Diego Giuliani (presenter)\textsuperscript{1}, Maria Michela Dickson\textsuperscript{2}, Danila Filipponi\textsuperscript{3}, Giuseppe Espa\textsuperscript{4}

Assessing the importance of spatial sampling designs in business surveys

Abstract

The paper aims at verifying the usefulness of employing spatial sampling designs in business surveys, with particular attention to the case of Italy. The paper is motivated by the fact that the official Italian business register, the \textit{Statistical Archive of Active Enterprises} (ASIA), has been recently geo-referenced in its entirety so that the spatial coordinates of each single business unit are now identified and hence proper spatial sampling designs can be implemented. The ASIA archive, which is managed and updated by the Italian Institute of Statistics (ISTAT), records the entire population of active enterprises operating in the manufacturing and services sectors. It represents the up-to-date reference register for most of the official business sample surveys.

At least since the seminal paper by Arbia (1993), literature on sampling has recognized that spatial heterogeneity in the population data and spatial autocorrelation of the target variable, if not properly taken into account, may negatively affect the efficiency and precision of the sample estimates. Referring to this problem, a recent stream of literature (see Stevens and Olsen, 2004; Cicchitelli and Montanari, 2012; Grafström, 2012; Grafström et al., 2012 among others) proposed different sampling procedures which incorporate the spatial location of units into the design and assure that the selected samples are spatially balanced, that is well spread over the spatial population. It has been proved, in the context of environmental data (see Wang et al., 2012 for a comprehensive review), that spatially balanced samples lead to more efficient estimates than samples selected without considering the spatial aspect.

By means of simulations of sample selections from an observed population of business units, we aim at showing that even in the context of firm data, where the feature of spatial dependence is typically relevant, spatial sampling designs perform better than the non-spatial ones. Furthermore, we also aim at individuating which is the most proper spatial sampling design, amongst those proposed in the literature, to conduct business surveys. In particular, we define a population of firms from the ASIA archive in which the target variable is the firm's sales and the goal is to estimate the population total. The different spatial sampling designs under comparison will be the Generalized Random-Tessellation Stratified method (GRTS) by Stevens and Olsen (2004), the Local Pivotal methods (LP1 and LP2) by Grafström et al. (2012), the Spatially Correlated Poisson Sampling method (SCPS) by Grafström (2012) and the Penalized Balanced Sampling method (PBS) by Breidt and Chauvet (2012). For all designs, the Horvitz–Thompson ratio estimator of the population total will be used both with equal and unequal inclusion probabilities proportional to the number of employees of business units. For GRTS, LP1, LP1 and SCPS the estimator will also be used with calibrated weights based on calibration variables defined by the spatial coordinates under different

1. University of Trento, Italy, diego.giuliani@unitn.it
2. University of Rome “La sapienza”, Italy, mariamichela.dickson@uniroma1.it
3. Italian National Institute of Statistics (ISTAT), Italy, dafilipp@istat.it
4. University of Trento, Italy, giuseppe.espa@economia.unitn.it
specifications. The performance of the designs will be evaluated in terms of efficiency of the HT-ratio estimator.

References