0.125			-0.011			0.619
			(3.11)**			-0.95			(6.53)**
N	6,184	6,184	6,184	6,184	6,184	6,184	6,184	6,184	6,184

* $p<0.05$; ** $p<0.01$

4 - Conclusion

The results confirm the potentiality of the use of indirect inference through instrumental variables in the analysis of the spatial urban structure. The topographical aptitude to urban uses can be considered as exogenous variable related to the spatial configuration of the urban structure. The standard deviation of the elevation used as an alternative to the ruggedness index presented a good performance in the first-stage estimation of the 2SLS model. The segregation indices showed a significant association to the urban form configuration as expected indicating that sprawled areas impact positively socio-spatial segregation. Besides, this association gets stronger near the inner limits of the urban spot. The methodology proposed is based on the use of geotechnologies such as remote sensing and geographical information systems. All the datasets used in the experiment were obtained from providers with open and free data access allowing the replication in different urban and geographical contexts.

The outcome of the models developed in this work indicates that there is a relationship between urban sprawl and socio-spatial segregation in the metropolis of São Paulo. However, the differences found in the results for distinct population group inform that the occupation and formation of urban areas occurs by specific processes for each group. The vast peripheries where the poor and less instructed population group dwells started its formation in the 50's by a process of centrifugal expansion that holds close relation with a strong monocentric structural organization. Historically, did not exist enough concentration of capital to generate an endogenous local development in these parcels of territory (Langenbuch, 1971, p.337). If this primordial period can be characterized as "suburban peripherical belt" formed as an expansion area of the future metropolis, today we see the consolidation of a heterogeneous space, densely occupied in most of the cases by segregated population originated from the tolerated illegality by the state or from its direct intervention in extensive social housing complexes (Rolnik, 1997).

The results reinforces the perspective introduced by Pedro Abramo (2009, p.63) describing the operation of the informal market, specifically what he calls the submarket of informal land allotment. For him, there is a tendency of extensification of the urban land uses derived from a spatial strategy of informal developers seeking for cheap and undeveloped land at the outskirts of the city. Our results indicate that it is exactly in these areas, near the urban fringe, where we found the strongest relationship between sprawl and segregation. However, if we look at the results for the subset with more central census tracts, dispersed areas can also be associated with segregation of low educated and low income population group. This finding goes in the opposite direction to what the author defines as compaction tendency of the informal territories in the inner parts of the urban structure (Abramo, 2009, p.66). Our results demonstrate that even considering the inner part of the city, the association of segregation of the poor and low educated group and dispersed urban configuration persists. Contrariwise, observing the results for the isolation of rich and highly educated households there is an association between less disperse, more compacted territories.

The methodology proposed aims to contribute to the debate of the consequences of housing policies and the role of the formal and informal market in the process of structuring the urban space and its impacts on the promotion of the extension of the urban area and the spatial differentiation according to population groups. From the vast territories occupied by informal allotments in the peripheries of the city, passing by the extensive housing complexes built in the 1970's and the 1980's, to the actual boom in the real estate sector motivated by the expansion of the credit, not much is said about the interferences of this developments on the urban structure, prevailing the limited debate about the housing deficit. Through the possibility to establish direct causality between urban form originated by the individual and collective decision and the resultant social fabric, we break with the myopic perspective that understand the urban space as a sole basis for the overlap of sector policies in layers of absence/presence of infrastructures and services.

5 -References

Abramo, P. **A cidade caleidoscópica: coordenação espacial e convenção urbana**. Rio de Janeiro: Bertrand Brasil, 2007. 368p.

Abramo, P. A cidade informal COM-FUSA: mercado informal em favelas e a produção da estrutura urbana nas grandes metrópoles latino-americanas. In: ABRAMO, P. **Favela e mercado informal: a nova porta de entrada dos pobres nas cidades brasielrias.** Porto Alegre: Antac, 2009.336 p.

Alonso, W. **Location and Land Use: Toward a general theory of land rent.** Cambridge: Harvard University Press, 1964. 204p.

Angel, S.; Parent, J.; Civco, D.L.; Blei, A.M. **Making Room for a Planet of Cities.** Lincoln Institute of Land Policy Focus Report, 2011. Available at: <http://www.lincolninst.edu/pubs/2094_Planet-of-Cities>

Angrist, J.D.; Pischke, J. **Mostly Harmless Econometrics**. Princeton: Princeton University Press. 2009. 373p.

Archer, R.W. Land Speculation and Scattered Development; Failures in the Urban Fringe Land Market. **Urban Studies**, n.10, 1973, p.367-372.

Bahl, R. W. A land speculation model: the role of the property tax as a constraint to urban sprawl. **Journal of Regional Science**, v.8, n.2, 1968. p.199-208.

Baun-Snow, N. Did highways cause suburbanization? **The Quarterly Journal of Economics**, May: 2007, p.775 – 801.
Burchfield, M.; Overman, H.G.; Puga, D.; Turner, M.A. Causes of Sprawl: A Portrait From Space. **The Quarterly Journal of Economics**, May: 2006, p.587-633.
Clawson, M. Urban Sprawl and Speculation in Suburban Land. **Land Economics**. v.38, n.2, 1962, p.99-111
Duncan, O.D.; Duncan, B. Residential Distribution and Occupational Stratification **American Journal of Sociology**, v.60, n.5, World Urbanism. March:1955, p.493-503.
Feitosa, F.; Câmara, G.; Monterio, A.M.V.; Koschitzki, T.; Silva, M.P.S. Global and local indicators of urban segregation. **International Journal of Geographical Information Science.** v.21, n.3, 2007, p.299-323.
Fujita, M.; Krugman, P.; Venables, A.J. **The spatial economy: cities, regions, and international trade**. Cambridge: MIT Press, 1999. 367p.
Fujita, M.; Ogawa, H. Multiple equilibria and structural transition of non-monocentric urban configurations. **Regional Science and Urban Economics**. n12. 1982. 161-196p.
Galster, G.; Hanson, R.; Ratcliffe, M.; Wolman, H.; Coleman, S.; Freihage, J. Wrestling Sprawl to the Ground: Defining and measuring an elusive concept. **Housing Policy Debate**, v.12, n.4, 2001, p. 681-717
Glaeser, E.; Khan, M.E.; Rappaport, J. **Why Do the Poor Live in Cities?** Harvard Institute of Economic Research: Discussion Paper. n1891. 2000.
Hall, P.G. **Cities of Tomorrow**. 3ªed. Malden: Blackwell publishing. 2002. 553p.
Harvey, R.O.; Clark, W.A.V. The Nature and Economics of Urban Sprawl .**Land Economics,** v.41, n.1, Feb:1965, p. 1-9
Jargowsky, P.A.; Take the Money and Run: Economic Segregation in U.S. Metropolitan Areas **American Sociological Review**, v.61, n.6, Dec: 1996, p. 984-998
Krugman, P. Space: The Final Frontier **The Journal of Economic Perspectives**, v.12, n.2 Spring, 1998, p.161-174
Lagenbuch, J.R. **A estruturação da Grande São Paulo: estudo de geografia urbana**. Rio de Janeiro: IBGE, 1971. 354p.
Leroy, S.; Sonstelie, J. Paradise Lost and Regained: Transportation Innovation, Income, and Residential Location. **Journal of Urban Economics**. n.13, 1983, p.67-89
Marques, E.; Torres, H. **São Paulo: segregação, pobreza e desigualdades sociais.** São Paulo: Editora SENAC, 2004.
Massey, D.S.; Denton, N.A. Trends in the residential segregation of Blacks, Hispanics, and Asians: 1970-1980. **American Sociological Review**, v.52, n.6, 1987, p.802-824.
Mills, E.; Hamilton, B. **Urban Economics**. 5th ed. Upper Saddle River: Harper Collins College Publishers. 1993. 480p. .
Muth, R. F. **Urban Economic Problems**. New York: Harper and Row, Publishers, Inc. 1975.402p.
Rolnik, R. **A cidade e a lei: legislação, política urbana e territórios em São Paulo.** São Paulo: Studio Nobel: FAPESP, 1997.
Richardson, H.W. **Economia Urbana**. Rio de Janeiro: Interciência, 1978. 194p.
Riley, S. J.; DeGloria, S.; Elliot, R. A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity. **Intermountain Journal of Sciences**, v.5, 1999, p.23–27.
Sabatini, F.; Salcedo, R. Gated communities and the poor in Santiago, Chile: functional and symbolic integration in a context of aggressive capitalist colonization of lower-class areas. **Housing Policy Debate,** v.18, n.3, 2007. p. 577-606
Schelling, T.C. Models of Segregation **The American Economic Review**, v.59, n.2, Papers and Proceedings of the Eighty-first Annual Meeting of the American Economic Association. 1969. p. 488-493

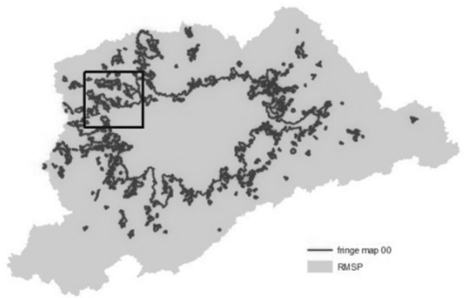
Smolka, M.; Biderman, C. Housing Informality: An Economist's Perspective on Urban Planning. In. **The Oxford Handbook of Urban Economics and Planning**. Ed: Brooks, N; Donaghy, K; Knaap, J.Y. Oxford University Press, Oxford, 2012. 814-833p.

Torres, H.G.; Marques, E.C.; Ferreira, M.P.; Bitar, S. **Poverty and space: pattern of segregation in São Paulo**. Paper presented at Workshop on Spatial Segregation and Urban Inequality in Latin America, Austin, Texas, 15-16 Nov 2002

Villaça, F. **Espaço Intra-Urbano no Brasil**. São Paulo: Studio Nobel, 1998. 373p.

Whyte, H. W. Urban Sprawl (1958) In. **The Essential William H. Whyte**. Ed. Albert LaFarge. New York: Fordham University, 2000. 383p.

APPENDICE



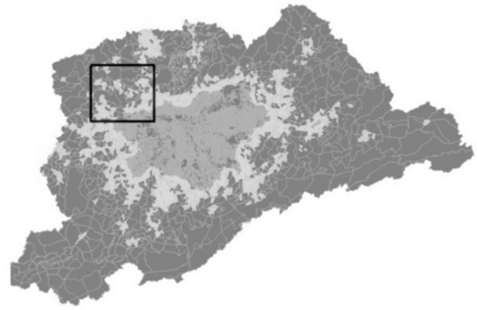
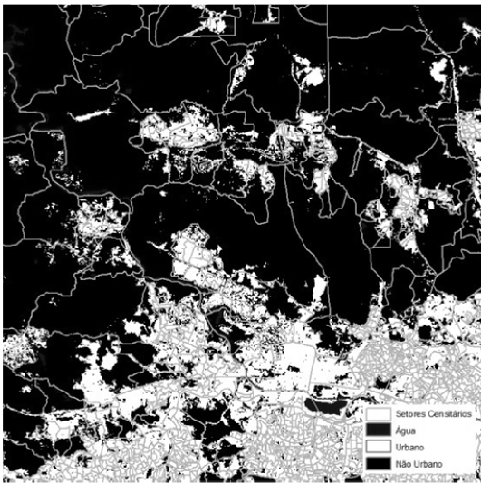


Image (A) **Image (B)**



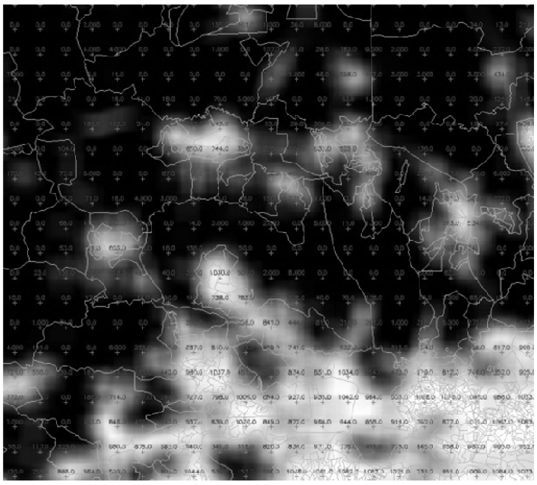
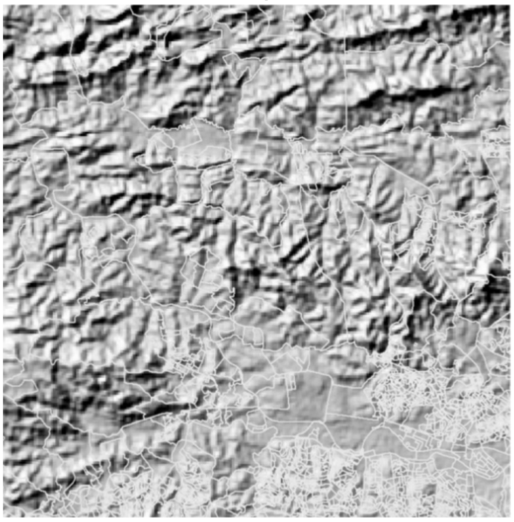


Image (C) **Image (D)**



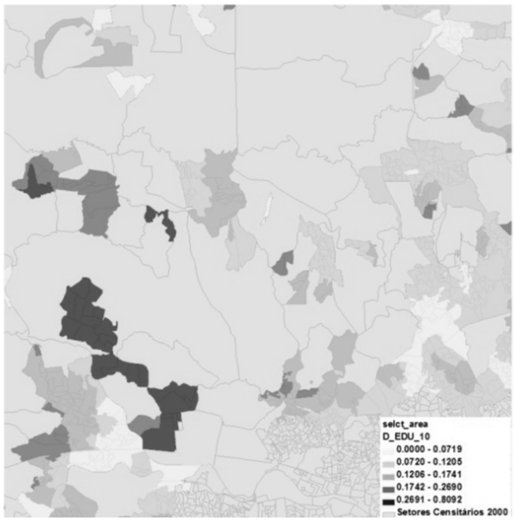


Image (E) **Image (F)**

Image (A) displays the urban limits of MRSP in 2000 used in the selection of census tracts on the edge/border. Image (B) displays the sectors selected for the 1Km inbound sample. Image (C) shows a detail of the image Landsat classified by landcover and overlayed by the tracts. Image (D) shows the resulting grid of the neighborhood operation for the calculations of s_i. Image (E) displays a detail of the SRTM data bank used for the surface ruggedness index. Image (F) shows a cartogram with the dissimilarity index for the selected sectors in detail.