Journal of Geographical Systems (2023) 25:469–488
https://doi.org/10.1007/s10109-023-00412-1

ORIGINAL ARTICLE

Introducing a spatially explicit Gini measure for spatial segregation

Umut Türk[1] · **John Östh**[2]

Received: 25 October 2022 / Accepted: 25 April 2023 / Published online: 14 June 2023

Abstract

This paper proposes an alternative measure of economic segregation by income that utilizes the Gini index as the basis of measurement. The Gini Index of Spatial Segregation (GSS) is a ratio of two Gini indices that compares the inequality between neighbourhoods to the inequality between individuals at the macro-level where neighbourhoods are nested. Unlike earlier measures of segregation found in the literature, the GSS uses individualized neighbourhoods, which can be defined as an area constituted within a radius or as a population count method around an individual geo-location, depending on the population density and proximity among individuals in the study area. The GSS can measure residential segregation by any continuous variable for both radii and k-nearest neighbours (knn with and without a decay factor) approaches to bespoke neighbourhoods. Therefore, it is sensitive to the spatial configuration of the area, easy to compute and interpret, and suitable for comparative studies of segregation over time and across different contexts. An empirical application of the index is illustrated using data from Sweden that covers the entire population for 1994, 2004, and 2014. We demonstrate how the definition and scale of the neighbourhood influence the measures of economic segregation. Overall, the GSS offers a flexible and robust framework for measuring segregation that can be used to inform policy decisions and research on inequality.

Keywords Segregation · Gini index · Sweden · Bespoke neighbourhoods

JEL Classification R10 · R23 · R52 · D63

✉ Umut Türk
umut.turk@agu.edu.tr

John Östh
johnosth@oslomet.no

1 Department of Economics, Abdullah Gül University, Sumer Campus, Kayseri, Turkey

2 Faculty of Technology, Art and Design, Department of Civil Engineering and Energy Technology, Oslo Metropolitan University, Oslo, Norway

1 Introduction

Measuring residential segregation by income, also known as Economic Segregation, has historically received less attention than measuring segregation by race, ethnicity, or occupation. However, in recent years, there has been an increasing number of publications thoroughly discussing economic segregation, such as works by Hardman and Ioannides (2004), Yang and Jargowsky (2006), Tammaru et al. (2014), van Ham et al. (2018), Östh et al. (2018), and Tammaru et al. (2021). Economic and racial segregation share many common factors, as both are spatial phenomena that may occur from similar dynamics and are often empirically intertwined (Reardon 2011). However, the primary methodological difference between the two is that economic segregation requires treatment of continuous variables that identify each individual with unique or similar values. Moreover, while racial segregation refers to the uneven distribution of groups of people who share a particular category across physical space, economic segregation quantifies income homogeneity or diversity in the areas of residents.

To measure the relative clusters of people based on socioeconomic characteristics, it is necessary to aggregate them into a spatial unit known as a "neighbourhood". However, the measurement may vary depending on the chosen definition of a neighbourhood. Relying on a predefined administrative unit such as census tracts or municipalities may lead to erroneous findings, which is known as the modifiable areal unit problem (MAUP) (Openshaw 1983; Wong 2004). The MAUP occurs due to both the scale problem, where the same data show different spatial patterns for varying levels of aggregation, and the zoning problem, where altering the grouping schemes produces different results even if the units are of the same scale.

In terms of the analysis of racial segregation, the scale problem has been recognized and addressed in several ways (Wong 1993, 1999, 2005; Reardon et al. 2008; Hong and Sadahiro 2014; Östh et al. 2015; Sadahiro 2015; Clark and Östh 2018; Olteanu et al. 2020). Wong and Shaw (2011) have made an interesting contribution to the literature on racial segregation by using activity spaces to measure exposure to reference populations beyond the residential space. Since residential segregation is related to clusters of people in space, how geography is handled becomes not only a statistical issue but may also be a crucial strategy to study the effects and causes of segregation. To address this issue, a way is to construct scalable bespoke local environments, where, depending on the definition of neighbourhood, either a set of varying radii (as seen in Lee et al. 2008, for racial segregation) or varying population sizes around an individual location, known as k nearest neighbours (knn), are used (as seen in Östh et al. 2015, for interaction among racial groups). Naturally, such analysis requires data at very fine geographical scales.

In many previous studies, economic segregation has been measured using indices designed for racial segregation, which often divide the population into two categories based on income levels: poor and not poor. In this paper, we introduce the Gini index of spatial segregation (GSS), a new index that can measure residential segregation by any continuous variable using both radii and k-nearest neighbours (knn) approaches to create custom neighbourhoods. As an addition to the spatial segregation literature, we propose a knn function with a decay factor. Unlike knn without decay, knn with decay takes into account the contribution of further away neighbours by applying a decay factor. This results in a neighbourhood definition falling between knn (without decay) and radii approaches. The GSS is sensitive to the spatial arrangement of the area, easy to calculate and interpret, and suitable for comparing segregation over time and across different contexts.

Residential segregation has been shown to result from household sorting across neighbourhoods with differential public goods and services that are excludable for location (Tiebout 1956; Epple et al. 1984). Additionally, segregation can arise from preferences for neighbourhood racial composition (Schelling 1969), as well as education and income (Jargowsky 1996), and from exogenous factors such as changes in the spatial distribution of opportunities and the shift from manufacturing activities to service-oriented economies (Morenoff and Tienda 1997), and demographic changes such as increased female participation in the labour market and an ageing population leading to changes in the demographic composition of cities (Wyly et al. 1998). Similarly, the effects of segregation have been shown to be related to inequality in growth (Burgers and Musterd 2002; Reardon and Bischoff 2011) and the distribution of top 1% income (Essletzbichler 2015). However, the issue of scale in economic segregation has only recently begun to attract scholarly attention (see Türk and Östh 2019).

The aim of this study is to develop an index to measure residential segregation by income while accounting for the scale problem. The proposed index compares the degree of inequality between customized neighbourhoods to that of the larger area in which the neighbourhoods are located and allows for flexibility in determining the size and definition of neighbourhoods. To demonstrate how neighbourhood definition and scale can affect measures of economic segregation, we conducted an empirical study in Sweden. We defined neighbourhoods using both radius and k-nearest neighbour approaches. Our results indicated that measures of segregation varied depending on the chosen neighbourhood definition. To address this issue, we propose a modified k-nearest neighbour algorithm that includes a spatial weights matrix based on the distance between individuals (see Getis 2009). This approach considers not only the spatial distribution of individuals but also the population density within each neighbourhood.

This paper is organized into several sections. The first section summarizes previous work on segregation, including studies that have used the Gini index as a basis for their analysis. In the second section, we introduce our proposed index, the Gini index of spatial segregation (GSS). The third and fourth sections describe the k-nearest neighbour and radii-based neighbourhood approaches we used in our analysis. The fifth section presents a synthetic example of how the GSS is computed. In the sixth section, we present an empirical application of the GSS in Sweden. Finally, the last section provides concluding remarks.

2 Previous work

In this paper, we use a group of segregation measures that define segregation by the ratio of between-neighbourhood variation in mean income to the total variation in income. This variation-over-variation measurement of segregation has been employed in several studies. For instance, Jahn et al. (1947) proposed a Gini measure of racial segregation as a ratio of between-neighbourhood and total inequality. Jargowsky (1996) developed the Neighbourhood Sorting Index (NHI), which computes the ratio of between-population weighted neighbourhood income variance and total variance to measure income segregation. Davidoff (2005) decomposed the variance of income distribution into within- and between-group components, then computed the ratio of the two to measure segregation. Kim and Jargowsky (2009) used the ratio of two Gini indices (between-neighbourhood and overall inequality) as a measure of economic segregation. This last study was based on Jahn et al.'s (1947) work.

In this paper, we adopt the Gini index to measure segregation, building on the approach taken by Jahn et al. (1947) and Kim and Jargowsky (2009). Originally developed to depict income inequality, the Lorenz curve shows the sorted cumulative percentage of total income as a function of the cumulative percentage of total households (Lorenz 1905; Dorfman 1979; Silber 1989). The Gini coefficient is a measure of the area between the Lorenz curve and the line of perfect equality, normalized by the total area under the 45-degree line (see also the Hoover index, Rogerson and Plane 2013). For categorical segregation, the Gini coefficient performs similarly to the dissimilarity index, with the Gini coefficient representing the area between the Lorenz curve normalized by the total area under the 45-degree line for the minority populations, weighted across all pairs of areal units (Massey and Denton 1988). However, this index is limited to measuring segregation between only two population groups. The Gini coefficient takes a maximum value of one when the minority and majority populations are perfectly segregated, and a minimum value of zero when there is no segregation.

An alternative approach to measuring segregation in multiple population groups was proposed by Reardon and Firebaugh (2002), who developed six segregation indices, including the Gini. These indices measure the disproportionality in group proportions across organizational units and are calculated as a weighted sum of the weighted average absolute difference in group proportions between all possible pairs of units across multiple groups. However, the principle of transfers, which applies to two-group segregation measures like the Gini, does not hold for these indices as they are extended to M population groups (Reardon and Firebaugh 2002).

The Gini index forms listed above do not consider the spatial configuration of neighbourhoods, suffer from the modifiable areal unit problem (MAUP), and are insensitive to spatial proximity between areas, also known as the "checkerboard phenomenon" (White 1983). To address these issues for the Gini index, Rey and Smith (2013) propose a spatially weighted version of the Gini that decomposes the index into within and between group differences in inequality. In similar lines, Panzera and Postiglione (2020) decompose the Gini coefficient expressed as the ratio of two covariances and weighted by both spatial proximity and also regional population.

For segregation research, Dawkins (2004) proposes a spatial version of the Gini index that measures categorical segregation with respect to the spatial proximity of neighbourhoods. The standardized spatial Gini index respects the principle of transfer for all segregation ratios and is constant for the multiplicative transformation of the racial composition margin. The decomposition of the index produces within and between components and a residual term that captures the correlation between a neighbourhood's position and the position of its neighbours when ranked by proximity among neighbourhoods.

To extend the standardized spatial Gini index to measure income segregation, Dawkins (2007) proposes a spatial ordering index calculated from either the nearest neighbour or a monocentric spatial ordering of neighbourhood per capita income and the Gini index between-neighbourhood income segregation. The index is represented as a ratio of two covariances, where the numerator is the covariance between neighbourhood aggregate income and spatial reranking of neighbourhoods, while the denominator is the covariance between neighbourhood aggregate income and the average ranking of a neighbourhood. While the spatial ordering index accounts for the spatial relationship between neighbourhoods, it relies on pre-determined administrative boundaries for neighbourhoods.

Below, we present the GSS, where neighbourhoods are constructed around each individual and at varying scales in the study area.

3 The GSS–segregation index

In this section, we introduce the Gini index of spatial segregation (GSS), which is defined as a ratio of two Gini indices as follows:

$$\mathrm{GSS}(y,n) = \frac{\frac{1}{N^2\mu_s}\sum_i \sum_j \left|\mu_{is} - \mu_{js}\right|}{\frac{1}{N^2\mu}\sum_i \sum_j \left|y_i - y_j\right|} \tag{1}$$

where $\mu_{\mathrm{is}} = \frac{\sum_{j\in s} y_j}{n_{\mathrm{is}}}$.

For a population of N individuals, y_i and y_j are the incomes of individuals i and j, μ_{is} and μ_{js} are the average income in individual i's and j's neighbourhoods, which are defined either by radii-based or k nearest neighbour approach around *individual* locations. Therefore, the shape (s) of the neighbourhood varies for the definition chosen and the size (n_{is}) can be set to meet various scales of geography. Note that for any population size $n_{is} \neq N$, μ_s will differ from μ.

The GSS is a weighted measure of income homogeneity/diversity in the areas of residents. That is the ratio between the inequality among bespoke neighbourhoods I_B and the individual-level inequality I_G. It takes a minimum of zero (no segregation) in two cases: if the numerator is zero, thus the between spatial inequality is zero or when the size of the neighbourhood is equal to the size of the whole study area: $n_{is} = N$. While the index takes the maximum value one (perfect segregation) if the distribution of individualized neighbourhood average incomes is identical to that of individual incomes (one individual in a neighbourhood represents whole neighbourhood) thus when $I_B = l_G$ or when the shape of the neighbourhood is too restrictive, i.e., each bespoke neighbourhood includes only one or few income units.

The index has several advantages. First, in contrast to Kim and Jargowsky (2009), the index is sensitive to the spatial configuration of neighbourhoods so that it tackles the checkerboard problem. Second, it does not require the use of predefined administrative units. Thus, it is not subject to the MAUP. Finally, as a ratio of two Gini indices, the GSS preserves several index properties. It respects the Pigou–Dalton principles of transfers, is less sensitive to outliers and deviations from normality and is suitable for the segregation measure of continuous variables. Finally, normalizing the index by individual-level income permits comparative studies among different contexts and over time.

Before the computation of economic segregation by the GSS, we need to define bespoke neighbourhoods. Below we present three alternative methods for constructing individualized neighbourhoods.

3.1 Defining neighbourhoods

In order to better describe exposure possibilities, we create individualized neighbourhoods for each individual's geo-location in the study area. Three methods are considered to define the shape of individual neighbourhoods: the radii-based method, and knn approach and the knn approach with a decay factor. The radii-based neighbourhood depicts the geography constituted within a radius. The shape is determined by the radius and, when applied to position p on the map, includes all observed locations within the specified radius distance from p. This approach is desirable when the analyses focus on locations and when population density is fairly even across space.

When population density varies considerably across geographies, the knn approach might be more appropriate. As a population-count method knn illustrates the interaction possibilities between individuals when the areas are not populated too sparsely. We can also apply an intermediate method by introducing a decay factor to the knn definition that benefits the advantages of both radii and the knn approaches.

Any GIS software can be used to create circles around individual coordinates and to obtain summary statistics within a given radius. In this case, we use ArcGIS software to construct radii-based neighbourhoods where the average disposable income in i's neighbourhood is measured as $\mu_{ir} = \frac{\sum_{j\in r} y_i}{n_{ir}}$ for radius sizes $r = 100m, 1\text{km}, 5\text{km}, 25\text{km}$. Note that in this way radii-based neighbourhoods are created for each individual and are allowed to overlap.

For the construction of the knn-based neighbourhoods, the computations are carried out using the EquiPop software (Östh 2014; Östh and Türk 2020).[1] Each knn-based neighbourhood contains the average disposable income earned by the nearest 100, 200, 400, …, 51,200 working-age (20–64) people for each residential location as follows: $\mu_{\mathrm{ik}} = \sum_{i} \frac{y_i}{k}$. Since we work on the whole of Sweden, the physical separation between neighbours is an issue that we must handle, especially for the country's northern parts. Accordingly, as a third method, we also introduce a distance decay model $f\left(\beta, d_{\mathrm{in}}\right)$, which is a function of the distance between individual i and their k nearest neighbours $n = 1, 2, 3, \ldots, k$ and the distance decay factor β. This operation spatially weights the observations so that as the distance between i and its k nearest neighbours increases, their relative contribution to the average income decreases. Therefore, for densely populated metropolitan areas, the average neighbourhood incomes remain like those produced by the knn algorithm without a decay factor.

Due to high computational burn involved in calculating individual specific distances and decay function, we derive β from half-life model. Half-life models use median value as departure. The median commuted distance always occurs at a distance where half of the population commute longer and half of the population commute shorter distances. Therefore, knowing the maximum distance from i to its kth neighbour, we can say that the probability of interacting with neighbours equals 0.5 at the observed median distance. Then, for the decay function to describe this probability at various distances, the probability-value will decay from one at no distance towards almost zero at far, far away (see Östh et al. 2016, for details). We estimate $\beta = 0.00001153$ with an exponential decay function. The associated weights function becomes $\exp\left(-\beta d_{in}\right)$ for each pair of neighbours.

It is important to note that the size and shape of the neighbourhood can have a significant impact on the results of the GSS calculation, as it determines which neighbouring units are included in the calculation of exposure to different income levels. Therefore, the selection of the appropriate neighbourhood size and shape should be based on the specific research question and the spatial scale of interest. The flexibility of using different catchment areas and neighbourhood shapes is one of the advantages of the GSS. This allows researchers to adapt the method to different spatial scales and characteristics of the study area.

It should also be noted that carrying out the GSS analysis at very local spatial units requires the availability of data at a fine spatial resolution, such as coordinates or street level. However, this can create challenges in applying the GSS method, as it requires the ability to define and operationalize neighbourhoods at a fine spatial resolution. Thus, researchers should carefully consider these limitations and potential data sources when applying the GSS method in practice. Moreover, income is not the only metric for measuring segregation, and there is a wealth of information now available online, for instance through open map services, that the GSS can use to operationalize neighbourhoods and measure segregation.

1 Here, we note that the second version of the software "EquiPop flow" can also integrate frictions (water body, etc.) in the road networks and can be used for defining more refined neighbourhoods.

4 Using the index to measure segregation

To demonstrate how the GSS is computed, we created a synthetic spatial distribution of continuous numbers in a 20×20 Excel matrix, which is provided as a supplementary.xlsx file (Online Source 1). This can be used by users to follow the steps and compute the GSS for any set of spatial distribution of numbers and for two neighbourhood sizes.

In the online version under the "Under the hood" tab, there are three panels showing random, spatially sorted, and checkerboard distributions, with associated Euclidean distances given in the first panel. As the user scrolls down, each local unit in the central 7×7 matrix is populated with neighbourhood averages for smaller and larger catchment areas. Simultaneously, Gini coefficients are calculated under the "Gini_part" tab for corresponding spatial distribution of values at the local unit (individual level) and by smaller and larger neighbourhood sizes with and without a decay factor. The calculation of the decay factor by half-life models is demonstrated under the "Half_life" tab. Finally, the "Control panel" tab shows the GSS values.

The final output from Online Source 1, Control panel tab is shown in Fig. 1, which displays three different spatial distribution systems. In the top system, values were randomly distributed across the matrix. In the middle system, values are spatially sorted such that greater values were clustered in the lower right section of the matrix. Finally, in the bottom system, values were distributed in a checkerboard fashion.

The GSS computation begins by defining the shape of the neighbourhood. For demonstration purposes, we used a smaller rook neighbourhood and a larger catchment area. Assuming 1-unit distances between local spatial units, we computed Euclidean distances to reach the nearest small or large catchment area. Results from four spatial configurations and three different value distributions are illustrated to the right. In the first local spatial distribution, only the local (i) are considered, while in larger catchment areas (2 and 3), a larger area surrounding each local unit is pooled for data. In the small catchment area graph, a small catchment area is shown. In the decay models, a distance decay is introduced using a median commuting time of four minutes, as explained in Online Source 1, "Half_life" tab. However, any median value could be used to derive the decay factor, such as distance, time, cost, and so on.

The graphed values/colours are the result of the computations conducted in the "Under the hood" tab. The "Control panel" values are copied to the Gini calculations, which can be viewed in the "Gini_Part" tab for details. The resulting Gini and GSS values can be reviewed in the Control Panel and shown in Table 1.

In Table 1, the "Local Gini" column corresponds to the individual-level Gini coefficient, which is the denominator of Eq. (1). The GSS values are displayed below the local Gini measures for each spatial distribution type and neighbourhood size. It should be noted that while random value distribution ensures randomness in values, it does not necessarily ensure randomness in the spatial setting. Therefore, a degree of spatial segregation can be observed, as seen in Fig. 1

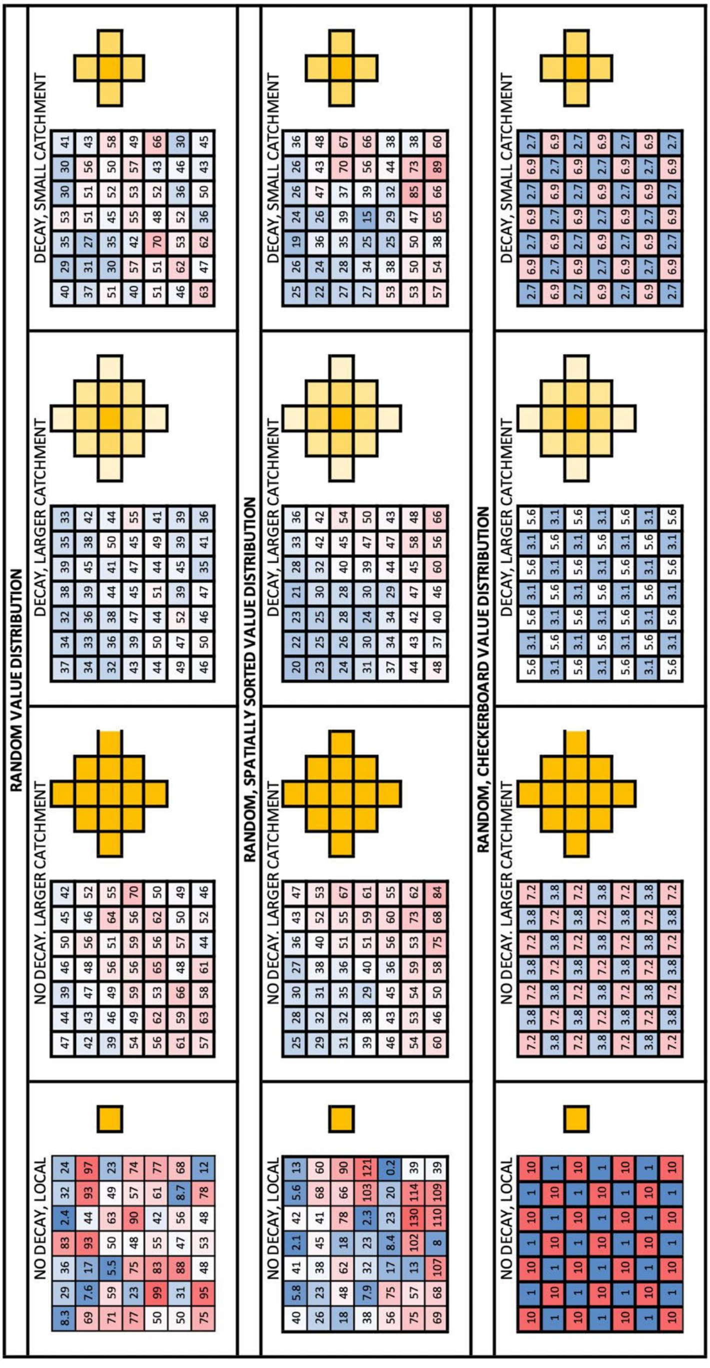

Fig. 1 Synthetic spatial distribution of values and neighbourhood averages based on smaller and larger catchment areas

Table 1 The GSS values computed for synthetic spatial distribution values in Fig. 1

Random value distribution		
	Local Gini	0.2910
GSS	No decay. larger catchment	0.2779
GSS	Decay, larger catchment	0.2737
GSS	Decay, small catchment	0.4347
Random, spatially sorted value distribution		
	Local Gini	0.4145
GSS	No decay. larger catchment	0.3974
GSS	Decay, larger catchment	0.4026
GSS	Decay, small catchment	0.5502
Random, checkerboard value distribution		
	Local Gini	0.4158
GSS	No decay. larger catchment	0.3757
GSS	Decay, larger catchment	0.3130
GSS	Decay, small catchment	0.4659

and Table 1. As per the design, the spatially sorted distribution shows the highest level of segregation, followed by the checkerboard distribution. Introduction of the decay factor results in higher segregation, as individual locations are further isolated due to the decay, resulting in less exposure based on distance. Similarly, smaller catchment areas can detect segregation in greater detail.

The table also demonstrates the level of detail in which GSS can compute segregation for spatial distributions of values that resemble a checkerboard pattern. In the next section, we provide an empirical example in Sweden.

5 Empirical application

This section presents an empirical analysis of the performance of the GSS index using Swedish register data from the PLACE database at Uppsala University. We examine the evolution of economic segregation across three decades, specifically the years 1994, 2004, and 2014. The choice of these years is partly driven by data availability (data for 1990–2017 are currently available). Still, more importantly, these years are representative of typical economic growth periods and allow us to study the reduction in effects connected to economic crises or strong growth (Östh and Lindgren 2012). We use register data that covers the entire resident population of Sweden, with families as our units of analysis. We extract measures of disposable income at the family level and their residential coordinates for each year under scrutiny. A common issue shared by previous studies of segregation is the reliance on predefined administrative units for analysis. This study uses residential coordinates that are available for each individual as 100×100 m grid units in the database.

5.1 Results

In this section, we present first economic segregation in Sweden as calculated by the GSS for the years 1994, 2004, and 2014, and their deviations. Then, the correlation analysis is shown and discussed in relation to public goods expenditures in Swedish municipalities.

Table 2 shows the GSS values for three years, where each value corresponds to a different neighbourhood size, and every second column of a given year reports the index weighted with decaying distance. While the last row shows the overall Gini of disposable income for corresponding year. The GSS reflects the slight increase in the Gini from 1994 to 2004 measured at neighbourhood size $k = 100$; for larger k values, a similar segregation pattern yields so that from 1994 to 2004, residential segregation by income has increased only at very small geography, i.e., among 100 nearest neighbours. Moreover, the GSS values for 2014 show that the increase in (individual level) inequality at the individual level is reflected at any scale of geography. It is plausible to assume that the increases in income inequality will translate into residential segregation by income (Reardon and Bischoff 2011). Our results indicate also that when the magnitude of the increase in income inequality is small, corresponding increase in segregation can only be detected at very local scales of geography.

Furthermore, the estimates for radii-based neighbourhoods are reported in Table 3. The results show increasing index values parallel to an increase in individual-level Gini over the years. There is no direct equivalence between r and k in how much geography is being depicted as we move from one definition to another; however, Fig. 2 offers a useful picture of how the GSS varies between years and for different neighbourhood sizes and definitions. A similar pattern is observed on the left-hand side, for the years 1994–2004, whereas the GSS in the year 2014 (grey line) lies above for all k values. Therefore, what we observe from the knn approach that the residential segregation by income remained at a similar rate from 1994 to

Table 2 The GSS values for 1994, 2004 and 2014 computed by the knn approach with and without a decay function

K	GSS_94	GSSdecay_94	GSS_04	GSSdecay_04	GSS_14	GSSdecay_14
100	0.315	0.319	0.323	0.327	0.328	0.333
200	0.287	0.294	0.287	0.293	0.298	0.306
400	0.264	0.272	0.260	0.269	0.274	0.286
800	0.241	0.253	0.235	0.247	0.255	0.270
1600	0.220	0.235	0.215	0.231	0.237	0.258
3200	0.200	0.220	0.199	0.218	0.221	0.248
6400	0.179	0.204	0.180	0.207	0.204	0.237
12,800	0.156	0.188	0.158	0.192	0.182	0.222
25,600	0.139	0.177	0.139	0.180	0.161	0.209
51,200	0.121	0.172	0.120	0.177	0.140	0.205
Gini(Individual)	0.257		0.262		0.332	

Table 3 The GSS values for 1994, 2004 and 2014 computed by the radii-approach

R	GSS 96(radius)	GSS 04(radius)	GSS 14(radius)
100 *m*	0.436	0.475	0.497
1000 *m*	0.290	0.327	0.373
5000 *m*	0.212	0.264	0.302
10000 *m*	0.186	0.236	0.267
25000 *m*	0.158	0.186	0.209
Gini(Individual)	0.257	0.262	0.332

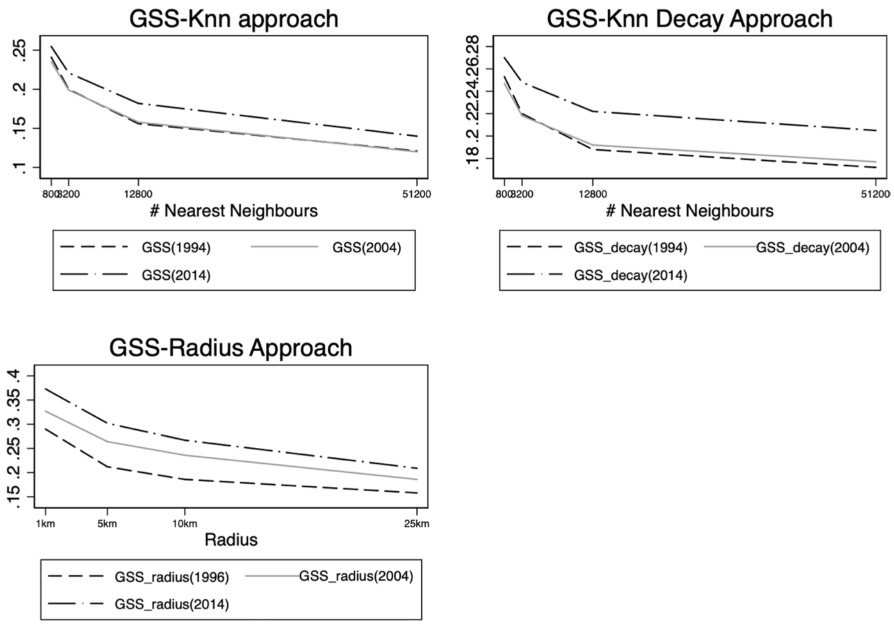

Fig. 2 The GSS measures computed by years and for different types and sizes of neighbourhoods

2004 despite a slight rise in the overall inequality and increased in 2014. The radii-based approach (below) instead shows a clear ranking among the years 1994, 2004 and 2014 with a similar response to different *r* values.

Figure 2 depicts the GSS pattern for various measures of neighbourhood sizes (horizontal axis) for the three years we consider. According to the figure, the radii-approach exhibits a higher level of segregation; however, the GSS values with decay factor display a similar pattern with respect to the radius as decreasing at a decreasing rate for larger sizes of neighbourhoods. The difference between k-nearest and radii-based neighbour approaches becomes clear as we move to the analysis at the municipality level.

We also study the implications of the choice of the individual neighbourhood shape to delimit levels of economic segregation across Swedish municipalities. We computed the GSS separately for 290 Swedish municipalities. Each value represents the ratio of the inequality between average incomes earned in the bespoke neighbourhoods of people who live in the same municipality to the total inequality in the country. We used both radii and k-nearest neighbour aggregates, and for the k-nearest neighbour approach, we reported values both with and without a decay factor (=0.0001153). The results for the year 2014 with different radii and k values are shown in Fig. 3. The colours correspond to the fixed intervals of GSS values for all maps. This makes it easier to compare the values obtained by the two approaches to neighbourhoods.

When examining the maps for different values of r and k, we can observe economic segregation at various levels, from small areas such as individual buildings to larger units such as census tracts. The GSS values display different patterns depending on the chosen approach, especially for lower r and k values. This is because the radii-based approach focuses solely on geography, which means that the number of people living within a given radius can vary between locations and over time. This is particularly noticeable in the first row of Fig. 2, where the GSS values are consistently high (between 0.4 and 0.6) for all municipalities with an r value of 100 m. However, even for a small k value such as 200 (equivalent to a densely populated block), much lower segregation is observed, approaching zero in some municipalities but still retaining high GSS values for metropolitan areas. In Appendix Fig. 5, we focus on the Stockholm greater metropolitan area for the same maps.

As opposed to radii, the knn approach focuses on people and neglects how far they live apart. This becomes a relevant issue, especially for the sparsely populated areas in the country's northern parts where kth neighbour might reside kilometres away from i. The second and the third rows of the figure offer a helpful comparison in this respect. Both maps display a similar pattern for a smaller value $k=200$, while for intermediate k levels, decayed GSS values capture some of the segregation pattern similar to the radii approach. Therefore, the maps on the third row—the knn with a decay factor—lie somewhere between radii and knn maps, rendering both the number and the geographic distribution of residents. Fig. 4 shows the deviation of the computed GSS values between 1994 and 2014. The maps are arranged in the same way as in Fig. 3. In the country's northern parts, economic segregation mostly remained the same and decreased in a couple of northern municipalities. While in the metropolitan areas such as Stockholm, Malmö and Göteborg, it increased, even for higher aggregates of people, i.e. for larger r and k values (see Appendix Fig. 6).

In this paper, we propose a novel index of segregation and demonstrate, using synthetic and real data, how the GSS can be employed to measure segregation for continuous variables such as income. Furthermore, we illustrate three distinct approaches to defining neighbourhoods. The radii-based approach offers a valuable tool for analysing segregation patterns in small, densely populated areas. In contrast, the knn approach is better suited for sparsely populated regions where residents may live far apart. Finally, the knn with a decay factor lies between the radii and knn maps, capturing both the number and the geographic distribution of residents. Additionally, we have provided insights into the temporal changes in

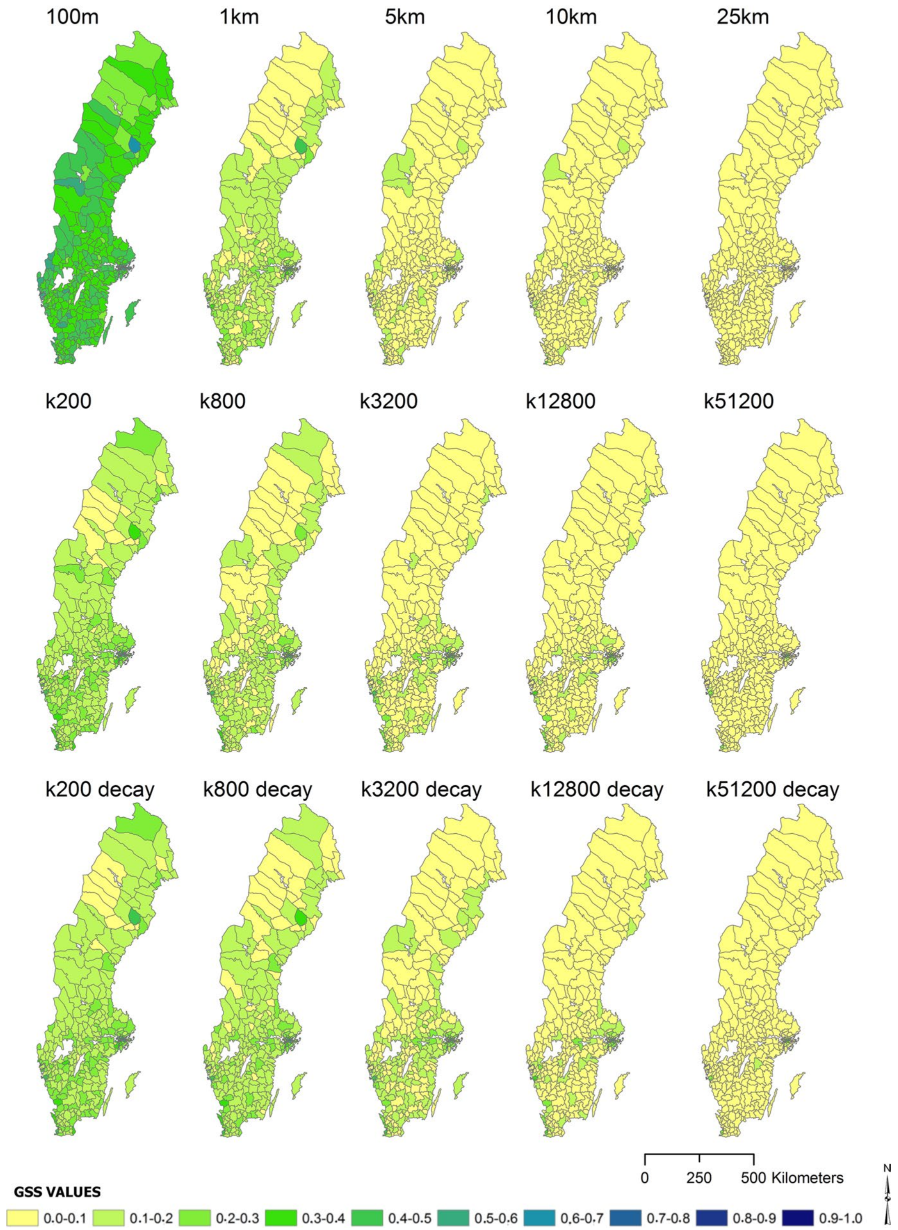

Fig. 3 GSS values for the year 2014. The first shows row radii approach, second row knn and third row knn with decay function

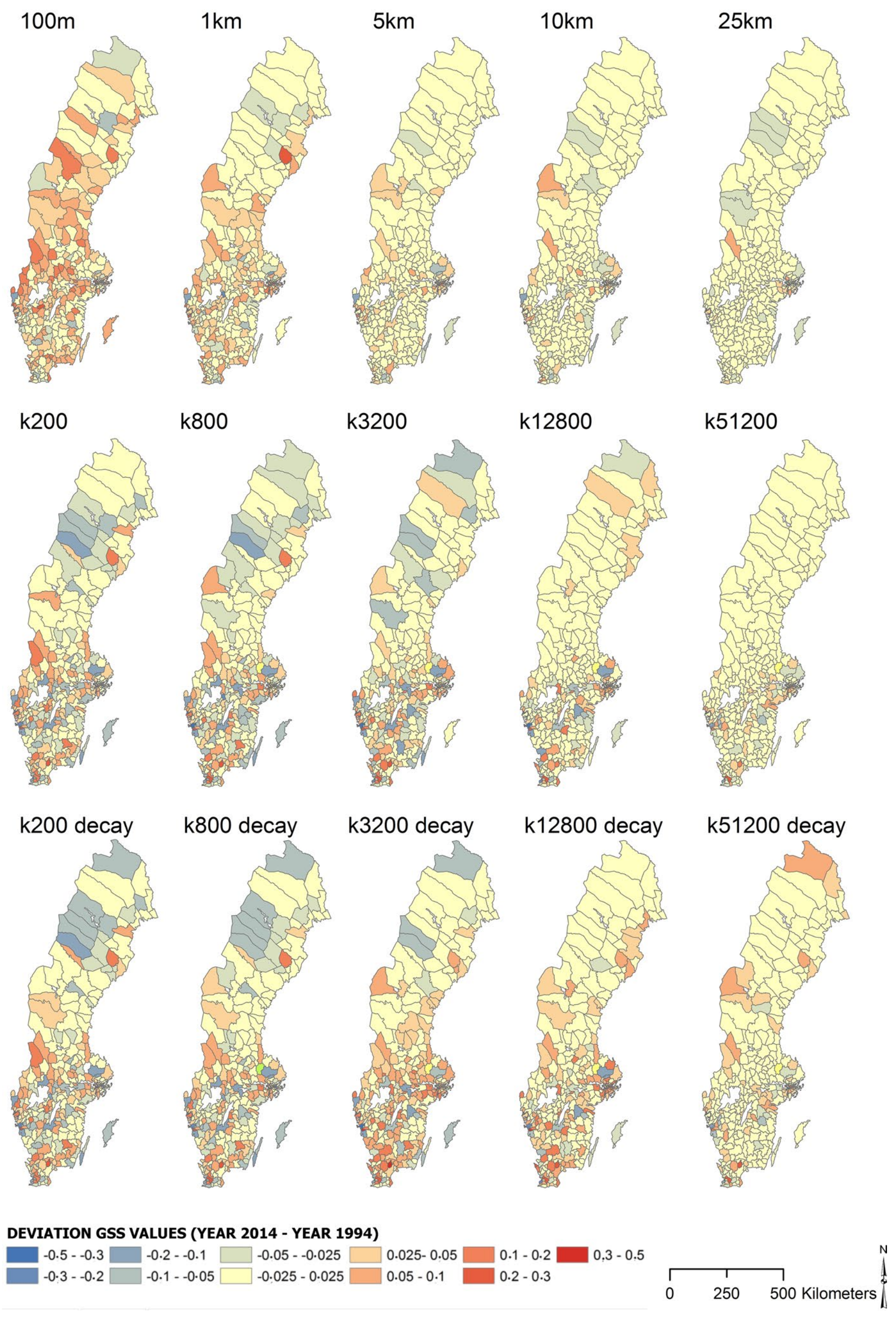

Fig. 4 Deviation of GSS values between 1994 and 2014. The first row shows radii approach, second row knn and third row knn with decay function

segregation patterns over a 20-year period, highlighting differences in segregation trends between densely populated urban areas and sparsely populated rural areas. Our findings suggest that economic segregation has remained relatively stable in the country's northern parts, while increasing in the metropolitan areas of Stockholm, Malmö, and Göteborg. These findings have important implications for policymakers seeking to address issues of economic segregation and promote greater social and economic equality across regions.

6 Conclusions

Traditionally, most of the existing studies of economic segregation use the indices originally developed for racial segregation by dividing the population into two categories: being under and above to a given level of income, i.e., poor, and not poor. By restricting the analysis to two groups, the indices do not make full use of the available information. Moreover, nearly all existing indices are aspatial in nature especially in the way neighbourhoods are considered and do not consider the distribution of individuals in space. Although there exit few spatial indices, they are rather difficult to compute and nearly all use some administratively defined area for the unit of analysis.

In this paper, we propose a new measure of residential segregation by income. The GSS builds upon the variation-over-variation type indices found in the literature. It is based on the individualized-neighbourhood approach and hence, makes use of the full information of the income distribution of residents and their distribution in space. The calculation of the GSS index consists of two steps. First, neighbourhoods are defined by either radii or knn approaches around each individual in the study area. Individualized neighbourhoods are constructed at different scales to study local contexts of various sizes. This last point allows the index to overcome issues associated with MAUP and checkboard phenomenon, the problems that may severely distort the sensitivity of the results of spatial analyses. In the second step, the ratio of inequality between individualized neighbourhoods and inequality between individuals is computed by two Gini indices. As a ratio of two Gini indices, the index has the advantage of preserving desirable properties of the Gini. It respects the Pigou–Dalton principles of transfers, it is less sensitive to outliers, deviations from normality, and finally, it is suitable for the segregation measure of continuous variables.

Moreover, using the Swedish register data, we have used both radii and knn approaches to individualized neighbourhoods and by employing spatial weights matrix based on the distance between neighbours, we have proposed an intermediate approach that benefits the advantages of both. Namely, while radii approach depicts a given neighbourhood by summarizing the characteristics of a population that happens to live from a given distance to individual i without controlling population density, the knn approach counts and summarizes population chrematistics of k number of nearest neighbours without controlling how far apart they live from each other. In a country with heterogeneous spatial distribution of the population, the two approaches produce different results. We have shown in this paper that introducing a decay factor into the knn approach accounts for both population density and distance among residents. Particular to Sweden, the estimates suggest that the economic segregation has remained at a similar degree from 1994 to 2004. While it has increased from 1994 in 2014 in parallel to rising inequality.

Appendix

See Figs. 5, 6

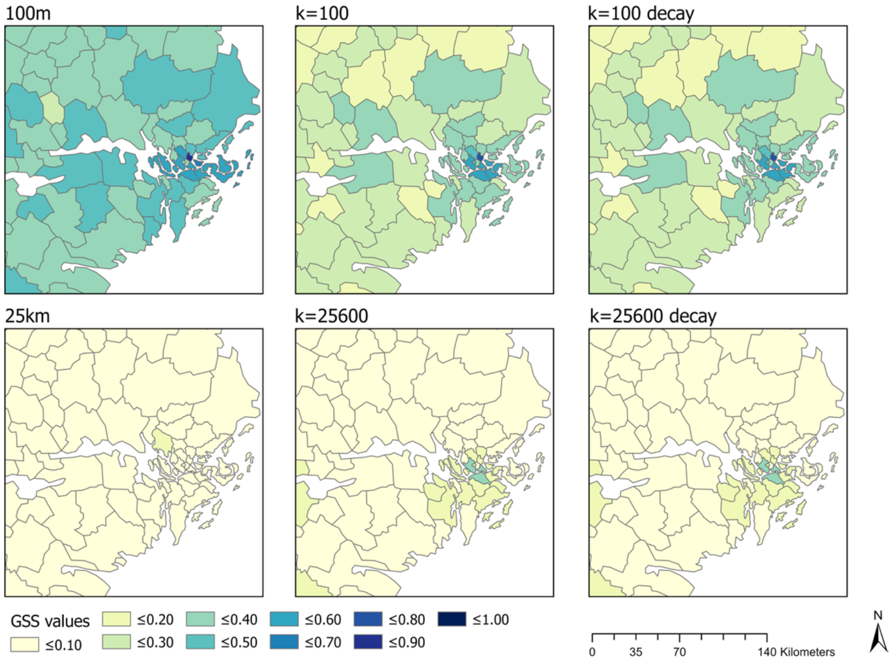

Fig. 5 The GSS values computed for Stockholm greater metropolitan area. The first column shows radii approach, second column the knn approach and the third column the knn with a decay factor

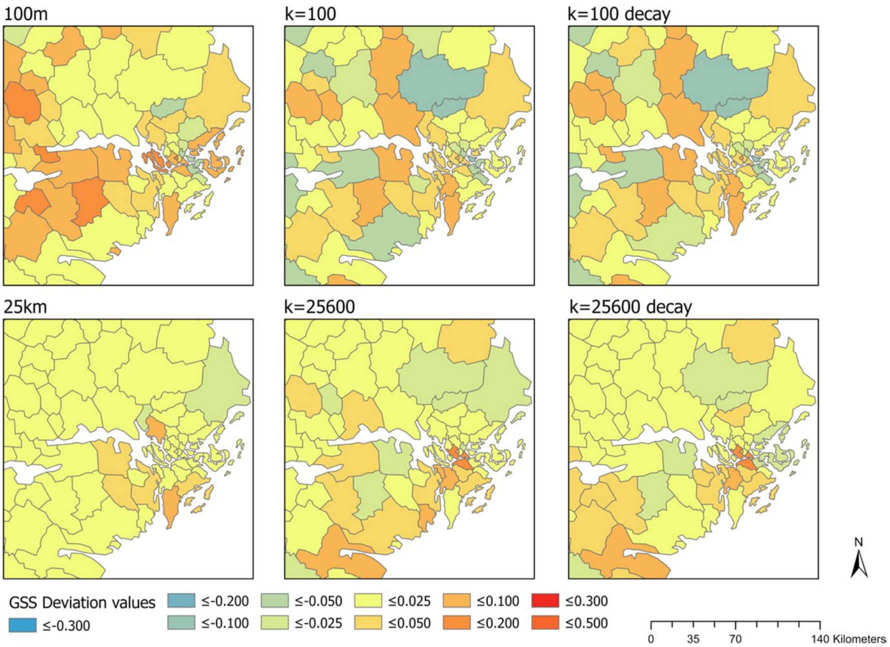

Fig. 6 Deviation of GSS values between 1994 and 2014 in Stockholm greater metropolitan area. The first column shows radii approach, second column knn and third column knn with decay function

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10109-023-00412-1.

Funding No funds, grants, or other support was received.

Data availability The datasets used in this study are register datasets and are subject to copyrights. As a result, they are not publicly available for sharing or distribution.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Burgers J, Musterd S (2002) Understanding urban inequality: a model based on existing theories and an empirical illustration. Int J Urban Reg Res 26(2):403–413. https://doi.org/10.1111/1468-2427.00387

Clark WA, Östh J (2018) Measuring isolation across space and over time with new tools: evidence from Californian metropolitan regions. Environ Plan B: Urban Anal City Sci 45(6):1038–1054. https://doi.org/10.1177/2399808318756642

Davidoff T (2005) Income sorting: measurement and decomposition. J Urban Econ 58(2):289–303. https://doi.org/10.1016/j.jue.2005.05.003

Dawkins CJ (2004) Measuring the spatial pattern of residential segregation. Urban Studies 41(4):833–851. https://doi.org/10.1080/0042098042000194133

Dawkins CJ (2007) Space and the measurement of income segregation. J Reg Sci 47(2):255–272. https://doi.org/10.1111/j.1467-9787.2007.00508.x

Dorfman R (1979) A formula for the Gini coefficient. Rev Econ Stat. https://doi.org/10.2307/1924845
Epple D, Filimon R, Romer T (1984) Equilibrium among local jurisdictions: toward an integrated treatment of voting and residential choice. J Public Econ 24(3):281–308. https://doi.org/10.1016/0047-2727(84)90012-4
Essletzbichler J (2015) The top 1% in US metropolitan areas. Appl Geogr 61:35–46. https://doi.org/10.1016/j.apgeog.2015.01.019
Getis A (2009) Spatial weights matrices. Geogr Anal 41(4):404–410. https://doi.org/10.1111/j.1538-4632.2009.00768.x
Hardman A, Ioannides YM (2004) Neighbors' income distribution: economic segregation and mixing in US urban neighborhoods. J Hous Econ 13(4):368–382. https://doi.org/10.1016/j.jhe.2004.09.003
Hong SY, Sadahiro Y (2014) Measuring geographic segregation: a graph-based approach. J Geogr Syst 16:211–231. https://doi.org/10.1007/s10109-013-0190-7
Jahn J, Schmid CF, Schrag C (1947) The measurement of ecological segregation. Am Sociol Rev 12(3):293–303. https://doi.org/10.2307/2086519
Jargowsky PA (1996) Take the money and run: economic segregation in us metropolitan areas. American Sociol Rev. https://doi.org/10.2307/2096304
Kim J, Jargowsky PA (2009) The GINI coefficient and segregation on a continuous variable. Occup Resid Segregation. https://doi.org/10.1108/s1049-2585(2009)0000017006
Lee BA, Reardon SF, Firebaugh G, Farrell CR, Matthews SA, O'Sullivan D (2008) Beyond the census tract: patterns and determinants of racial segregation at multiple geographic scales. Am Soc Rev 73(5):766–791. https://doi.org/10.1177/000312240807300504
Lorenz MO (1905) Methods of measuring the concentration of wealth. Publ Am Stat Assoc 9(70):209–219. https://doi.org/10.1080/15225437.1905.10503443
Massey DS, Denton NA (1988) The dimensions of residential segregation. Soc Forces 67(2):281–315. https://doi.org/10.2307/2579183
Morenoff JD, Tienda M (1997) Underclass neighborhoods in temporal and ecological perspective. Ann American Acad Polit Soc Sci 551(1):59–72
Olteanu M, de Bezenac C, Clark W, Randon-Furling J (2020) Revealing multiscale segregation effects from fine-scale data: a case study of two communities in Paris. Spatial Demogr 8(3):255–267. https://doi.org/10.1007/s40980-020-00065-4
Openshaw S (1983) The modifiable areal unit problem. Geo Books, Norwick, Norfolk. https://doi.org/10.1177/0002716297551001005
Östh J (2014) Introducing the EquiPop software – an application for the calculation of k-nearest neighbour contexts/neighbourhoods. Available from url: http://equipop.kultgeog.uu.se
Östh J, Lindgren U (2012) Do changes in GDP influence commuting distances? A study of Swedish commuting patterns between 1990 and 2006. Tijdschr Econ Soc Geogr 103(4):443–456. https://doi.org/10.1111/j.1467-9663.2011.00697.x
Östh J, Türk U (2020) Integrating infrastructure and accessibility in measures of bespoke neighbourhoods. In: Musterd S (ed) Handbook on Urban segregation. Edward Elgar, pp 378–394. https://doi.org/10.4337/9781788115605.00031
Östh J, Clark WA, Malmberg B (2015) Measuring the scale of segregation using k-nearest neighbor aggregates. Geogr Anal 47(1):34–49. https://doi.org/10.1111/gean.12053
Östh J, Lyhagen J, Reggiani A (2016) A new way of determining distance decay parameters in spatial interaction models with application to job accessibility analysis in Sweden. ENTER 16(2):344–363. https://doi.org/10.18757/ejtir.2016.16.2.3142
Östh J, Shuttleworth I, Niedomysl T (2018) Spatial and temporal patterns of economic segregation in Sweden's metropolitan areas: a mobility approach. Environ Plan a: Econ Space 50(4):809–825. https://doi.org/10.1177/0308518x18763167
Panzera D, Postiglione P (2020) Measuring the spatial dimension of regional inequality: an approach based on the Gini correlation measure. Soc Indic Res 148(2):379–394. https://doi.org/10.1007/s11205-019-02208-7
Reardon SF, Bischoff K (2011) Income inequality and income segregation. Am J Sociol 116(4):1092–1153. https://doi.org/10.1086/657114
Reardon SF, Firebaugh G (2002) Measures of multigroup segregation. Sociol Methodol 32(1):33–67. https://doi.org/10.1111/1467-9531.00110
Reardon SF, Matthews SA, O'Sullivan D, Lee BA, Firebaugh G, Farrell CR, Bischoff K (2008) The geographic scale of metropolitan racial segregation. Demography 45(3):489–514. https://doi.org/10.1353/dem.0.0019

Reardon SF (2011) Measures of income segregation. Unpublished Working Paper. Stanford Center for Education Policy Analysis. 1(7)
Rey SJ, Smith RJ (2013) A spatial decomposition of the Gini coefficient. Lett Spat Resour Sci 6(2):55–70. https://doi.org/10.1007/s12076-012-0086-z
Rogerson PA, Plane DA (2013) The Hoover index of population concentration and the demographic components of change: an article in memory of Andy Isserman. Int Reg Sci Rev 36(1):97–114. https://doi.org/10.1177/0160017612440811
Sadahiro Y (2015) A method for analyzing the segregation between point distributions: statistical tests and consideration of attributes. J Geogr Syst 17:29–60. https://doi.org/10.1007/s10109-014-0205-z
Schelling TC (1969) Models of segregation. Am Econ Rev 59(2):488–493
Silber J (1989) Factor components, population subgroups and the computation of the Gini index of inequality. Rev Econ Stat. https://doi.org/10.2307/1928057
Tammaru T, Marcińczak S, Van Ham M, Musterd S (2014) Socio-economic segregation in European capital cities. Routledge, Abingdon. https://doi.org/10.2139/ssrn.2713024
Tammaru T, Sinitsyna A, Akhavizadegan A, Ham M V, Marcińczak S, Musterd S (2021) Income Inequality and Residential Segregation in European Cities. In Urban Inequality and Segregation in Europe and China. Springer, Cham. https://doi.org/10.1007/978-3-030-74544-8_3
Tiebout CM (1956) A pure theory of local expenditures. J Polit Econ 64(5):416–424. https://doi.org/10.1086/257839
Türk U, Östh J (2019) How much does geography contribute? Measuring inequality of opportunities using a bespoke neighbourhood approach. J Geogr Syst 21:295–318. https://doi.org/10.1007/s10109-019-00297-z
van Ham M, Tammaru T, Janssen HJ (2018) A multi-level model of vicious circles of socio-economic segregation. Divided cities: understanding intra-urban disparities. OECD, Paris https://doi.org/10.1787/9789264300385-en
White MJ (1983) The measurement of spatial segregation. American J Sociol. https://doi.org/10.1086/227768
Wong DW (1993) Spatial indices of segregation. Urban Studies 30(3):559–572. https://doi.org/10.1080/00420989320080551
Wong DW (1999) Geostatistics as measures of spatial segregation. Urban Geogr 20(7):635–647. https://doi.org/10.2747/0272-3638.20.7.635
Wong DW (2005) Formulating a general spatial segregation measure. Prof Geogr 57(2):285–294. https://doi.org/10.1111/j.0033-0124.2005.00478.x
Wong DW, Shaw SL (2011) Measuring segregation: an activity space approach. J Geogr Syst 13:127–145. https://doi.org/10.1007/s10109-010-0112-x
Wong DW (2004) The modifiable areal unit problem (MAUP) World minds: geographical perspectives on 100 problems, 571–575. https://doi.org/10.1007/978-1-4020-2352-1_93
Wyly EK, Glickman NJ, Lahr ML (1998) A top 10 list of things to know about American cities. Cityscape. https://doi.org/10.2139/ssrn.156910
Yang R, Jargowsky PA (2006) Suburban development and economic segregation in the 1990s. J Urban Aff 28(3):253–273. https://doi.org/10.1111/j.1467-9906.2006.00291.x