

The Use of Spatial Sampling Designs in Business Surveys

Maria Michela Dickson¹, Roberto Benedetti², Diego Giuliani³, Giuseppe Espa³

¹Department of Social Science, University "Sapienza", Roma, Italy

²Department of Economics, University "G. D'annunzio", Pescara, Italy

³Department of Economics and Management, University of Trento, Trento, Italy

Email: mariamichela.dickson@uniroma1.it

Received 5 June 2014; revised 7 July 2014; accepted 16 July 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

An innovative use of spatial sampling designs is here presented. Sampling methods which consider spatial locations of statistical units are already used in agricultural and environmental contexts, while they have never been exploited for establishment surveys. However, the rapidly increasing availability of georeferenced information about business units makes that possible. In business studies, it may indeed be important to take into account the presence of spatial autocorrelation or spatial trends in the variables of interest, in order to have more precise and efficient estimates. The opportunity of using the most innovative spatial sampling designs in business surveys, in order to produce samples that are well spread in space, is here tested by means of Monte Carlo experiments. For all designs, the Horvitz-Thompson estimator of the population total is used both with equal and unequal inclusion probabilities. The efficiency of sampling designs is evaluated in terms of relative RMSE and efficiency gain compared with designs ignoring the spatial information. Furthermore, an evaluation of spatially balancing samples is also conducted.

Keywords

Spatial Sampling Design, Business Surveys, Horvitz-Thompson Estimator, Spatial Balance

1. Introduction

An important fact has recently started to occur in the context of business surveys. More and more official national business registers in the world have been georeferenced in their entirety. That is, they record the precise spatial coordinates (in terms of longitude and latitude) of each single business unit. For example, to name a few, one should mention the *US Census Bureau's Longitudinal Business Database* (LBD), the *OFS Suisse Geo Stat STATENT* and the *Italian Statistical Archive of Active Enterprises* (ASIA). This detailed kind of spatial infor-

mation has never been exploited in establishment surveys in order to select spatially balanced samples of firms.

Literature on sampling, since the seminal paper by Arbia [1], has recognized that spatial heterogeneity in the population data and spatial autocorrelation of the target variables, if not properly taken into account, may negatively affect the efficiency and precision of the samples estimates. More recently, a stream of literature proposed different sampling procedures which incorporate the spatial location of units into the design and assure that the selected samples are spatially balanced, which are well spread over the spatial population. In the context of environmental data, it has also been proved that spatially balanced samples lead to more efficient estimates than samples selected without considering the spatial aspects.

It is argued that, even in the context of establishment surveys, the feature of spatial dependence is relevant as firms located near in space are more similar than those located far apart. Therefore, it is reasonably expected that spatial sampling designs should perform better than the non-spatial ones, even when conducting business surveys. In this paper, it is aimed for verifying this hypothesis by means of simulations. In particular, referring to an observed population of business units, the purpose is to measure the performance of spatial sampling designs for firm data. Furthermore, it is also aimed for individuating which is the most proper spatial sampling design, amongst those proposed in the literature, to conduct business surveys. In particular, the population of firms has been defined from the Italian 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 Spatially Correlated Poisson Sampling method (SCPS) by Grafström [2], the Local Pivotal methods (LPM1 and LPM2) by Grafström *et al.* [3], the Balanced Sampling (BS) by Deville and Tillé [4] and the Doubly Balanced Spatial Sampling (DBSS) by Grafström and Tillé [5]. For all designs, the Horvitz-Thompson estimator of the population total will be used both with equal and unequal inclusion probabilities proportional to the number of employees of business units. The performance of the designs will be evaluated in terms of efficiency of the HT estimator.

The structure of the paper is the following. In Section 2, the spatial sampling designs that have been tested on the firm data have been briefly introduced. Section 3 contains the results of the simulation study. Finally, Section 4 contains some conclusions and directions for further studies in the field.

2. Spatial Sampling Designs

In the last three decades, spatial sampling field has been involved in a flourishing of proposals to incorporate spatial information in sampling designs. For example, very ample is literature on the exclusion of contiguous units in selecting samples (Hedayat *et al.*, [6] and [7]; Wright and Stufken, [8]). Moreover, another field of research has been concentrating his effort in study how is possible to guide selection of samples giving geographic coordinates as part of available information on the study population. The most important turning point is definitively attributable to Stevens and Olsen [9], which proposed an innovative method to select spatial samples.

The Generalized Random-Tessellation Stratified method (GRTS), proposed by Stevens and Olsen [9], essentially consists on assigning to sampling units an order according to a recursive hierarchical randomization process which preserves the spatial relationships of the sample units. Then, the sampling units are arranged in order and mapped from two- or multi-dimensional space in one-dimensional space, through a quadrant recursive function (Mark, [10]). The samples are selected in one dimension, using systematic πps sampling and then mapped back in the original two- or multi-dimension, preserving the spatial order. GRTS design ensures that drawn samples are well spread over the space.

The GRTS has been the most utilized probability-based method to give spatially-balanced sample for data on a linear, areal or not contiguous space. GRTS supports sampling with unequal selection probabilities and produces samples that are much more evenly distributed over space than an ordinary unequal probability design. On the other hand, the mapping used in GRTS is not perfect because it does not exclude the possibility that units that are close in distance in the original complex space may be mapped rather far apart in the one dimensional space where the sampling is made.

GRTS has been considered as the first success in spatial sampling. Nevertheless, it could result obsolete in terms of time of computing and implementation if compared with new methods exposed in prosecution. Due to these reasons, it is not implemented in simulation in Section 4.

2.1. Spatially Correlated Poisson Sampling Method

The Spatially Correlated Poisson Sampling method (SCPS), proposed by Grafström [2], allows to overcome the

problems related to the mapping procedure of GRTS. The SCPS is a modification of Correlated Poisson Sampling (CPS): a list sequential method for real time sampling which provides the selection of units with unequal inclusion probabilities (Bondesson and Thorburn, [11]). The CPS visits units one by one and give to every units a definite inclusion probability. The inclusion probability assigned to each unit is then updated recursively, based on the previous units sampling outcome and in accordance to a specific updating rule characterized by some weights that control how the inclusion probabilities for a unit should be affected by the sampling outcome for the previous units (Bondesson and Thorburn, [11]).

It is possible to define the vector of inclusion indicators of a random sample \mathbf{x} as $\mathbf{I} = (I_1, I_2, \dots, I_N)$ for a population of N units. Given that, the probability function for CPS can be written as

$$\Pr(\mathbf{I} = \mathbf{x}) = \prod_{i=1}^N \left(\pi_i^{(i-1)} \right)^{x_i} \left(1 - \pi_i^{(i-1)} \right)^{1-x_i}, \quad \mathbf{x} \in \{0, 1\}^N,$$

where $\pi_i^{(i-1)}$ is a function of the first $i-1$ components of the vector \mathbf{x} (Grafström, [12]).

The SCPS design is a modification of CPS, achieved by the introduction of a known distance function $d(i, j)$ between units in the updating rule. It essentially consists on a set of strategies for choosing weights for CPS. The distance function may be the Euclidean distance or another general distance measure. Grafström [2] proposed maximal weights and Gaussian preliminary weights as strategies for choosing the weights. If the order of units in the list is changed, for maximal weights and Gaussian preliminary weights, the second order inclusion probabilities may be changed. Differently from GRTS, the order of the units is not relevant since the weights depend on distances amongst units. The samples drawn will have sample locations well spread over the population, because the efficiency of the method is independent of the ordering of units. The algorithm visit units (one by one and all at once) and decide whether or not the unit should be sampled. The final intent is to create correlations between the inclusion indicators. In presence of negative correlation, units that are close in distance rarely appear simultaneously in the sample.

2.2. Local Pivotal Methods

The Local Pivotal methods (LPM1 and LPM2), proposed by Grafström *et al.* [3], are a spatial extension of the Simple Pivotal method (Deville and Tillé, [13]). LPM methods draw samples considering distances between units and in accordance to the updating rule of the Pivotal method, for two nearby units at each step (see for details, Deville and Tillé, [13]). To choose the two nearby units i and j , it is possible to choose between LPM1, which is more balanced, and the LPM2, which is less balanced but computationally more feasible.

The LPM1 randomly chooses the first unit i and then the nearest neighbor unit j (if two or more unit have the same distance to i , the method randomly choose between them). The inclusion probabilities are updated according to the updating rule of pivotal method. If j is not the nearest neighbor of i , the method restart from the beginning. When all units have been visited, the method stops. To select a LPM1 sample, the needed expected number of iterations is at most proportional to N^3 .

The LPM2 works similarly to the LPM1, but the inclusion probabilities are directly updated with the updating rule of the pivotal method. The expected number of iterations needed to select a sample is in this case proportional to N^2 . The Local Pivotal methods give more spatially balanced samples than GRTS or SCPS for some populations. They imply a smaller variance when a spatial trend exists (Grafström *et al.*, [3]).

In case of a high amount of available auxiliary information, Local Pivotal methods allow to select samples which are well spread in the space. Moreover, if units near in space have similar values, the estimation could be subjected to an improvement. Similarly, it is possible to assign bigger weights to auxiliary variables which have a strong relationship with the variable of interest.

2.3. Balanced Sampling

Balanced Sampling (BS) is a popular method to select random units from a finite population, taking into account a fixed sample size and stratification. Balanced sampling is very important in the case when the auxiliary variables are correlated with the interest variable. A large family of methods to select balanced samples is available, which comprehends those proposed by Yates [14], Royall and Herson [15], Deville *et al.* [16], Hedayat and Majumar [17], Breidt and Chauvet [18].

However, the Cube method proposed by Deville and Tillé [4] changed completely the perspective of balanced

sampling. It is an algorithm to select balanced samples which can use many different auxiliary variables with equal or unequal inclusion probabilities. The Cube method, named after a geometric representation of a sampling design, satisfies predefined inclusion probabilities and drawn samples which are better balanced than those obtained with other balancing methods. It is based on a random transformation of inclusion probabilities vector π to draw a sample s which exactly satisfies the inclusion probabilities and the equation of balancing (see for details, Deville and Tillé, [4]). The Cube method is composed by two phases: the *flight phase* and the *landing phase*. The first one is a random walk which starts on the point of the inclusion probabilities vector and stays within the intersection area between the cube and the constraint subspace. The random walk stops when it arrives in a vertex of the subspace. If the sample is not obtained, then a landing phase is applied where the sample is selected as near as possible to the constraint subspace.

The BS method used in the empirical application of this paper is a particular application of the Cube method, where geographic coordinates have been used as auxiliary variables.

2.4. Doubly Spatial Sampling Method

The Doubly Balanced Spatial Sampling method (DBSS), proposed by Grafström and Tillé [5], combines the LPM2 (Grafström *et al.*, [3]) and the Cube methods (Deville and Tillé, [4]; Tillé, [19]).

DBSS is a method that succeeds in drawing samples which are balanced on auxiliary variables and well spread in the topographical space. First of all, this method selects a cluster of units through LPM2. In particular, the first unit i is selected randomly with equal inclusion probability and then its nearest unit j is subsequently selected. The mean position of the $j+1$ units is computed and then the next unit is selected as the nearest to the mean position. The procedure is repeated as long as the sum of the squares of the distances between cluster units and their mean decreases. When the cluster is selected, DBSS applies the Cube method flight phase on a cluster of units near in distance. Then the sampling outcomes are decided for the units according to the balancing conditions. The second order inclusion probabilities are updated locally. So that they are smaller among nearby units because LPM creates negative correlation between inclusion indicators of nearby units. At the end of the decision process for sampling outcomes, the landing phase of the Cube method is applied.

The DBSS provides better balanced samples, as those selected by the Cube method. In addition, at each selected unit, the inclusion probability is updated for the remaining units in a decreasing order based on the distance between units. It prevents the selection of units near in distance, giving a sample that is well spread in the space.

To implement the DBSS design a vector of geographic coordinates, to spread the population over the space, and a vector of auxiliary variables for balancing procedure are required. In the simulation study of this paper, geographic coordinates are used both to balance and to spread the population, giving a “*dual*” spatial connotation to the algorithm.

3. Data

The data used in this paper derive from the ASIA archive. This database, 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 establishment surveys as it implements the Eurostat’s legislation (art. 1 and 7 of Council Regulation no. 2186/93, [20]) about the statistical harmonization of the business registers. For each business unit, the database currently contains yearly information about firm code, sector of activity (according to the NACE, which is the Statistical Classification of Economic Activities in the European Community), number of firm’s employees, firm’s sales, legal status (according to the current Italian classification), firm’s birth and death dates and the precise spatial locations (in terms of GMT longitude and latitude coordinates).

4. Simulation Study

In this section the performances, in terms of efficiency and spatial balancing, of the various spatial sampling designs, are assessed in an observed population of business units. The reference population used is a subset of the ASIA archive and it is composed by 822 georeferenced non-specialized retail stores located in the Province of Trento, Italy, in 2009. It is assumed that the goal of sampling is to estimate the population total of firm’s sales.

The number of firm's employees is available as an auxiliary variable.

Figure 1 shows the spatial distribution of stores, where it can be clearly noted that the population units are not evenly distributed in space but they tend to concentrate in three particular areas, corresponding to the three biggest towns of the region. Moreover, **Figure 2**, which depicts the empirical semivariogram, shows the presence of relevant spatial autocorrelation in the stores' sales. Denoting the store's sales with y , for any generic stores i and j that have a function distance d , the empirical semivariogram ordinate is the quantity $v(d) = \frac{1}{2}(y_i - y_j)^2$. If the sales of neighboring stores are spatially independent, then $v(d) \cong 1$ for all $d_{ij} > 0$. On the other hand, if the values of sales tend to be similar amongst firms located at a distance d , then $v(d) > 1$. **Figure 2** shows that $v(d)$ has a peak relevantly greater than 1 at $d < 5000$ meters, indicating that stores' sales are characterized by

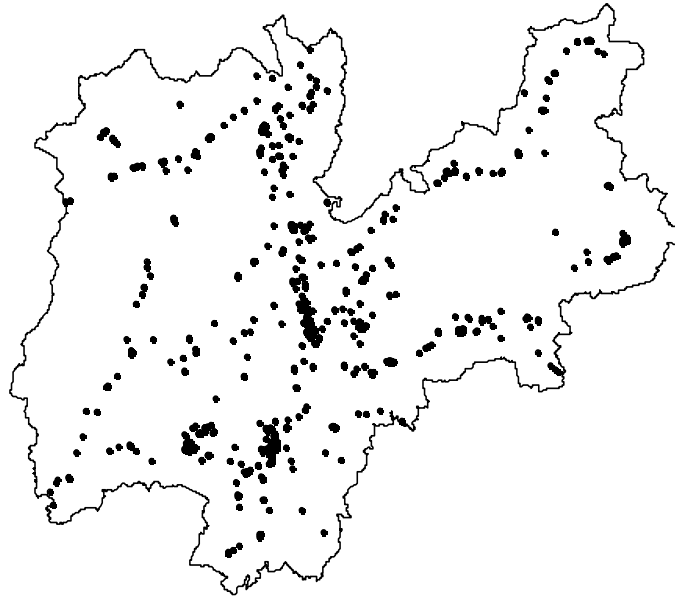


Figure 1. Spatial distribution of population units.

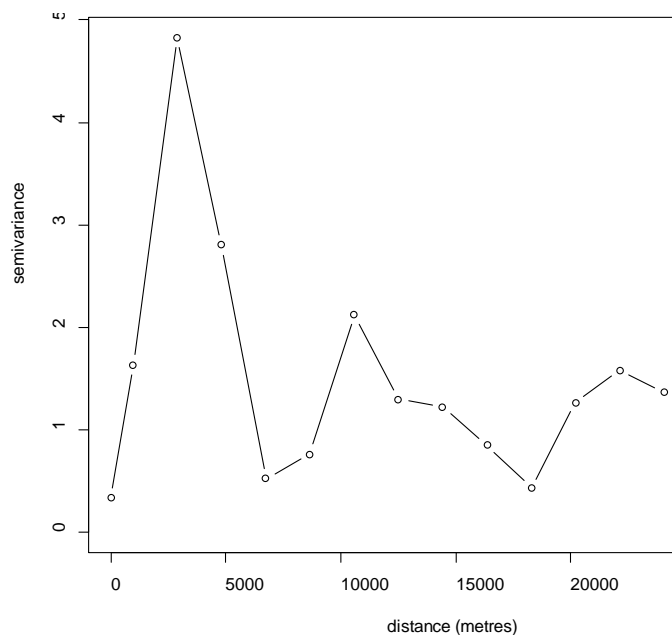


Figure 2. Empirical semivariogram of sales.

a small scale spatial trend which, in turn, suggest that spatial sampling designs may have an important role in producing relatively more precise estimates of population total of sales. Note that the semivariogram ordinates have been averaged in bins of constant width.

In this simulation study, the following sampling designs are compared: LPM1, LPM2, SCPS, BS and DBSS. Six different scenarios are taken into consideration. Indeed, from the population of size $N = 822$, samples of size $n = 50$, $n = 75$ and $n = 100$ units with both equal and unequal (as proportional to the number of employees) inclusion probabilities have been selected. Observing how the results change with respect to the different scenarios allows assessing, on one hand, the presence of some trends in the fixed sample size and, on the other hand, the ability of sampling designs to exploit auxiliary information.

For all designs, the Horvitz-Thompson estimator has been used (Horvitz and Thompson, [21]):

$$\hat{Y}_{HT} = \sum_{i=1}^n \frac{y_i}{\pi_i}$$

where y_i and π_i are the values, respectively, of the target variable and inclusion probability of sample unit i .

For LPM1, LPM2 and SCPS the Horvitz-Thompson estimator has been also used with calibrated weights based on the spatial coordinates taken both in linear and quadratic forms. In the linear case, the calibrating variables are the horizontal and vertical coordinates. The quadratic specification considers also the square of the coordinates and their cross product. On the other hand, the spatial coordinates, in the linear and quadratic specifications, are also used as balancing variables for the BS and DBSS designs. All designs have been implemented using the *R* package “Balanced Sampling” (Grafström, [22]).

A total of 10,000 samples were selected with each design and the relative root mean square error (rRMSE) was estimated as

$$\text{rRMSE} = \frac{\sqrt{\sum_j (\hat{Y}_j - Y)^2 / 10000}}{Y},$$

where Y represents the known population total of firm’s sales and \hat{Y}_j indicates the sample estimate of the j -th sample.

The values of the rRMSE of the different sampling designs have also been compared with the rRMSE’s of the Simple Random Sampling design (SRS) and the Random Pivotal Method design (RPM) in order to better assess the advantage of including space in the sampling procedure. Moreover, it has been computed the following indicator

$$\left(1 - \left(\frac{\text{rRMSE}_{\text{spatial design}}}{\text{rRMSE}_{\text{SRS or RPM}}} \right) \right) 100 (\%)$$

to measure the efficiency gains of spatial designs with respect to SRS or RPM more precisely (see **Table 1** and **Table 2**).

The general indication from **Table 1** and **Table 2** is that rRMSE’s of spatial designs are lesser than those of non-spatial designs.

In the context of equal inclusion probabilities, DBSS proved to be the most efficient design. Good results are also obtained by LPM1, LPM2 and BS, while SCPS has results slightly worse, for all sample sizes. In this context, the calibration of Horvitz-Thompson estimator in quadratic form provided very similar results as those obtained without any calibration or with the linear calibration on the coordinates. Furthermore, balancing on quadratic spatial coordinates has resulted in a high loss in terms of efficiency of sample estimates.

In the unequal inclusion probability scenario, results are markedly different. Firstly, the efficiency gain of spatial designs with respect to RPM design is very high. Secondly, DBSS has been confirmed as the best method for establishment surveys, with the efficiency gain that increase with the increasing of sample size. Thirdly, LPM1, LPM2 and BS are confirmed as efficient methods also in this case. Finally, the performance of SCPS increases greatly. On the other hand, calibrating and balancing on quadratic spatial coordinates do not produce a notable increase in efficiency of estimation.

Since DBSS is a combination of LPM2 and BS, it shows the theory expressed by Grafström and Tillé [5] and by Grafström and Lundström [23], for which samples well spread in the space are also balanced. This is especially true in the present case, where the “doubly spatial” connotation which has been given to DBSS leads to an increase in efficiency, whether in equal or unequal inclusion probability situations.

Table 1. Relative root mean square error of population and efficiency gain for equal inclusion probabilities.

Design	Equal inclusion probabilities					
	Relative RMSE			Efficiency gain		
	$n = 50$	$n = 75$	$n = 100$	$n = 50$	$n = 75$	$n = 100$
SRS	2.7492	2.1797	1.8700			
LPM1	2.6296	2.1297	1.8282	4.35%	2.29%	2.23%
LPM1 (linear calibration)	2.6322	2.1307	1.8287	4.25%	2.25%	2.21%
LPM1 (quadratic calibration)	2.6238	2.1263	1.8275	4.56%	2.45%	2.27%
LPM2	2.5829	2.1332	1.8045	6.05%	2.13%	3.50%
LPM2 (linear calibration)	2.5853	2.1342	1.8050	5.96%	2.09%	3.48%
LPM2 (quadratic calibration)	2.5747	2.1306	1.8039	6.35%	2.25%	3.53%
SCPS	2.6891	2.1237	1.8449	2.19%	2.57%	1.34%
SCPS (linear calibration)	2.6917	2.1246	1.8453	2.09%	2.52%	1.32%
SCPS (quadratic calibration)	2.6777	2.1202	1.8416	2.60%	2.73%	1.52%
BS (linear balancing)	2.6097	2.1426	1.8263	5.08%	1.70%	2.33%
BS (quadratic balancing)	2.6601	2.1472	1.7980	3.24%	1.49%	3.85%
DBSS (linear balancing)	2.5470	2.0832	1.7808	7.36%	4.42%	4.77%
DBSS (quadratic balancing)	2.6445	2.0910	1.7783	3.81%	4.07%	4.90%

Table 2. Relative root mean square error of population and efficiency gain for unequal inclusion probabilities.

Design	Unequal inclusion probabilities					
	Relative RMSE			Efficiency gain		
	$n = 50$	$n = 75$	$n = 100$	$n = 50$	$n = 75$	$n = 100$
RPM	0.4580	0.4343	0.4255			
LPM1	0.2399	0.1298	0.0142	47.61%	70.12%	96.67%
LPM1 (linear calibration)	0.2759	0.1609	0.0712	39.76%	62.95%	83.26%
LPM1 (quadratic calibration)	0.2933	0.1744	0.0875	35.97%	59.83%	79.44%
LPM2	0.2398	0.1299	0.0142	47.64%	70.08%	96.66%
LPM2 (linear calibration)	0.2774	0.1617	0.0716	39.43%	62.77%	83.18%
LPM2 (quadratic calibration)	0.2951	0.1745	0.0876	35.56%	59.82%	79.40%
SCPS	0.2406	0.1297	0.0139	47.46%	70.14%	96.73%
SCPS (linear calibration)	0.2802	0.1619	0.0704	38.82%	62.72%	83.47%
SCPS (quadratic calibration)	0.3005	0.1775	0.0875	34.39%	59.14%	79.43%
BS (linear balancing)	0.2397	0.1316	0.0139	47.66%	69.69%	96.73%
BS (quadratic balancing)	0.2388	0.1325	0.0142	47.85%	69.50%	96.67%
DBSS (linear balancing)	0.2350	0.1280	0.0139	48.70%	70.53%	96.74%
DBSS (quadratic balancing)	0.2365	0.1300	0.0138	48.36%	70.06%	96.76%

The behavior of simulations using geographic coordinates in quadratic explanation is ascribable to the specific nature of the present study, where there is not a quadratic relationship between coordinates. Probably in different kinds of studies about firms it is possible to recognize an advantage to consider coordinates in quadratic form.

Concerning the ability of spatial sampling designs to produce well spread samples, the spatial balance has been measured following the Voronoi polygons approach suggested by Steven and Olsen [9] and recalled by Grafström *et al.* [3]. Supposing that $i \in s$ is a sample unit of sample s , the Voronoi polygon is a subarea of the space which includes all population units located closer to i than to any other j -th sample unit. Let v_i be the sum of the inclusion probabilities of units in the Voronoi polygon of the i -th sample unit, then $E(v_i) = 1$ and $\sum_{i \in s} v_i = \sum_{j \in U} \pi_j = n$ (Grafström *et al.*, [3], pp. 516-517). To have a spatially balanced sample, all v_i in the sample s must be close to 1, and the variance

$$SpB = \frac{1}{n} \sum_{i \in s} (v_i - 1)^2$$

could be used as a measure of spatial balance for the sample s . Therefore, to assess the ability of spatial sampling designs to produce well spread samples, the mean of SpB over the 10,000 drawn samples has been considered (see **Table 3**). Lower is its value, higher is the degree of spatial balance.

As can be noted from **Table 3**, in both the contexts of equal and unequal inclusion probabilities, LPM1, LPM2, SCPS and DBSS tend to produce samples that are much more spatially balanced than the non-spatial designs (SRS/RPM). On the other hand, performances of BS in terms of spatial balance are in line with SRS/RPM results, because spatial coordinates have been used in BS only as auxiliary variables and not as spread variables. The results for BS spatial balance could also be explained by the fact that balancing on the geographical coordinates allows only capturing large-scale spatial trend. However, as shown by the empirical semivariogram of **Figure 2**, the data used in this simulation study are characterized by small-scale spatial autocorrelation that, in contrast, is perfectly captured by LPM1, LPM2, SCPS and DBSS which are the designs that take into account all distances amongst units.

5. Discussion

Spatial sampling designs give a new important point of view to economic analysis. As shown, all spatial designs proposed have a greater efficiency gain with respect to non-spatial designs. In the context of business surveys, results of sampling are strongly related with the peculiarities of the population, wherein it is relevant if there is relevant spatial autocorrelation or not.

Furthermore, a comparison between methods which select samples spread in the space and methods which are balanced on geographic coordinates has been conducted. The result tells that the only method which is spread and balanced in the same time is more efficient, even in the case where the relationship between geographical variables is not linear, *i.e.* quadratic form. Moreover, if coordinates are used both as spread variables and balancing variables, results are even more fulfilling in terms of efficiency in estimation.

Table 3. Spatial balance of drawn samples.

Design	Spatial balance					
	Equal inclusion probabilities			Unequal inclusion probabilities		
	$n = 50$	$n = 75$	$n = 100$	$n = 50$	$n = 75$	$n = 100$
SRS/RPM	0.5718	0.5725	0.5672	0.5649	0.5682	0.5604
LPM1	0.2112	0.2031	0.1980	0.2156	0.2094	0.2055
LPM2	0.2110	0.2038	0.1983	0.2153	0.2094	0.2053
SCPS	0.2262	0.2212	0.2191	0.2310	0.2331	0.2322
BS	0.5304	0.5411	0.5396	0.5310	0.5240	0.5087
DBSS	0.2086	0.2022	0.1955	0.2675	0.2723	0.2601

Future researches in spatial establishment surveys could regard different contexts. It may indeed be possible that in other spatial distributions of establishments with different characteristics, the spatial design shows a different behavior. The availability of more complete information about firms could give an improvement in estimation using spatial sampling designs.

Acknowledgements

The authors are grateful to Anton Grafström (University of Umeå) for his research about spatial sampling which gives the possibility to make progress in economic spatial sampling. They are also grateful to Prof. Yves Tillé (Université de Neuchâtel) for his invaluable guidance and support during the preparation of this paper. Last but not least, they express a sincere thanks to Danila Filipponi of the Italian Institute of Statistics (ISTAT) for her help and assistance with the ASIA archive.

References

- [1] Arbia, G. (1993) The Use of GIS in Spatial Statistical Surveys. *International Statistical Review*, **61**, 339-359. <http://dx.doi.org/10.2307/1403632>
- [2] Grafström, A. (2012) Spatially Correlated Poisson Sampling. *Journal of Statistical Planning and Inference*, **142**, 139-147. <http://dx.doi.org/10.1016/j.jspi.2011.07.003>
- [3] Grafström, A., Lundström, N.L.P. and Schelin, L. (2012) Spatially Balanced Sampling through the Pivotal Method. *Biometrics*, **68**, 514-520. <http://dx.doi.org/10.1111/j.1541-0420.2011.01699.x>
- [4] Deville, J.-C. and Tillé, Y. (2004) Efficient Balanced Sampling: The Cube Method. *Biometrika*, **91**, 893-912. <http://dx.doi.org/10.1093/biomet/91.4.893>
- [5] Grafström, A. and Tillé, Y. (2013) Doubly Spatial Sampling with Spreading and Restitution of Auxiliary Totals. *Environmetrics*, **24**, 120-131. <http://dx.doi.org/10.1002/env.2194>
- [6] Hedayat, A.S., Rao, C.R. and Stufken, J. (1988) Sampling Plans Excluding Contiguous Units. *Journal of Statistical Planning and Inference*, **19**, 159-170. [http://dx.doi.org/10.1016/0378-3758\(88\)90070-5](http://dx.doi.org/10.1016/0378-3758(88)90070-5)
- [7] Hedayat, A.S., Rao, C.R. and Stufken, J. (1988) Designs in Survey Sampling Avoiding Contiguous Units. In: Krishnaiah, P.R. and Rao, C.R., Eds., *Handbook of Statistics, Vol. 6*, Elsevier, Amsterdam, 575-583.
- [8] Wright, J.H. and Stufken, J. (2008) New Balanced Sampling Plans Excluding Adjacent Units. *Journal of Statistical Planning and Inference*, **138**, 3326-3335. <http://dx.doi.org/10.1016/j.jspi.2006.10.020>
- [9] Stevens Jr., D.L. and Olsen, A.R. (2004) Spatially Balanced Sampling of Natural Resources. *Journal of the American Statistical Association*, **99**, 262-278. <http://dx.doi.org/10.1198/016214504000000250>
- [10] Mark, D.M. (1990) Neighbor-Based Properties of Some Orderings of Two-Dimensional Space. *Geographical Analysis*, **2**, 145-157.
- [11] Bondesson, L. and Thorburn, D. (2008) A List Sequential Sampling Method Suitable for Real-Time Sampling. *Scandinavian Journal of Statistics*, **35**, 466-483. <http://dx.doi.org/10.1111/j.1467-9469.2008.00596.x>
- [12] Grafström, A. (2010) Entropy of Unequal Probability Sampling Designs. *Statistical Methodology*, **7**, 84-97. <http://dx.doi.org/10.1016/j.stamet.2009.10.005>
- [13] Deville, J.-C. and Tillé, Y. (1998) Unequal Probability Sampling without Replacement through a Splitting Method. *Biometrika*, **85**, 89-101. <http://dx.doi.org/10.1093/biomet/85.1.89>
- [14] Yates, F. (1949) *Sampling Methods for Census and Surveys*. Griffin, London.
- [15] Royall, R. and Herson, J. (1973) Robust Estimation in Finite Populations I. *Journal of the American Statistical Association*, **68**, 880-889. <http://dx.doi.org/10.1080/01621459.1973.10481440>
- [16] Deville, J.-C., Grosbras, J.-M. and Roth, N. (1988) Efficient Sampling Algorithms and Balanced Sample. In: Payne, R. and Green, P., Eds., *COMPSTAT*, Physica Verlag, Heidelberg, 255-266. http://dx.doi.org/10.1007/978-3-642-46900-8_35
- [17] Hedayat, A. and Majumdar, D. (1995) Generating Desirable Sampling Plans by the Technique of Trade-Off in Experimental Design. *Journal of Statistical Planning and Inference*, **44**, 237-247. [http://dx.doi.org/10.1016/0378-3758\(94\)00044-V](http://dx.doi.org/10.1016/0378-3758(94)00044-V)
- [18] Breidt, F.J. and Chauvet, G. (2012) Penalized Balanced Sampling. *Biometrika*, **99**, 945-958. <http://dx.doi.org/10.1093/biomet/ass033>
- [19] Tillé, Y. (2006) *Sampling Algorithms*. Springer, New York.
- [20] Council Regulation (EEC) No 696/93 of 15 March 1993 on the Statistical Units for the Observation and Analysis of

the Production System in the Community.

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31993R0696:EN:HTML>

- [21] Horvitz, D.G. and Thompson, D.J. (1952) A Generalization of Sampling without Replacement from a Finite Universe. *Journal of the American Statistical Association*, **47**, 663-685.
- [22] Grafström, A. (2014) Balanced Sampling: Balanced and Spatially Balanced Sampling. R Package Version 1.4. <http://www.antongrafstrom.se/balancedsampling>
- [23] Grafström, A. and Lundström, N.L.P. (2013) Why Well Spread Probability Samples Are Balanced. *Open Journal of Statistics*, **3**, 36-41. <http://dx.doi.org/10.4236/ojs.2013.31005>

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).

