

# Small Area Housing Deficit Estimation: A Spatial Microsimulation Approach

Flávia da Fonseca Feitosa<sup>1</sup>, Roberta Guerra Rosemback<sup>2</sup>, Thiago Correa Jacovine<sup>1</sup>

<sup>1</sup> Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas - Universidade Federal do ABC (CECS/UFABC) - São Bernardo do Campo, SP - Brazil

<sup>2</sup> Centro de Desenvolvimento e Planejamento Regional - Universidade Federal de Minas Gerais (CEDEPLAR/UFMG) - Belo Horizonte, MG - Brazil.

flavia.feitosa@ufabc.edu.br, rosemback@cedeplar.ufmg.br,  
thiago.jacovine@aluno.ufabc.edu.br

***Abstract.** This paper presents our first attempts to develop a new methodology for measuring housing deficit at small areas. It combines the advantages of two types of census data: (a) individual-level sample data, which are very useful for depicting many dimensions of the housing deficit, but do not present detailed geographic information; and (b) universal data with detailed spatial resolution (census tracts), but aggregated. For that, we explore an approach based on spatial microsimulation. We simulate spatial microdata by using aggregate data as constraints to expand and allocate individual-level data to census tracts. This procedure allowed us to estimate a particular dimension of the housing deficit (housing cost) at higher spatial resolution.*

## 1. Introduction

The lack of decent housing for families with limited means represents a major social problem in Brazil and other developing countries. To support the development of appropriate housing policies, it is essential to characterize and quantify the multiple dimensions of housing deficit. According to the Brazilian National Housing Policy (PNH - *Política Nacional de Habitação*), federal funding towards municipal social housing projects will only be allocated to municipalities that have developed a Local Plan of Social Interest Housing (PHLIS - *Plano Local de Habitação de Interesse Social*), which includes a local diagnosis of the current housing situation.

The PNH assigns to the municipalities a role that was previously attributed to the federal and state levels. This leading role of municipalities in the development of housing diagnosis raises new methodological demands, which include a refinement of scale and a thorough discussion on the potential of available data sources (Rosemback et al. 2014).

Building a methodology for estimating housing deficit requires the definition of which data sources will be used and the identification of their limitations and potentialities. The census is the most comprehensive statistical survey carried out in Brazil. It collects data on the composition and characteristics of population, households, dwellings and their surroundings. It is available to all municipalities and is therefore a unique data source for understanding the Brazilian housing conditions.

The census survey relies on two types of questionnaires: a sample one, which is applied to a fraction of households (about 11% of the population), and a simplified one,

to the remaining households. As a result, the census provides different types of data for public use, two of which are of particular interest for this work: (a) individual-level sample data (microdata), which are very useful for depicting many dimensions of the housing deficit, but are not universal and, for confidentiality reasons, do not present detailed geographic information; and (b) universal data with detailed spatial resolution (small areas known as census tracts), but aggregate.

This paper presents our first attempts to develop a new methodology for measuring housing deficit at small areas. For that, an approach based on spatial microsimulation is explored to combine the advantages of the two types of census data (sample microdata and universal aggregate data). Spatial microsimulation, in this work, is understood as the process of generating spatial microdata by taking data at the individual level and using aggregated level constraints to allocate these individuals to small areas (Lovell, 2014).

In the next section, we introduce conceptual dimensions of the housing deficit and the two main measurement approaches adopted for Brazilian cities – *place-based* and *household-based*. Along the description of these two approaches, we point out the potential and shortcomings of census data for capturing the multiple dimensions of the housing deficit. Afterwards, a spatial microsimulation method named "iterative proportional fitting" (IPF) is presented as a valuable resource to address the limitations identified in the current housing deficit estimation approaches. To illustrate this point, an experiment is conducted with census data from a small region of the city of São Bernardo do Campo, Brazil. In this experiment, we estimate one particular dimension of the housing deficit (housing cost) at higher spatial resolution by using universal aggregate data as constraints to expand and allocate individual-level sample data to small zones (census tracts).

## 2. The Housing Deficit: Dimensions and Measurement Approaches

The development of a diagnostic that is suitable for supporting social housing policies demands a multidimensional view of the housing issue. According to Rosembach et al. (2014), it is possible to point out at least seven dimensions of adequacy that must be considered in the assessment of housing needs, as presented in Table 1.

**Table 1. Housing Needs: Dimensions of Housing Adequacy (Rosembach et al., 2014).**

<b>Dimension</b>	<b>Description</b>
1. Housing Cost	The household spending on housing should not severely compromise the total household income.
2. Physical Suitability of the Dwelling Unit	Dwellings should be made of materials that permanently ensure weather protection, the health, privacy, and security of their residents.
3. Dwelling Unit Suitability to the Household	The household density in a building should not be excessively high. Families should not cohabit for lack of choice.
4. Environmental Safety	Dwellings should not be located in areas of environmental risks, including risks of flooding or landslides, contaminated areas, etc.
5. Legal Security	Households must have legal security of tenure.
6. Infrastructure and Public Services	Dwellings should be served by sewage, water supply, electricity network, street lighting, paving, trees, curb, sidewalk, etc.
7. Location and Accessibility	The location of dwellings should promote the integration into the city, including appropriate access to employment options, efficient public transportation, health services, school, culture and leisure;

Building a methodology that is able to measure these multiple dimensions of the housing deficit remains as an important challenge to be faced. In Brazil, it is possible to identify two different measurement approaches to address this issue. The first is a *place-based* approach that relies on the identification of so-called 'squatter settlements', i.e., inadequate human settlements occupied by low-income residents. The second approach, which is *household-based*, adopts the concept of 'housing needs' and addresses more explicitly some of the dimensions shown in Table 1.

In the *place-based approach*, the housing deficit is estimated from the count of the families living in areas demarcated as squatter settlements. It demands, therefore, data with detailed spatial information. These diagnostics often use local data provided by the municipalities and/or census data. In the latter case, the census data used is the one obtained from the simplified questionnaire and aggregated by census tracts, which is the smallest spatial unit of analysis available for public use. Due to its high spatial resolution, this data can be more easily combined with auxiliary data on dimensions that are not covered by the census survey, such as maps of risk areas ('Environmental Safety') and municipal data on legal status of land tenure ('Legal Security').

However, the variables obtained from the simplified census questionnaire can only depict the dimension 'Infrastructure and Public Services'. Another disadvantage of this data comes from the fact that it is aggregated. In this case, it is possible to know how many dwellings without connection to water supply networks *or* without garbage collection service can be found within the area of a certain census tract. Nevertheless, it is not possible to know how many dwellings are not served by both services. Thus, the identification of squatter settlements and the use of aggregated data prove to be insufficient for the characterization and quantification of the housing deficit.

The *household-based approach* demands individual-level data. The João Pinheiro Foundation (FJP) adopted this approach to develop a methodology for estimating the Brazilian housing deficit that is considered as a reference among social housing experts. This methodology relies on sample data from the 2010 Census to quantify the housing deficit of all municipalities in the country (FJP, 2013). In fact, the data obtained from the long census questionnaire represents a valuable source of information for measuring 5 of the 7 dimensions presented in Table 1. Only the dimensions 'Environmental Safety' and 'Legal Security' cannot be captured by any of the variables available in this dataset.

This data is available as microdata, which allows us to consider the household as an analytical category and also provides a richer set of information about their living conditions. It includes many variables that are not available in the aggregate data and allow multiple combinations among them. The dimension 'Housing Cost', for instance, demands a combination between information on the household spending on housing and the total household income, which can only be obtained from individual-level data.

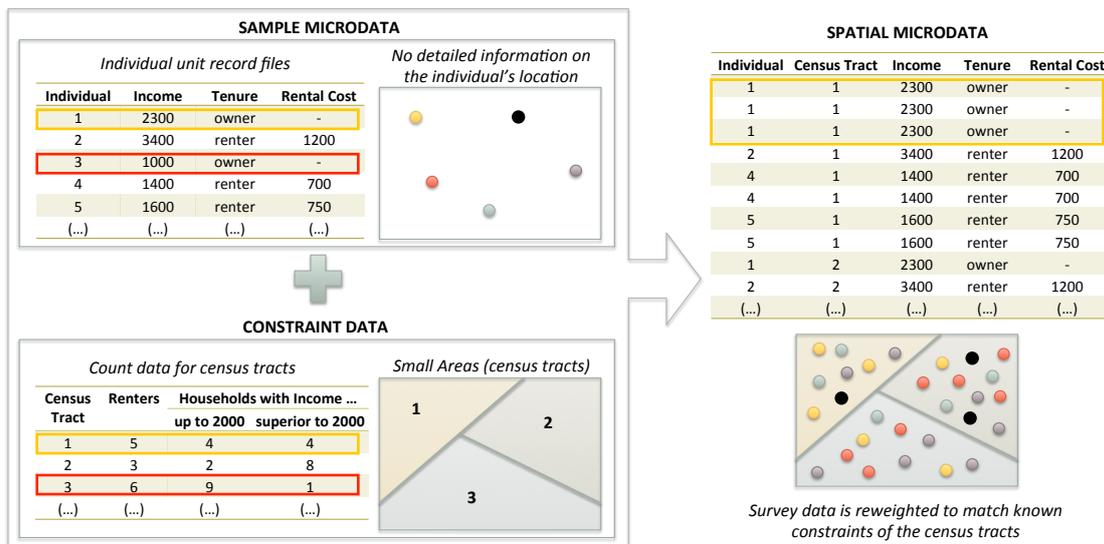
Nevertheless, census microdata does not provide detailed spatial information. The housing deficit estimates that are calculated from this dataset are only made available for the municipality as a whole or, at best, for weighting areas, which are large geographical partitions that are used in the census sampling weighting procedures. To incorporate the two dimensions that demand auxiliary data into these estimates ('Environmental Safety' and 'Legal Security'), it is essential to improve the spatial

resolution of these results. In other words, it is important to obtain more detailed information on the location of the sampled households.

### 3. Spatial Microsimulation for Housing Deficit Estimation: An Experiment

Since detailed spatial information about individual-level data cannot be provided due to confidentiality reasons, we advocate that spatial microsimulation represent a valuable resource to address the problem. To demonstrate that, this work uses spatial microsimulation to combine the advantages provided by the two types of census data presented in this paper: sample microdata (more variables and individual-level, but lower spatial resolution) and universal data (higher spatial resolution and easier integration with auxiliary data sources, but less variables and aggregated). This procedure introduces new possibilities for the development of hybrid approaches (individual and place-based) for measuring housing deficit.

In general, spatial microsimulation takes microdata at the individual level (e.g., sample microdata) and uses aggregate level constraints to allocate these individuals to small areas (e.g., data aggregated by census tracts) (Lovelace, 2014). In this work, we explore a spatial microsimulation method named "iterative proportional fitting" (IPF), which is based on deterministic reweighting. Basically, the weights are calculated and adjusted for each individual observation in every census tract until the known marginal distribution of the census tract population is matched by the weighted survey microdata (Hermes and Poulsen, 2012). Information about the marginal distribution is obtained from count data aggregated by census tracts (constraint data). The output is a "spatial microdata" – in other words, a dataset that contains a single row per individual and also an additional variable that indicates the small area (census tract) where the individual is located, as the Figure 1 shows.



**Figure 1. Creating spatial microdata using the IPF method. Adapted from Hermes and Poulsen, 2012.**

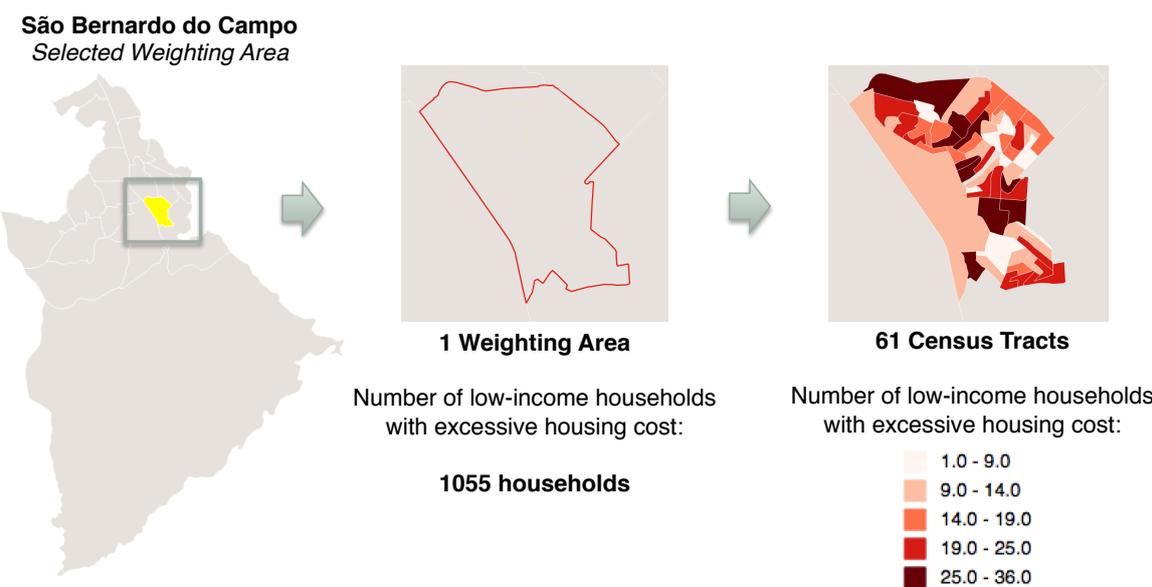
In the example presented in Figure 1, the results obtained from the IPF method show us that the individual 1 from the sample microdata received a weight equal to '3' for census tract 1. This means that this individual (1) represents three individuals of the census tract 1 (that is why individual 1 appears 3 times at the Spatial Microdata table).

Meanwhile, a weight equal to zero was assigned to the individual 3 because its characteristics are untypical for this small area. Thus, the individual 3 is not allocated to this census tract.

In this work, we present an experiment that uses the IPF method to estimate the spatial variability of one of the dimensions of housing adequacy presented in Table 1, the ‘housing cost’. This dimension is often represented by the ‘number of households with total income of up to 3 minimum wages that spend more than 30% of their income with rental costs’, which can be measured using microdata, but not using data aggregated by census tracts (higher spatial resolution). Nevertheless, we have per census tract data on the number of households with a certain tenure status (renters, owners, etc.) as well as the number of households with per capita income between certain boundaries. In the IPF method, these data can be used as aggregate constraints.

The experiment was conducted for a small portion of the city of São Bernardo do Campo, located in the Metropolitan Region of São Paulo, Brazil. Considering the sample data from the 2010 population census, it is estimated that there were 9811 low-income households with excessive rental cost in the city. Of this total, 1055 households are located in the weighting area that was chosen for the experiment (Figure 2).

The aim of the experiment is to estimate this particular dimension of the housing deficit at higher spatial resolution by using universal aggregate data as constraints to expand and allocate the individual-level sample data into the 61 census tracts contained in the selected weighting area. The procedure was conducted at the statistical and modeling software R (package 'ipfm' - Lovelace, 2014).



**Figure 2. Experiment results: Spatial distribution of low-income households with excessive rental cost, based on simulated data.**

To expand and allocate the individual-level sample data, we used the variables 'income per capita' (8 intervals of income) and 'tenure status' (6 classes) as *constraint variables*, since they are available in both datasets (individual-level and aggregate). As *target variables*, we selected variables that are only obtained from the sample microdata: "total household income" and "rental cost". With these two target variables,

it is possible identify the low-income households with excessive rental cost. The results are summarized in Figure 2.

#### **4. Concluding Remarks**

In this paper we advocate that spatial microsimulation techniques introduce new possibilities for the development of a methodology for measuring housing deficit at small areas. While the count by census tracts presents a good spatial resolution but lacks details on the households, the census sample microdata presents a richer dataset that is suitable for capturing different dimensions of the housing deficit but lacks information on the spatial location of households.

By addressing the shortcoming of both data types, and therefore both place-based and individual-based approaches for measuring housing deficit, we expect to develop a hybrid approach that is able to better depict the housing deficit in Brazilian cities. Such approach should be able to explore not only the full capability of census data, which is available for all municipalities, but also allow the integration with auxiliary data that may be available at the local level, such as natural hazard/risk assessments and municipal data on legal status of land tenure.

Additional tests must be conducted to ensure that the simulated spatial microdata is as representative as possible of the aggregate constraints. For that, we intend to explore the choice of different constraint variables and validate the resulting estimates. In addition, it is important to test and compare different methods for spatial microsimulation by exploring its main features, variability and validity comparing with external datasets.

#### **Acknowledgements**

The authors would like to thank the financial support received from CNPq (Grant 443052/2014-0) and FAPEMIG.

#### **References**

- Hermes, K., Poulsen, M. (2012) "A review of current methods to generate synthetic spatial microdata using reweighting and future directions". *Computers, Environment and Urban Systems*, 36, 281-290.
- Lovelace, R. (2014) "Introducing Spatial Microsimulation with R: A Practical". National Centre for Research Methods Working Paper 08/14. University of Leeds, Leeds, UK.
- Rosemback, R., Rigotti, J., Feitosa, F. and Monteiro, A. (2014) "As dimensões da questão habitacional e o papel dos dados censitários nos diagnósticos municipais: uma sugestão de análise frente às novas exigências da Política Nacional de Habitação", XIX Encontro Nacional de Estudos Populacionais, ABEP, São Pedro, SP.
- Fundação João Pinheiro – FJP. (2013). "Déficit habitacional municipal no Brasil 2010", Fundação João Pinheiro, Centro de Estatística e Informações, Belo Horizonte, MG.