AN INTERDISCIPLINARY METHODOLOGY FOR SOCIO-ECONOMIC SEGREGATION ANALYSIS

Antonio De Falco [1], Antonio Irpino[2]

[1] Department of Economics and Statistics, University of Napoli Federico II, (e-mail: antonio.defalco3@unina.it)

[2] Department of Mathematics and Physics, University of Campania "L. Vanvitelli", (e-mail: antonio.irpino@unicampania.it)

ABSTRACT: This paper proposes an original methodology for the analysis of socio-economic residential segregation. The strategy involves employing areal interpolation methods to create population grids, applying a compositional data approach to quantify categorical distributions, and utilising principal component analysis to define an index of socio-economic class composition for each cell in the study area. By combining index values with spatial autocorrelation tools, it is possible to identify and map segregated areas. To test our method, we rely on the latest UK census data (2021) for the Liverpool metropolitan area, using social groups defined according to the National Statistics Socio-economic Classification.

KEYWORDS: residential segregation analysis, grid cells, compositional data analysis, PCA, spatial analysis

1 Introduction

Residential segregation refers to the spatial separation of social groups within urban areas based on factors such as socio-economic status or ethnicity. While not inherently negative, segregation can lead to the formation of urban areas with distinct social compositions and unequal distribution of resources and services. These factors shape the opportunity/constraint structure of individuals, perpetuating and transmitting social inequalities (Musterd, 2020). Over the years, different indices have been proposed to measure the phenomenon according to its dimensions. However, recent re-conceptualisations and more effective measures have introduced new analysis approaches. Within this framework, there is particular interest in developing indices that incorporate the spatial dimension as they are better able to capture population patterns and the variability of segregation across urban space. Segregation measures typically rely on categorical data provided by national statistical agencies and reported for different spatial units, such as census tracts. As a result, segregation studies often use ecological or aggregated units for analysis. Summary statistics describing these spatial units often involve compositional data with a fixed-sum constraint. However, applying standard statistical methods designed for unconstrained data to compositional data, which are constrained to a simplex can introduce bias (Aitchison, 1986). Additionally, the use of aggregated units poses challenges in spatial analysis due to their arbitrary scale of aggregation and delineation of boundaries, which may not align with meaningful divisions relevant to the studied phenomenon. This issue is commonly known as the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984). Moreover, employing aggregated units with irregular and changing geometries further complicates the comparison of urban areas over time or synchronously. To overcome these challenges, the next section presents a novel methodology for analysing socio-economic segregation, addressing the measurement complexities, and ensuring comparability across urban areas.

2 Methodology

The methodological proposal is based on an interdisciplinary approach, incorporating statistical, sociological, and geographical knowledge. The first phase involves using areal interpolation methods, commonly employed in quantitative geography to improve the estimation of population distribution across a territory. Starting with census aggregated units, a dasymetric binary interpolation procedure (Langford, 2013) using satellite data on land use and land cover is applied to enhance the estimation process. This procedure defines a new set of regular hexagonal grids with higher spatial resolution. The use of grid cells allows for diachronic and/or synchronic comparative analyses between urban areas that report different administrative subdivisions. Utilising grid cells instead of standard units provides a flexible tool to effectively address the MAUP, as the spatial resolution of the cells can be easily modified according to the research objectives. After estimating population grid data, a compositional data analysis strategy, as defined by Aitchison (1986), is employed in the second phase. Population data categorical distributions are quantified by performing the clr-logratio transformation, enabling a subsequent correlation-based statistical analysis. Next, the strategy for measuring socio-economic segregation is implemented in the third phase. A weighted principal component analysis (PCA) (Greenacre, 2018) is performed on the clr-coordinates to synthesise the distributions of socio-economic classes into a single factor while reducing the influence of sparsely populated grid cells on the results. Subsequently, the socioeconomic composition index is defined using the scores derived from the first component. For ease of interpretation, the scores are normalised to values ranging from 0 to 100. In the fourth phase, to detect the spatial structure of the index, a spatial autocorrelation analysis is conducted using the Moran index (Moran, 1948). Index values are used as input data, and a different definition of proximity is incorporated in the spatial weight matrix to define the spatial relationships between areal units. This criterion, based on temporal distances, utilises the median time taken by an individual to travel from one cell to another using four different modes of transportation (walking, biking, driving, and transit). This approach may offer a more realistic representation of the degree of connection between areal units and the potential spatial interaction between social groups compared to criteria based on adjacency and geographical distance. Furthermore, the Moran index can consider the degree of clustering of only one population group at a time, but this limitation is overcome by using an index that summarises the distributions of all socio-economic groups. To assess the intensity of the spatial structure of the socio-economic composition index, the global Moran index is first calculated using different temporal distances as thresholds. Then, the local Moran index (Anselin, 1995) is applied to the index by selecting the specification of the spatial weight matrix that maximises the autocorrelation value.

3 Results

The proposal was applied to the metropolitan area of Liverpool. Data were collected from various sources, including the UK Census data for the year 2021, the UK Corine land cover dataset for the year 2018, the Traveline National Dataset (TNDS), and the Open Street Map data. The methodology was implemented in R. Figure 1 displays the local Moran map of the socio-economic composition index, illustrating the spatial patterns of socio-economic groups defined according to the UK National Statistics Socio-economic Classification in the study area. Based on significant local Moran values and PCA loadings, cells were classified into different categories: HH (High-High) spatial clusters indicate higher socio-economic class segregation, LL (Low-Low) spatial clusters indicate lower socio-economic class segregation, HL (High-Low) spatial outliers represent high index values surrounded by low index values, and LH (Low-High) spatial outliers represent low index values surrounded by high index values. Cells with no significant local Moran values are considered non-segregated areas.

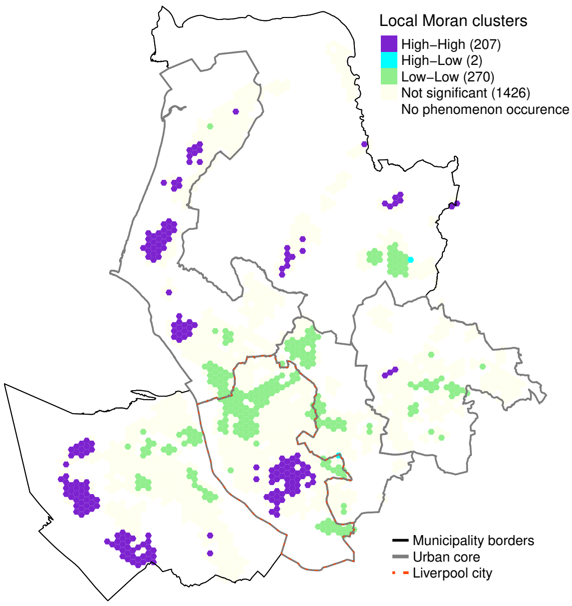

Figure 1. *Local Moran map of the socio-economic composition index. Liverpool, 2021*

References

AITCHISON, J. 1986. *The statistical analysis of compositional data*. London: Chapman and Hall.

ANSELIN, L. 1995. Local Indicators of Spatial Association – LISA. *Geographical Analysis*, **27**, 93–115.

GREENACRE, M. 2018. *Compositional Data Analysis in Practice*. New York: Chapman and Hall/ CRC.

LANGFORD, M. 2013. An evaluation of small area population estimation techniques using open access ancillary data. *Geographical Analysis*, **45**, 324–344.

MORAN, P. A. P. 1948. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society*, **10**, 243–251.

MUSTERD, S. 2020. *Handbook of Urban Segregation*. Cheltenham: Edward Elgar Publishing.

OPENSHAW, S. 1984. *The Modifiable Areal Unit Problem*. Norwich: Geo-Books.