Spatial distribution of sanitation and income inequality in Brazilian Slums

Vanessa Nadalin – IPEA - Institute for Applied Economic Research
Lucas Mation IPEA - Institute for Applied Economic Research

Preliminary and incomplete – please contact authors before citing it
To be presented at the Regional Studies Association Global Conference 2014

1. Introduction

During the second half of the XIX century, Brazil experienced rapid urbanization that posed the challenge of providing good quality housing and urban infrastructure to the fast growing population in big cities. This challenge rarely have been tackled properly, and the country reached the XX century with serious problems of slums: housing with poor urban services and some sort of informal land tenure.

To help to study the issue, data is needed. The main source used for the detection of the phenomena of slums is the data available from population censuses, which are not comparable over time. The statistics office (IBGE) alerts that changes in classification procedures improved considerably the identification of slums in the 2010 Census compared to previous census. This article overcomes that obstacle, using a methodology that enables the comparison of census data on slums living conditions between 2000 and 2010. The method is based on the possibility to compare the census tracts between 2000 and 2010, creating "minimum comparable areas" of census tracts. We then explore the information of the slum areas of the census 2010 to reclassify the slum areas in the 2000 census.

The first goal of the article is to use census data on water provision and sewage collection for the two Census years to test if there was a positive change in the spatial distribution of sanitation in slums. We evaluate specifically the local spatial disparity of neighboring slum and regular areas. We look for the existence of more local spatial auto-correlation (LISA) clusters of low values in slums and if the amount of low value clusters decreased from 2000 to 2010. There have been attempts of the government to extend sanitation for as much dwellings as possible in the country and considerable improvement has been achieved. Still it is important to check if this improvement has reached slums areas.

Then, as a second goal, we analyze income inequality Gini coefficient inside each "minimum comparable area" to evaluate if slums are more or less internally unequal than regular areas. We also analyze if this internal income inequality increased on average, from 2000 to 2010. This is done using data aggregated by income range, neglecting the inequality inside each income range. We use a calculation of Gini coefficient for this situation proposed by Kakwani (1980). To our knowledge, it is the

first time Gini coefficient is calculated for such small areas as census tracts and inside slums for Brazil.

2. Slum identification in Brazilian Census

The correct detection of the phenomena of slums is critical to the formulation of a number of public policies in different spheres of government. However, identifying these areas requires very detailed studies of each local situation, which are costly. This is due to the fact that there small and large slums that may be organized in different types of urban conformations.

The population census of the IBGE have great potential to detect these phenomena, because of its reach, coverage of several topics, rigor in the collection process and possibility of detailed geographical breakdown of the data. Moreover, its high cost is justified by the various uses of the information. However, the measurement of slums in a nationwide census operation is not a trivial task. There is the difficulty of defining a precise concept that applies to the entire territory, as well as the difficulty for its operationalization.

In the 1950s an IBGE work on living conditions of Rio de Janeiro slums used for the first time the concept of subnormal clusters (IBGE, 1953). The concept remained essentially the same since then, despite small changes in terminology. According to IBGE (2011), subnormal clusters are census tracts in with some of the following characteristics:

- At least 51 households;
- Occupation of lands without formal titles (even if regularized in the last 10 years), including invasions and irregular and illegal settlements;
- irregular urbanization, narrow lanes of irregular alignment, uneven lots, buildings outside urban standards;
- Precarious public services such as water, sewer or garbage collection;
- Inadequate housing topography due to the steep slopes, or the propensity to flooding.

According to the official census data, there has been a considerable evolution of the population in subnormal clusters over the last thirty years, from 2.2 million in 1980 to 11.4 million in 2010. However, the IBGE warns that these figures are not comparable due to changes in the information used for classifying census tracts as subnormal.

Most studies on the subject indicate that there was an underestimation of slum areas in the 2000 Census. Taschner (2001) and Costa and Nascimento (2005) emphasize the minimum limit of 51 households in a subnormal sector as a key factor for such underestimation. Marques et. al (2007) identify "precarious census tracts" with socioeconomic characteristics of residents similar to those of the subnormal tracts. They

find that this population is as big as the population in census tracts classified as subnormal, indicating how the true population in slums could be underestimated.

The classification of sectors as subnormal presents difficulties of two kinds. On one side are the problems related to the generality of the concept that captures poor urban and housing conditions. On the other side are the problems of operationalization of this classification for the entire national territory, since this process is done in a decentralized manner at the local units of the IBGE.

One of the conceptual problems is that, with the exception of formal land titles, the other criteria have no clear threshold. What counts is the amount of certain characteristics in a certain area. This flexibility in the definition is necessary to accommodate the differences in slum phenomenon across the country. It is the local IBGE team that classifies the areas. The accuracy of this classification depends on the correct appropriation of the concept of subnormal clusters by these local teams in the preparatory phase of each census, which may involve thousands of IBGE employees.

Operational difficulties arise from the fact that slums can be small and that local knowledge is required, including field work, to identify them. Cardoso (2013) points out that because of the difficulties of using urban and housing characteristics to distinguish subnormal clusters, "land ownership has been, officially, the criterion that defines the classification of census tracts as subnormal clusters or slums." However, until the 2010 Census rarely information concerning formal land titles was available to IBGE teams.

This ability to identify slums and similar settlements has been improved in the 2010 Census. The delimitation of census tracts began to be made with the aid of digital road maps and the use of high resolution satellite images. These were complemented with field visits to observe the morphological patterns of urban areas. Meetings with municipalities and local stakeholders were performed as well as a survey of the formal land titles situation in each area.

A direct comparison between the 2000 and 2010 censuses indicates the degree of incompatibility of subnormal clusters data. The population residing in slums would have grown 75%, from 6.5 million in 2000 to 11.4 million in 2010. That is, residents of slums would have grown from 3.8% of the national population in 2000 to 6% in 2010. Such an intense increase in slums would be incompatible with the improvement of the socioeconomic indicators, poverty reduction and inequality recorded in Brazil during the last decade.[1]

3. Data

The questions of population censuses are divided into two blocks. The basic questionnaire has a few questions about age, gender, income, literacy, some

[1] See IPEA (2010) and Furtado (2013).

characteristics of the home and is applied to all surveyed individuals. For reasons of confidentiality, this questionnaire information is disclosed aggregated by census tract. The sample questionnaire consists of more detailed and comprehensive questions and is applied only to individuals selected in the sample. The information of this questionnaire is disclosed without aggregation but their spatial identification is less precise, at the level of groups of census tracts.

Because the census tracts are the ones classified as subnormal or regular, this is the unit of analysis used in this work. The data comes from the IBGE publications on information aggregated by census tract for the years 2000 and 2010 (IBGE, 2002, 2012). It was necessary to make some adjustments to make census tracts of both years comparable, as well as the slum classification and the income data. These procedures are described below.

a. Comparability of census tracts

The first and crucial adjustment was to find the same space partition in both periods. For that we use the Minimum Comparable Areas (MCAs) for census tracts created by Mation (2013). These are the smallest possible areas formed by aggregations of census tracts whose outer perimeter is common in all periods of time.

The IBGE changes the layout of census tracts over time to accommodate the growth in the number of households and changes in urban fabric, since each sector should be between 250 and 350 households in urban areas and their perimeters should be easily identified by enumerators.

These changes are recorded in a table filled at the moment of redefining the layouts of census tracts. Between 2000 and 2010, 45% of the sectors from the 2000 Census kept their layout in 2010. In the other sectors changes were mostly simple divisions, such as a census tract of 2000 that was divided in two in 2010. In only 3% of cases there were more complex divisions.

In cases of simple divisions, the 2010 census tracts were aggregated to restore the original 2000 ones. But due to the possibility of more complex divisions, Mation (2013) proposed an automatic method using the recordings of the changes and graph theory. For the period between 2000 and 2010, 212,164 MCAs were created, from the original 215 812 census tracts of 2000 and the 316,574 census tracts of 2010.

Because subnormal areas take place only in urban areas, MCAs in which all census tracts were empty or rural in both periods were excluded. After these exclusions, we had 164,190 urban MCAs that exactly matched the layout of census tracts in 2000 in 99.64% of cases. These are the ones used in this work.

The IBGE original digital census tracts map of 2010 was aggregated to form a map of the MCAs. The 2000 Census included in the census tracts layouts empty urban areas

such as parks and lakes adjacent to occupied areas. These areas were separated in empty areas in the 2010 census, and were excluded from the MCAs map.

b. Comparability of slum classification

With the improvement in the identification of slums in 2010, many more census tracts were classified as subnormal. Since the characteristics of subnormality have low probability of change between the two census periods and expansion of slums occurs in areas not previously occupied, the information from the 2010 Census were used to reclassify the 2000 census tracts.

In cases where there is no change in the official classification between the two periods, it is maintained. In cases where subnormal areas in 2000 change classification to normal in 2010 the original classification is also maintained. They correspond only to 15% of the slum population in 2000, and might correspond to processes of slum upgrading.

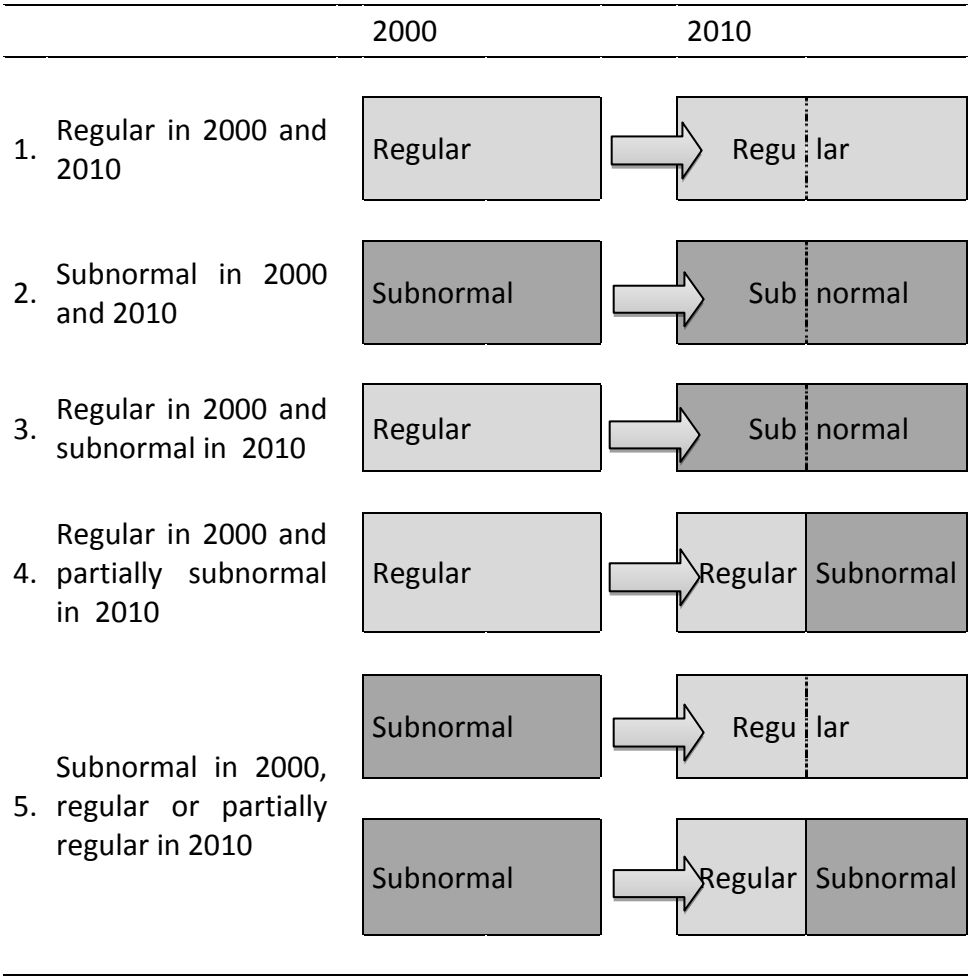
The cases that demand more attention are those in which normal census tracts of 2000 became subnormal in 2010. Firstly, as already noted, the concept of subnormality includes irregularity of the urban fabric, and lack of formal land titles. It seems reasonable to assume that hardly a land occupied regularly would be converted to a subnormal area.

Slums usually emerge in areas without previous urban occupation, whether in census tracts entirely empty or partially empty in 2000. Accordingly, there is strong evidence of misclassification in 2000 if an occupied census tract classified as normal in 2000 changed its classification to subnormal in 2010. If there has not been subdivision from 2000 to 2010, or if all the 2010 sectors are subnormal, the corresponding census tract of 2000 was reclassified as subnormal.

Apart from that case, a more complex case occurs, in which a regular census tract of 2000 has been subdivided into more than one units in 2010 with at least one of them classified as subnormal and one of them classified as regular. This may have happened for some reasons. The subnormal census tract of 2010 may have filled an empty space of the 2000 census tract. There may have been a small slum, with less than 51 households, embedded in the regular census tract, that during the decade grew to more than 51 households. Finally, the slum may have existed, with more than 51 households but have not been classified as subnormal, an error corrected in the 2010 classification.

Figure 1 depicts these possible types of change in the classification of a census tract from 2000 to 2010, considering the possibility of subdivision in 2010. This possibility is indicated by the dotted lines of the blocks to the right in the figure.

Figure 1: Minimum Comparable Areas according to the original classification of census tracts as regular or subnormal in the years 2000 and 2010

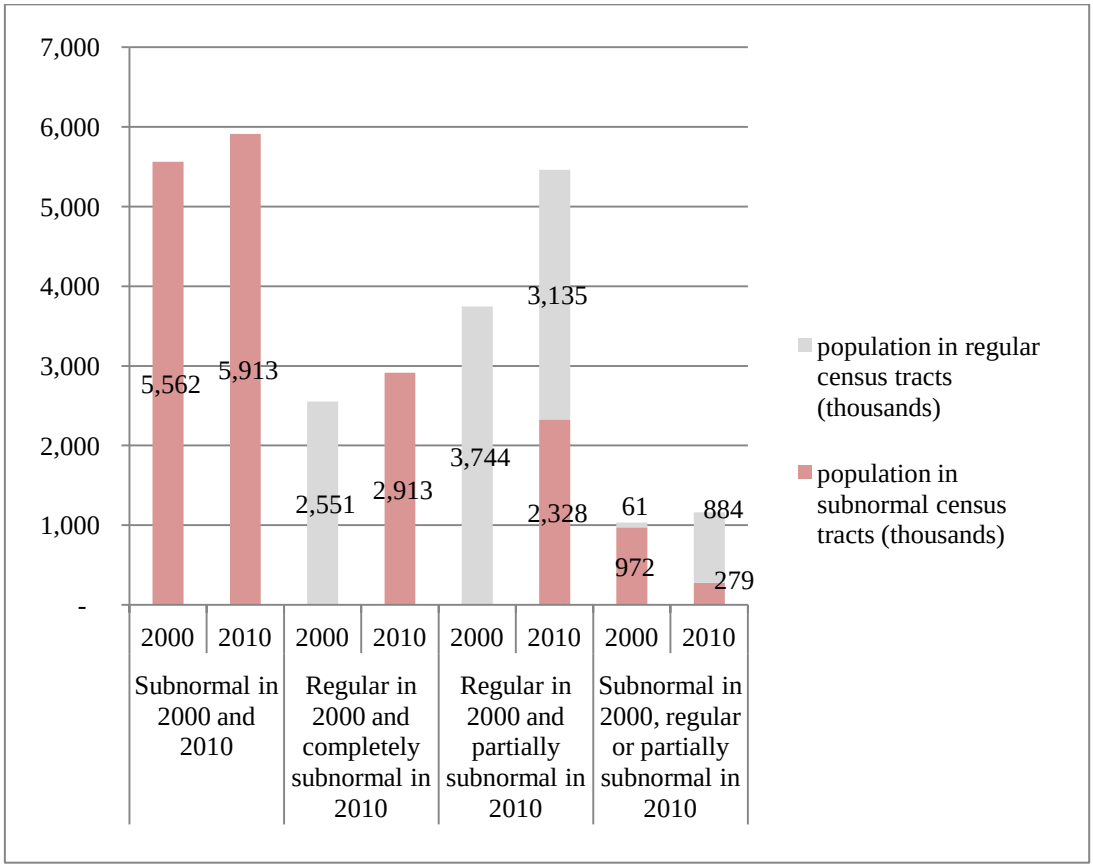


Note: dotted lines indicate the possibility of subdivision of the original 2000 census tract

In this paper we only consider as slums the MCAs composed entirely of subnormal sectors in 2010. These comprise the originally classified as subnormal in 2010 and 2000 as well as the originally classified as subnormal in 2010 but totally reclassified as subnormal in 2000.

Therefore, we will not analise the MCAs composed of both the subnormal and regular census tracts in 2010. Figure 2 depicts the size of this group not analyzed in relation to the others. Its slum population corresponds to 20.4% of the 11.4 million people living in slums in 2010.

Figure 2: Brazilian population in regular and subnormal areas in Minimum comparable areas with some subnormal census tract (2000 and 2010)



c. Comparability of income data

The way census data was collected changed from 2000 to 2010. In 2010, IBGE started to use portable electronic device while in 2000 they still used paper. There is an issue with the income variable. It is quite usual that the surveyed person refuses to answer this question. In paper, the enumerator leaved the field blank. But in 2010 the portable electronic device did not allow the option to leave blank. Some of the enumerators filled the question with a zero. Therefore, IBGE assumes that the number of zero income households includes some of the people that failed to declare their income.[2]

To adjust for this difference and be able to compare income data from 2000 Census with the 2010 Census we impute income for some of the zero income household. Some people has positive income but chose to lye and declared having zero income. These cases also occurs in the 2000 Census, so we perform the same imputation in the 2010 and 2000 Census.

[2] See IBGE 2012: ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_Gerais_da_Amostra/Estudo_e_tratamento_rendimentos.pdf

The procedure consists of two stages, one in the sample data and the other in the census tract data. The first stage is when the imputation is done, with individual microdata. Chief of households with declared zero income are separated into "true" and "false" poor households. A cluster analysis divides zero income households into these two groups according to gender, education, living in urban areas and the home quality. Then the hot deck imputation was made taking the income from donor of the same class of the recipients. These classes also took into account socioeconomic variables to find suitable donors.

The second stage is to bring this imputed income in the sample data into the census tract data. This income data is aggregated into income ranges. We know the total number of households inside each income range and the sum of their incomes, for each income range. Sample data is representative of a group of census tracts. For this set of census tracts we found the households that ceased to have zero income and how many were added in each of the income ranges. Then we distributed these households into the separated census tracts according to how many zero income households they originally had. Next, we added the imputed household income to the original data on the sum of households incomes. The imputed income was the average imputed income in that income range for that group of census tracts in the sample microdata.

As a result, from the 7,3 million households with zero income in 2010, 1,4 million were left with zero income. In 2000, the original 4 million gave place to 1.4 million. Indeed, in 2010 there were more "false" poor households than in 2000, what confirms the IBGE issue with the portable electronic device.

4. Spatial Distribution of sanitation

In this section we analyze the evolution on the share of households with water provision and sewage collection. We calculate local spatial auto-correlation, LISA (Anselin, 1995) spatial clusters for all the MCAs in the whole Brazil, for each year. We presume that clusters of low value are more frequent in slums because the official sanitation network does not reach its territory.

Table 1 brings the results for slums (subnormal MCAs), their contiguous neighbors and Brazil. Overall, low value clusters (LL) decreased for both water and sewage. A good result is that for water it decreased more for slums, whereas a not so good result is that for sewage it decreased less. As slums corresponds to 5.4% of the MCAs we do not confirm that LL clusters are more frequent in slums. Instead, we see that LL clusters are more frequent in slums neighbors. For instance, 20,5% of sewage LL clusters in 2010 are slum neighbors, while this kind of MCAs corresponds to 13,3% of the total of MCAs.

In general both public services increased their provision in slums, their neighbors and the whole of Brazil. Slums do have a lower level of both services provision, yet this

level grew more on slums. The gap of provision is bigger in relation with slums neighbors than national average, and is getting narrower.

Table 1: Low value clusters of water provision and sewage colletion in Brazilian minimum comparable areas (2000-2010)

		water			sewage		
		2000	2010	growth	2000	2010	growth
Slums	% of households with provision	85.9%	89.2%	3.4%	60.2%	70.6%	10.4%
	LL clusters	823	522	-36.6%	690	553	-19.9%
	% of total LL clusters	6.2%	5.3%		3.7%	4.5%	
Slums contiguous neighbors	% of households with provision	90.1%	91.8%	1.7%	78.7%	81.5%	2.8%
	LL clusters	2690	1955	-27.3%	3298	2496	-24.3%
	% of total LL clusters	20.2%	19.8%		17.5%	20.5%	
Brazil	% of households with provision	87.9%	90.5%	2.6%	70.7%	74.2%	3.5%
	LL clusters	13339	9862	-26.1%	18890	12195	-35.4%
	% of LL clusters over total MCAs	8.1%	6.0%		11.5%	7.4%	

To test the significance of these means differences for slums and their neighbors we performed a difference-in-differences regression. Slums are the treated group, since they received specific policies to improve infrastructure and their neighbors the control group. The estimated equation is:

$$share\ of\ service\ provision_{it} = \beta_0 + \beta_1 slums_i + \beta_2 2010_i + \beta_3 slums_i * 2010_i$$

where "i" indicates the MCA, "t" the year, "slums" and "2010" are dummy variables. Table 2 shows the results which corroborates results in table 1, indicating that slums are worse off than their neighbors, but that they improved more than them.

This means analysis does not capture the spatial pattern issue. Although there was an improvement in all means and even better improvement in slums, table 1 shows that for sewage collection low value clusters decreased less for slums. That is to say that the spatial distribution of that improvement is unequal in space. Because sewage collection requires a network infrastructure it is reasonable that the improvements are concentrated in some parts of the territory. The network expands to some portion of the territory first.

Table 2: Results of differences in differences estimation for water provision and sewage collection

	water	sewage
slums	-0.028	-0.185
	(8.62)**	(40.68)**
2010	0.029	0.039
	(13.56)**	(13.89)**
slums*2010	0.018	0.07
	(4.18)**	(11.30)**
Constant	0.889	0.793
	(538.76)**	(375.15)**
Observations	61462	61462
R-squared	0.01	0.05

Robust t statistics in parentheses
* significant at 5%; ** significant at 1%

Brazilian Federal government between the years 2007 and 2010 invested in sanitation through its program of growth speeding, "programa de aceleração do crescimento", PAC. By 2010, this program finished improvements of US$212 million through the slum upgrading policy and US$ 580 million through the general sanitation policy[3]. Hence, at least some of this improvement might be due to this federal investment in sanitation, and the better improvement of slums because of the specific investment on slums.

5. Slums income inequality

The disclosure of income distribution data from grouped observations at the census tract level enables the calculation of income inequality for such small spatial units. We will assume that inequality of income within each income range is zero. Thus, we will observe the lower bound of the true Gini index. If the income ranges are too wide, Gini will be underestimated. We use this lower Gini bound formal definition from Kakwani (1980):

$$Gini\ lower\ bound = 1 - \sum_{t=1}^{T+1} f_t(q_t + q_{t+1})$$

Where T+1 is the number of income classes, f_t is the relative frequency of the t^{th} income class and q_t is the cumulative proportion of income received by the households inside

[3] Source: PAC report on http://www.planejamento.gov.br/secretarias/upload/Arquivos/noticias/pac/Pac_7_4.pdf

Accessed on 4/20/2014

each income class. It is derived from a geometric approach that assumes that the true Lorenz curve is the connection of the intervals limit values.[4]

The inspection of figures 3 and 4 tells us that slums have a greater concentration of households of lower income than their contiguous neighbors. The gini distribution is quite similar, but in general the inequality is lower for slums. The differences are clear, and validate the idea that slums are a parallel "society" inside the formal city.

Figure 3: Histogram of the number of households by income ranges (2000, 2010)

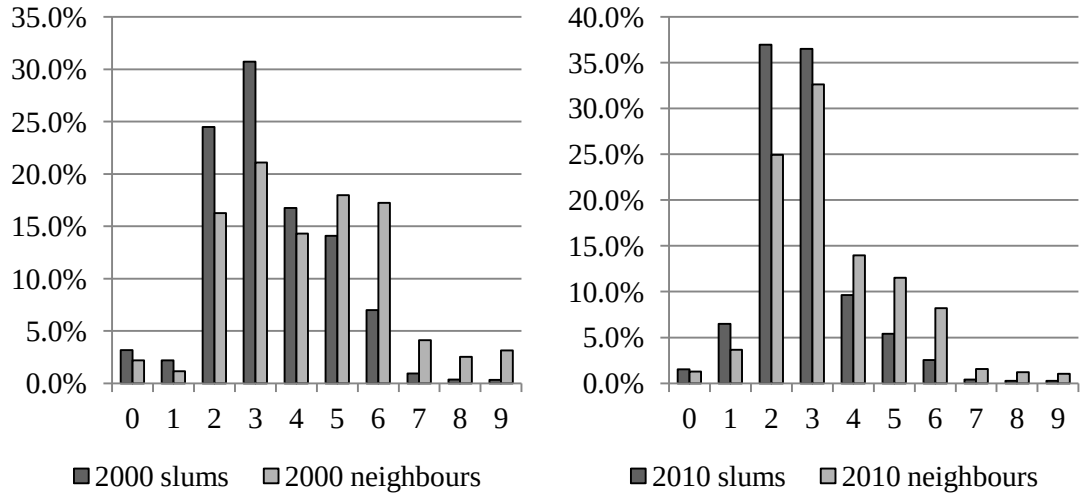
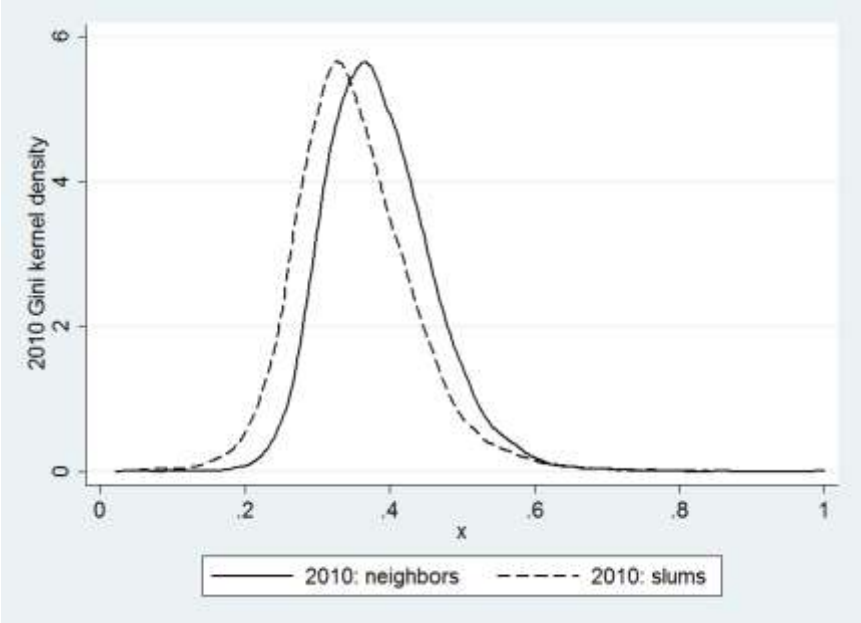


Figure 4: Kernel density estimation of the distribution of 2010 Gini coefficient



[4] There are 9 positive income classes defined with respect to the minimum wage (mw). Up to 0,5 mw, from 0,5 to 1 mw; from 1 to 2 mw; from 2 to 3 mw; from 3 to 5 mw; from 5 to 10 mw; from 10 to 15 mw; from 15 to 20 mw and more than 20 mw. In 2000 census, the minimum wage was 151 brazillian reais of 2000, or 286 of 2010, while the minimum wage in 2010 was 510 brazilian reais of 2010.

Table 4 brings the gini average for MCAs, how it changed in the decade and their medium income. We can see that there was an overall fall in inequality and this fall was larger in slums than in its neighbors. Moreover, there was an overall increase in average income, and that increase was greater for slums.

Table 4: Gini coefficient and average income descriptive statistics

		2000	2010	delta/growth
Slums	Gini mean (x100)	37.30	35.20	-2.10
	Gini Std. Dev. (x100)	6.60	8.21	
	Average income	767.85	898.19	17.0%
Slums contiguous neighbors	Gini mean (x100)	40.56	38.62	-1.94
	Gini Std. Dev. (x100)	7.12	7.64	
	Average income	1469.24	1519.03	3.4%
Brazil	Gini mean (x100)	41.60	39.16	-2.44
	Gini Std. Dev. (x100)	8.23	8.57	
	Average income	1510.76	1530.86	1.3%

The differences between these groups in these statistics are statistically significant by an ordinary t test. This tells us that slums are more homogeneous than the national average, and that they are considerably more homogeneous than their neighbors. Because inequality has fallen and income risen, all slums households must have increased their income, and lower income households increased even more. This is compatible with social policies for income distribution of Brazilian government aiming the poorer families. Still, if we see that immediate neighbors houses poor people at the same time as wealthier people, this can be viewed as an evidence that slums are poverty traps. One of the most classic critics to minimum income social policies is the lack of promotion of social mobility, and this spatial comparison with slum neighbors may be telling us this story.

6. Concluding remarks

This article presented how sanitation and income distribution changed from 2000 to 2010 in Brazilian slums and their contiguous neighbors. It describes a great effort on assembling the data set, to make spatial units, slums classification and income data comparable. The construction of that panel data set is a considerable contribution and it can be used for far more analysis than the ones made here.

With respect to the goals set for this work, the analysis on the local spatial disparity of sanitation indicated that there was an overall improvement, and that improvement was greater in slums. In general 2010 presented less low value clusters than 2000, indicating a better spatial distribution of infrastructure. Still, for slums that spatial distribution on sewage collection was not as widespread as in regular areas. That indicates that

improvements are concentrated in space, because infrastructure networks are constructed in some portions of the territory to provide the whole territory.

With respect to income inequality, slums are less unequal than the national average and less unequal then their neighbors. Yet, their average income is lower, although their income has increased more. We suspect this is a consequence of minimum income policies, since Gini index decreased more for slums. Comparing with slums neighbors, which present wealthier household as well as poorer ones, we suspect this might be a signal that slums are poverty traps.

Finally, as next steps for this study, we recognize the importance of repeating the analysis disaggregating for regions, especially metropolitan regions. On the issue of income inequality we intend on furthering the investigation on within neighborhoods inequalities and between neighborhoods inequality, focusing on spatial income distribution and slums.

7. References

ANSELIN, L., Local Indicators of Spatial Association - LISA Geographical Analysis 27 (2), 1995.

CARDOSO, A. L. Assentamentos precários no Brasil: discutindo conceitos. In: Caracterização e Tipologia dos Assentamentos Precários Brasileiros – Estudos de Caso. [s.l.] IPEA, 2013. .

COSTA, V. G.; NASCIMENTO, J. O conceito de favelas e assemelhados sob o olhar do IBGE, das prefeituras do Brasil e da ONU. ENCONTRO DE GEÓGRAFOS DA AMÉRICA LATINA, X, 2005.

FURTADO, B. A. Índice de Vulnerabilidade das Famílias (2000-2010): Resultados. [s.l: s.n.]. Disponível em: <http://ideas.repec.org/p/ipe/ipetds/1835.html>. Acesso em: 18 jun. 2013.

IBGE. As favelas do Distrito Federal e o Censo Demográfico de 1950. [s.l.] IBGE, 1953.

___. Instruçõeses para delimitação dos setores censitários, 1970.

___. Censo Demográfico 1980: resultados do universo agregados por setor censitário, 1980a.

___. Instruções para delimitação dos setores censitários, 1980b.

___. Censo Demográfico 1991: resultados do universo agregados por setor censitário, 1991a.

___. Manual de Delmitacao de Setores e Zonas de trabalho de 1990, 1991b.

___. Manual de delimitacao dos setores, censo demografico 2000, 2000.

___. Censo Demográfico 2000: resultados do universo agregados por setor censitário, 2002.

___. Censo Demográfico 2010 – Aglomerados Subnormais – Primeiro ResultadosIBGE, , 2011.

___. Censo Demográfico 2010: resultados do universo agregados por setor censitário, 2012.

___. Aglomerados subnormais: Informações Territoriais, 2013.
IPEA. Objetivos de Desenvolvimento do Milênio – Relatório nacional de acompanhamento. n. Brasília: Ipea, 2010.
KAKWANI, N. Income Inequality and poverty Oxford University Press, International Bank for Reconstruction and Development, 1980.
MARQUES, E. et al. Assentamentos precários no Brasil urbano. Brasília: Ministério das Cidades, 2007.
MATION, L. Criação de Áreas Mínimas Comparáveis (AMCs) para Setores Censitários de 1980 a 2010texto para discusão - IPEA (no prelo), , 2013.
TASCHNER, S. P. Favelas em São Paulo – censos, consensos e contra-sensos. Cadernos Metrópole, n. 05, p. 09–27, 2001.