ARTICLE IN PRESS

Spatial Statistics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Cluster-based measures of regional concentration. Critical overview

Katarzyna Kopczewska

University of Warsaw, Faculty of Economic Sciences, Ul.Dluga 44/50, 00-241 Warszawa, Poland

ARTICLE INFO

Article history:
Received 15 November 2017
Accepted 29 July 2018
Available online xxxx

Keywords:
Modifiable Areal Unit Problem (MAUP)
Cluster-based measures
Sectoral and geographical concentration
Theory of measures
Sample-size bias
Mantel test

ABSTRACT

This paper provides an overview of the available cluster-based measures of geographical and sectoral concentration (often referred to as specialisation) and tests their statistical behaviour using the Monte Carlo simulation. The study proves that the degree of the aggregation of the dataset matters in the result and that this sensitivity to the Modifiable Areal Unit Problem is inherited in most of the measures. Gini, Krugmann, Theil's H, and Ogive together with other analysed indices are proved to be non-absolute measures that are dependent on the values in the surroundings. Two regions with the same internal industrial (sectoral) structure but with a different share in the overall volume will have a different sectoral concentration index, which limits the inter-regional comparability of these measures. The results also indicate that the information capacity of the measures could be the same, mainly due to the construction of the measures. Thus, in regional comparisons a justified selection of the measures from the different information clusters is a necessity. The empirical ranges of the measures are narrower than the expected theoretical ranges, which causes the interpretation to be more restrictive. The Mantel test and the correlation analysis show, that the innovations in the input data, such as rescaling or permutation, do not alter the results significantly.
© 2018 Elsevier B.V. All rights reserved.

0. Introduction

Over the past 60 years, the regional science literature has developed many cluster-based indicators[1] to measure the sectoral concentration (SC) (often referred to as specialisation) and geographical

E-mail address: kkopczewska@wne.uw.edu.pl.

[1] The literature distinguishes two groups of concentration measures: cluster-based measures, which operate on a two-dimensional table with data aggregated by sectors and regions (such as Ogive, Krugman, entropy measures etc.), and distance-based measures, which operate on individual geo-located point data (e.g. Ripley's K) (more in Duranton and Overman, 2005).

https://doi.org/10.1016/j.spasta.2018.07.008
2211-6753/© 2018 Elsevier B.V. All rights reserved.

concentration (GC). This creates a wide toolbox of methods based on the two-dimensional table by n sectors ($i = 1, 2, \ldots, n$) and m regions ($j = 1, 2, \ldots, m$) for any cumulative target variable y_{ij} (e.g., the employment, number of firms). The most popular measures are the Gini index, the Krugman dissimilarity index, the Theil entropy measure, and the National Average Index, while dozens of others exist. Their detailed overview can be found in Kopczewska (2017a).

A comparison of these measures in terms of their properties is not a novelty. Schmalensee (1977) wrote 40 years ago, "*An infinite number of alternative concentration measures can be computed from typically available data. As long as correlations are not perfect, different measures will sometimes give different signals. Thus, in any particular study, it may matter which of the infinite set of available measures is selected, and the existing literature provides no guidance as to how this selection should best be made*". This statement is still true and furthermore the number of measures has increased. The literature gives some examples of comparisons of the measures, however mainly for the Herfindahl index and other one-dimensional market structure measures (Rosenbluth, 1955; Scherer, 1970; Bailey and Boyle, 1971; Hart, 1971; Kwoka, 1981; Sleuwaegen and Dehandschutter, 1986; Palan, 2010; Latreille and Mackley, 2011 etc.), as well as for Theil's H, Gini and Ellison–Glaeser (López et al., 2012) but they are mostly empirical.[2] Dewhurst and McCann (2002) empirically compared the regional industrial specialisation measures and identified correlations and overlapping information among the measures, as well as the problem of ranking changes as a function of the indicator applied. Even if they observe an interdependence among the indices and in terms of the region sizes, they explain it using agglomeration forces.

The literature on the statistical characteristics and joint relations of the concentration measures is rather modest and predominantly refers to the Gini index. Deltas (2003) demonstrates, using a Monte Carlo simulation, that the one-dimensional (non-regional) Gini index is biased in a small sample (or better, a small number of units). Deltas (2003) recommends using adjusted Gini or other, non-specified, statistics for which the expected value is independent of the sample size. Interestingly, Deltas (2003) does not define the direction of the bias, as the Gini is biased upward in the case of a uniform distribution and downward for a log normal distribution of the data. Fabrizi and Trivisano (2016) estimated the unbiased Gini for small areas based on the Beta regression and applied it to the disposable household income.[3] For other measures, there is a study by Ellison and Glaeser (1997), who claim that a "dartboard effect" exists in traditional concentration measures, which results from discrete zoning; thus, they introduced a correction using the Herfindahl component. Menon (2012) filters out the industry-specific "noise" from the industrial concentration measures, which is estimated with a Monte Carlo simulation and allows for an unbiased estimate of the industry concentration indices. However, there is a gap in the literature on properties and comparisons of the wider range of concentration measures on non-real data.

Studies on the Modifiable Areal Unit Problem (MAUP) in the context of concentration measures appear to be the best developed issue. The current literature, starting from Openshaw (1978, 1984), throughout the problem overviews (e.g. Wong, 2009) provides a large amount of evidence on the significance of this phenomenon. Stillwell et al. have only recently proved the MAUP failure for demographic measures in empirical migration analysis. Briant et al. (2010) analysed the MAUP empirically for the Gini and Ellison–Glaeser and proved that the scaling effect is much stronger than the zoning effect.[4] Arbia (1989) and Amrhein (1995) proved that zoning and scaling do not impact the average of data generated randomly and without spatial autocorrelation. Dewhurst and McCann

2 One should be aware of the concurrence of the names of measures, which have in fact different characteristics. In industrial economics the concentration measures are one-dimensional and they summarise the share of firms in the market; thus, they use the firm-level data in a given sector (implicitly on a given territory). In regional economics, the concentration measures summarise the distribution of industries within the region (or opposite), using two-dimensional firm-level data aggregated by industries and territories. The similarity is that since the 1990s and the appearance of the Ellison–Glaeser index, the Herfindahl index (industrial economics concentration measure) was incorporated into other concentration measures from regional science.

3 Gini can be also treated as an equivalence for variance and association measures, and it can be applied in other methods instead of typical variance and association measures (Yitzhaki and Schechtman, 2012).

4 They also test if the aggregation and zoning impacts the different variables in the same way following Amrhein (1995), who shows that the regression coefficients depend on the size and shape of the analysed regions.

(2007) claim that "*The results of these findings suggest that the testing of agglomeration economies and new economic geography models appears to be somewhat dependent on both the quality and also level of disaggregation of the data. Unfortunately, there is no theoretical guidance as to what is the most appropriate level of sectoral disaggregation.*" More than a decade after Dewhurst and McCann (2007), this gap still exists. Thus, the goal of this paper is to address simulation analysis of the available indicators of the sectoral and geographical concentration, which could shed some light on and give some guidance to the sectoral and regional aggregation issues.

From an analytical perspective, there are few differences between those measures:

(a) The cluster-based measures when calculated for territories (by industry), are referred to as the sectoral concentration (often specialisation), because they analyse the sectoral distribution of the economic activity within a single region;
(b) The cluster-based measures when calculated for industries (by regions), are referred to as the geographical concentration, because they analyse the geographical (inter-regional) distribution of the economic activity within a single industry;
(c) The measures are mostly constructed as the ratio of the observed distribution of activity to some benchmark distribution, which could be the empirical one given by global marginal distributions or the theoretical one given by the equal distribution (also referred to as a random uniform distribution);
(d) Most of the measures are calculated for a given industry or given region, but one can find the local measures for each cell (as the Location Quotient), as well as the global ones, which give a single number for the whole table.

There are a few important points about those measures, that require further study and clarification, which constitutes the purpose of this paper. **First**, there are no studies that indicate the sensitivity of the cluster-based measures to the degree of the dataset aggregation and the MAUP (Modifiable Areal Unit Problem), which stems from the number of regions and industries used in the analysis. This paper tests the robustness of the results for different scales of data aggregation, which is also referred to as sample-size bias. **Second**, one can question the extent to which the scale of the phenomenon measured (average and standard deviation of the variable studied) impacts the results. This paper tests whether the indices are comparable for large and small regions. **Third**, there are no studies on the sensitivity of these measures for permutations of data inside the table. This paper studies whether measures can reflect the changes in the inter-regional location pattern and structural inter-sectoral shifts. Permutation also acts as a counterfactual study, equivalent to Pseudo Statistical Areas,[5] to study the zoning effect (Menon, 2012). **Fourth**, even if the extreme values of these measures are commonly known, and applied to interpret the phenomenon, the practice shows that many measures very rarely achieve their extreme values. This paper studies the range in which the results appear most often, which is helpful in rescaling the intervals for interpretation. **Fifth**, the current body of literature offers in fact a random selection of the measures, which will be used for comparisons in empirical studies and consequently to analyse the similarities (or differences) of the results. This paper studies the correlations between the measures and thus the similarity of the information content. It gives a clue as to which measures should be selected in order to obtain robust results and to avoid doubling the information. **Sixth**, the literature assumes that the absolute values of the measures are fully comparable and independent of other factors such as the size of the region or sector ("scale-free measure" — see e.g., (Eisenberg, 2015). This paper compares the results under different scenarios and tests this independence.

This paper is thus a statistical simulation in-depth analysis of the cluster-based measures of the geographical and sectoral concentration offered in the regional science literature. For simulated data, it compares the sectoral concentration (SC) and geographical concentration (GC) indices in many cross-sections. The paper answers important questions on the properties of the measures: structural stability, information capacity, bias, volatility, correlations, the sensitivity to the MAUP, data scale, dispersion and permutations. Knowledge of these features is necessary for the deliberate selection of measures to be used in concentration analysis.

[5] From Menon (2012): "*The pseudo statistical areas (PSAs) are internally connected spatial units, created by randomly aggregating contiguous counties (or municipalities) in order to build a counterfactual for the functional regions*".

Table 1
Two-dimensional table as the basis for most of the cluster-based measures.
Source: Own concept.

		m regions, $j = 1,2,.., m$					
		Territory A	Territory B	Territory C	Territory D	...	Total
n sectors, $j = 1,2,.., n$	Industry 1	y_{ij}	y_{ij}	y_{ij}	y_{ij}	...	Σy(ind1) $= \sum_j y_{ij}$
	Industry 2	y_{ij}	y_{ij}	y_{ij}	y_{ij}	...	Σy(ind2) $= \sum_j y_{ij}$
	Industry 3	y_{ij}	y_{ij}	y_{ij}	y_{ij}	...	Σy(ind3) $= \sum_j y_{ij}$
	Industry 4	y_{ij}	y_{ij}	y_{ij}	y_{ij}	...	Σy(ind4) $= \sum_j y_{ij}$
	...	...	...	...	...	...	...
	Total	Σy(terrA) $= \sum_i y_{ij}$	Σy(terrB) $= \sum_i y_{ij}$	Σy(terrC) $= \sum_i y_{ij}$	Σy(terrD) $= \sum_i y_{ij}$	...	$\Sigma\Sigma$y $= \sum_i \sum_j y_{ij}$

1. Measures of geographical and sectoral concentration

The starting point for the cluster-based measures is the two-dimensional contingency table (see Table 1) by n sectors ($i = 1, 2, .., n$) and m regions ($j = 1, 2, \ldots, m$) for the target variable (e.g. employment, number of business units). It aggregates the variable y_{ij} within each "cell". It enables the inter-sectoral and inter-regional comparisons of business activity but limits the analysis inside the "cell" because of data aggregation.[6]

The first class of cluster-based measures is calculated by industries for a given region to explain the **sectoral concentration**[7] of the economic activity within a given region. The main difference is the underlying benchmark distribution, to which empirical values are compared. The benchmark distribution, usually in the denominator of the index, can be empirical — based on the total values from the marginal distributions of the cross-table (as in Table 1), uniform — based on equal values from uniform distribution (global total divided by the number of units), transformed empirical — based on the non-linear modifications (e.g., ranks) of the empirical marginal distribution. The analytical details are shown in Appendix 1. Thus, there are four groups of these measures, depending on the benchmark distribution used in the construction of the measure:

(a) Based on the empirical benchmark distribution, where the empirical values are compared with empirical global distributions. Here belong the National Averages index (NAI), the Krugman Dissimilarity index, the Relative Diversity index (RDI) (calculated as the inverse Krugman dissimilarity index), the Hachman index, the Hallet index, the Kullback–Leibler Divergence (KLD) and the Lilien index (the dynamic index).
(b) Based on the uniform benchmark distribution, where the random (equal) values of the target variable in the sector/region are assigned as the reference point and the empirical values are compared with the hypothetical ones. Here belong the Shannon's H, the Theil's H, the Relative H, the Ogive index and the Refined Diversification index.
(c) Based on the transformed empirical distribution, which erases the straightforward features of the empirical marginal distribution. Here belong the Gini coefficient and the Relative Specialisation index (RSI) based on the maximum value of the Location Quotient.
(d) Based on no marginal distribution of the target variable table but using the number of firms. Here belong the Herfindahl index (HH) and the Absolute Diversity index (ADI) (calculated as an inverse HH index).

The second class of cluster-based measures, calculated by regions for a given industry, explains the **geographical concentration** of the economic activity within the given industry. There are also four groups of these measures:

6 For this reason, to enable the inside-cell analysis, distance-based measures have been developed well in the previous years.

7 Many authors in the literature call these "*specialisation measures*". As proven by Churski et al. (2018), regional specialisation cannot be simply derived from the sectoral concentration. Thus, we suggest limiting the interpretation of this group of measures to the sectoral concentration, which is a feasible one that results from the information content.

(a) Based on the empirical benchmark distribution. Here belong the Krugman Concentration index, the Bruhlart & Traeger index, the Agglomeration V and the Clustering index (also the Bergstrand index).
(b) Based on the uniform benchmark distribution, where random (equal) values of the target variable in the sector/region are assigned as the reference point. Here belongs: the Relative H, the Theil's H, the Shannon's H and the Kullback–Leibler Divergence (KLD).
(c) Based on the transformed empirical distribution, which erases the straightforward features of the empirical marginal distribution. Here belong the Gini index, the Locational Gini index, the Ellison–Glaeser index (EG) and the Maurell–Sedillot index (MS).
(d) Based on no marginal distribution of the target variable table. Here belongs the Guillain & LeGallo index, calculated as Moran's I for LQ.

Moreover, there are the **overall concentration** measures for the whole economy, such as the Theil total index (using a uniform marginal distribution) and the Geographic Concentration index (using an empirical marginal distribution). Finally, there is the Location Quotient (LQ), calculated for each cell (region and industry), based on the empirical marginal distribution. Table 2 summarises the extreme values of these measures and their interpretation. Appendix 1 gives the details of the measures.

2. The design and methodology of the study

In this study, there are six working hypotheses to be tested. **The first hypothesis** ($H1$) is that the size of the data table (the number of sectors and regions) does not matter for the measure values. In other words, this is the sensitivity to the MAUP, which is crucial for the decisions about the aggregation scale.[8] The MAUP encompasses two effects: *zoning* (grouping), which is to relocate the borders of a fixed number of sub-regions and *scaling*, which is to set the level of aggregation for the fixed borders of the sub-regions (rich literature after (Openshaw, 1984); e.g., (Manley, 2014)). Both issues will be tested in this context, scaling by using different levels of dataset aggregation and zoning by applying permutation design.[9] **The second hypothesis** ($H2$) is that the scale and the variation of the input data do matter for the result. This concern is the issue of sensitivity to the different distributions, the extreme concentration, data dispersion and the empty cells.[10] **The third hypothesis** ($H3$) is that the measures are location-invariant, and thus, the allocation (mathematical permutation) of the values does not matter for the overall result. This approach is to model the situation of relocating the already existing businesses among the regions (horizontal shifts) as well as sectoral transformation when business changes the activity profile (vertical shifts). When the measures are location-invariant and the permutation does not change the result, this serves as proof that the measures are a poor reflection of the changes in the data and are imperfect tools for tracking slight inter-regional or inter-sectoral shifts in business. This approach follows also the concept of pseudo statistical areas (Menon, 2012), which act as a counterfactual to run comparisons. **The fourth hypothesis** *(H4)* is that the extreme values of the measures, although theoretically achievable, appear relatively rarely in the empirical results and the measures have extra their "real" intervals of values, which are more suitable for interpretation purposes. **The fifth hypothesis** *(H5)* is that because of the similar information content, the information capacity of the measures is similar within some groups. Thus, the more comprehensive summary of the empirical data analysed requires using the measures from the different information clusters as well as the coherent interpretation consistent with the information content of the measure. **The sixth hypothesis** *(H6)* is that the absolute values of the measures are fully comparable between cases and they are independent of the other factors and characteristics of the regions/sectors as e.g., from the share in the region/sector in the overall volume.

8 Deltas (2003) treats this as sample-size bias, because his analysis is for the one-dimensional Gini index.

9 The permutation allows for a counterfactual study, and it is in fact the equivalent of Pseudo Statistical Areas that allow for studying the zoning effect.

10 Even if this is a case of affine transformation and can be studied analytically, for comprehensiveness of the paper and in line with the approach used, it was studied here in simulation. Deltas (2003) simulation shows that bias in Gini is independent of scale (mean of the distribution) but increasing with the dispersion.

Table 2
Typology and interpretation of the cluster-based concentration measures.
Source: Own summary.

Type of concentration	Benchmark marginal distribution	Indicator	Rules of interpretation
sectoral	uniform	Shannon's H	From 0 to ln(n) 0 for full concentration of industry Max for equal share (full diversification)
		Relative H	From 0 to 1 0 for full concentration of industry 1 for equal shares of industries within region
		Theil's H	From 0 to ln(n) 0 for equal (uniform) distribution of sectors within the region Max for fully unequal distribution of sectors within the region
		Ogive index	From 0 to max 0 for equal shares of industry The more unequal the shares, the higher the Ogive measure
		Refined Diversification index	From 0 to 1 0 for the full diversification of region (equal shares) 1 for complete non-diversification
sectoral	empirical	National Averages index (NAI)	From 0 to max 0 for low disparity between regional and national structure Max for significant disparity between regional and national structure
		Kullback–Leibler Divergence KLD	From 0 to max 0 for regional structure most similar to the national one Max for regional structure most dissimilar to the national one
		Lilien index	For dynamic data, From 0 to max 0 for structural stability over time Max for significant shifts between industries over time
		Krugman Dissimilarity index	From 0 to max 0 for industrial structure fully consistent with the referential one The more dissimilar the structure, the higher the Krugman measure
		Hallet index	From 0 to 0.5, Half of Krugman index 0 for industrial structure fully consistent with the referential one (no sectoral concentration) 0.5 when the structures differ significantly (usually sectoral concentration).
		Hachman index	From 0 to 1 0 when region has completely different structure than the country 1 when region has exactly the same industrial structure as country
		Relative Diversity Index (RDI)	inverse Krugman dissimilarity index, from 0 to max 0 for similar structure of regional and national economy Max for dissimilar structures

(*continued on next page*)

Please cite this article in press as: Kopczewska K., Cluster-based measures of regional concentration. Critical overview. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.07.008.

ARTICLE IN PRESS

Table 2 (*continued*)

Type of concentration	Benchmark marginal distribution	Indicator	Rules of interpretation
sectoral	transformed empirical	Gini index	From 0 to 1 0 for uniform distribution of activity among sectors within the region 1 for full sectoral concentration in the region
		RSI (max LQ)	Max LQ in region (by sectors), From 0 do max 0 for underrepresentation of all sectors in the region Max for the degree of over/under-representation of sector in the region
sectoral	no distribution	Herfindahl (HH) index	For firm's size data, From 0 to 1 0 for even distribution (e.g., of employment among the firms) 1 for extreme concentration (e.g., of employment in few (or single) firm), monopolisation
		Absolute Diversity Index (ADI) (inverse HH)	Inverse Herfindahl index, From 0 to max 0 for even distribution (e.g., of employment among the firms) max for extreme concentration (e.g., of employment in few (or single) firm), monopolisation
geographical	uniform	Kullback–Leibler Divergence (KLD)	From 0 to max 0 for complete spatial dispersion of business max for extreme geographical concentration
		Shannon's H	From 0 to ln(n) 0 for full geographical concentration of industry (e.g. all firms from given sector in one region) Max for equal share (fully equal allocation of sectoral business to regions)
		Relative H	From 0 to 1 0 for full geographical concentration of industry in one region 1 for equal regional shares in the industry
		Theil's H	From 0 to max 0 for equal (uniform) distribution of activity between regions within the sector Max for fully unequal distribution of activity among the regions
geographical	empirical	Krugman Concentration index	From 0 to max 0 for activity from given industry allocated proportionally to the region's size Max for activity from industry allocated to a single region only
		Bruelhart&Traeger index	From 0 to max 0 for dispersion of industry (equal shares) Max for geographical concentration of industry
		Agglomeration V	From 0 to max V<1 for differences in sector are smaller than differences in country, which indicates that the given sector is less geographically concentrated than the overall economy. V>1 for larger regional than national differences, which proves that the given sector is more geographically concentrated than the overall economy
		Clustering index (Bergstrand)	From 1 to max

(*continued on next page*)

Please cite this article in press as: Kopczewska K., Cluster-based measures of regional concentration. Critical overview. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.07.008.

ARTICLE IN PRESS

Table 2 (*continued*)

Type of concentration	Benchmark marginal distribution	Indicator	Rules of interpretation
			1 for similar distribution of activity in the sector and in the whole economy, weighted with the distance between regions. C > 1 for neighbouring regions that have similar shares of given activity; the higher the C value, the stronger the clustering
geographical	transformed empirical	Gini index	From 0 to 1 0 for uniform distribution of activity among regions within the sector 1 for full geographical concentration in the sector
		locational Gini	From 0 to 0.5 0 for equal distribution (between regions) of activity in sector and whole economy 0.5 full concentration of activity in single region only
		Maurell–Sedillot index (MS)	From –max to +max, includes Herfindahl index MS<0 for spatial dispersion MS=0 for random distribution of activity among regions MS>0 geographical concentration of business: MS<0.02 low concentration MS>0.05 high concentration
		Ellison–Glaeser index (EG)	From –max to +max, includes Herfindahl index EG<0 for spatial dispersion EG=0 for random distribution of activity among regions EG>0 geographical concentration of business: EG<0.02 low concentration EG>0.05 high concentration
geographical	no distribution	Guillain & LeGallo (Moran for LQ)	Joint interpretation of Moran's I & Gini or LQ index, includes information on spatial location and spatial weights matrix - high Gini and low Moran's I — apparent agglomeration does not sprawl over the territory and is located only in a single region - high Gini and high Moran's I — sectoral concentration appears and is present in neighbouring regions - low Gini and high Moran's I — there are some slight spatial clusters, but the sectoral concentration is not strong - low Gini and low Moran's I — proves the uniform or even distribution of activity over the territory
Overall concentration	empirical	Geographic Concentration index	0 for no geographic concentration (full geographic diversification) 1 for full geographic concentration (no diversification)
	uniform	Overall Theil's index	Theil=0 for even distribution Theil=max for full concentration Overall Theil — the gap to full diversification, the higher the value, the larger the gap Proportion of regional and inter-regional components shows the source of concentration
Single "cell"	empirical	Location Quotient (LQ)	LQ> 1 concentration of activity in the region LQ> 1.25 potential exporter LQ< 1 underrepresentation of activity in the region, potential importers

Details of calculating the above measures in the Appendix.

Please cite this article in press as: Kopczewska K., Cluster-based measures of regional concentration. Critical overview. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.07.008.

Table 3
Summary of the experimental design.
Source: Own concept.

Distribution	Type of sampling	"Large" table $m = 25$ regions, $n = 100$ sectors	"Middle" table $m = 10$ regions, $n = 25$ sectors	"Small" table $m = 5$ regions, $n = 5$ sectors
		drawn $t = 1000$ iterations (separate data tables) of 25·100=2500 cells	aggregation of the "large" table, irregular groups of regions and sectors	aggregation of the "middle" table, irregular groups of regions and sectors
$N_1(\mu\in(0,50)$, $\sigma\in(0,25))$	Double randomisation	Scheme 1	Scheme 7	Scheme 13
$N_1(\mu\in(0,50)$, $\sigma\in(0,25))$	Single randomisation	Scheme 2	Scheme 8	Scheme 14
$N_1(\mu\in(0,50)$, $\sigma\in(0,25))$	Permutation	Scheme 3	Scheme 9	Scheme 15
$N_2(\mu\in(0,500)$, $\sigma\in(0,100))$	Double randomisation	Scheme 4	Scheme 10	Scheme 16
$N_2(\mu\in(0,500)$, $\sigma\in(0,100))$	Single randomisation	Scheme 5	Scheme 11	Scheme 17
$N_2(\mu\in(0,500)$, $\sigma\in(0,100))$	Permutation	Scheme 6	Scheme 12	Scheme 18

The above hypotheses will be verified on the basis of the Monte Carlo simulation experiment. Each scenario has three parameters: the size of the table, the parameters the of distribution, and the type of sampling. By using the different schemes of drawing the random numbers and sampling, one obtains for each scenario $t = 1000$ iterated sets of the input data tables (as in Table 1) of different sizes (n sectors and m regions) and data properties. For those tables, the full set (k) of measures for each of $t = 1000$ iterations for every scenario is calculated.[11] Computationally, there are $k = 13$ measures of sectoral concentration in the result matrix SC_d of size [t iterations, $13{\cdot}m$ regions] and $k = 9$ measures of geographical concentration in the result matrix GC of size [t iterations, $9{\cdot}n$ sectors]. The experiment assumes $D=1,\ldots d$ designs, which will generate d SC matrices and d GC matrices.

Randomised data to the input data tables (as in Table 1) are based on few patterns. The Data Generating Process (DGP)[12] is always the normal distribution, but its parameters μ and σ are drawn from the uniform distribution under two scenarios: $N_1(\mu\in(0,50),\sigma\in(0,25))$ and $N_2(\mu\in(0,500)$, $\sigma\in(0,100))$. The parameters of the distributions were selected arbitrarily, but in agreement with the parameters observed in real-data samples.[13] $N_2(\mu, \sigma)$ has a larger scale and dispersion of data than $N_1(\mu, \sigma)$, which makes it possible to test the second hypothesis on the importance of the data scale and the variability.[14] Data are generated for a "large" table with $m = 25$ regions and $n = 100$ sectors (2500 cells) and further aggregated to a "middle" table with $m = 10$ regions and $n = 25$ sectors (250 cells) and to a "small" table with $m = 5$ regions and $n=5$ sectors (25 cells), which makes it possible to test the first hypothesis on the MAUP.[15] There are three scenarios of randomisation: (i) double randomisation,

11 The number of iterations (replications) $t = 1000$ is a standard in the case of spatial Monte Carlo simulations (Arbia et al., 2013; Santi et al., 2017).

12 Random numbers were generated with the Mersenne — 'Twister' algorithm, which is suitable for the Monte Carlo simulations and commonly accepted ((Cassey and Smith, 2014). All computations have been made in $R \times 64$ 3.3.3.

13 Regional sectoral statistics for employment and number of firms were used as structural benchmarks in setting the parameters of the distributions. A similar scale and dispersion as proposed in the N_1 and N_2 distributions can be observed, respectively, in the cases of NUTS4/NUTS5 or NUTS2 data.

14 Two-dimensional tables by sectors and territories with generated random data inside were tested for multivariate normality using the Henze–Zirkler's Multivariate Normality Test. Large tables, generated with the DGP, were all multivariate normal, while aggregated tables are not, mainly because of the non-equal aggregation in clusters of different size. Thus, the assumption for the location models on the multivariate normality is fulfilled (Krzanowski, 2000).

15 Openshaw and Baxter (1977) present an algorithm for zoning and aggregation effects to generate random zoning systems, which is based on a few core regions to which adjacent non-core regions are clustered. Holt et al. (1996) give the rules of the

in which for each iteration, both the parameters of the normal distribution (μ, σ) as well as the values from the normal distribution are being drawn; (ii) single randomisation, in which the parameters (μ, σ) of the "cell" are drawn once, and then, each iteration is to draw values from the normal distribution for these fixed parameters (μ, σ); (iii) permutation, whereby both the parameters (μ, σ) and the values from the normal distribution are drawn once, and then each iteration is to permutate the "cells" completely over the whole table. Randomisation scenarios differ in the degree of control over the DGP, from the lowest in double randomisation to the highest in permutation. Additionally, the permutation allows for a counterfactual study, and it is in fact the equivalent of Pseudo Statistical Areas that allow for studying the zoning effect. It is also to test whether the measures are sensitive to changes in the location pattern and structural shifts. This experimental design is to create $t = 1000$ matrices each for 18 schemes (see Table 3). This process in total creates 18,000 matrices with a total of 15 mln "cell" drawings for "large" tables and further aggregation[16] for the "middle" and "small" tables. For all of those 18,000 input data tables one calculates 13 SC measures and 9 GC measures (multiplied by the number of regions and sectors).[17] These results will be summarised as described below.[18]

The obtained matrices of the SC and GC measures will be used in verifying the six working hypotheses: *H1* — sensitivity to MAUP and data aggregation, *H2* — sensitivity of scale and variation of data, *H3* — location invariant measures, *H4* — other "real data" intervals for measures, *H5* — information capacity in groups, and *H6* — absolute measures independent of other factors. They are to be compared in different cross-sections with five quantitative methods (A–E) (see Fig. 1). For the purpose of simplifying the description, the specification below applies to the SC measures only, and exactly the same procedure will be applied to the GC measures.

Methods used in this research:

METHOD A: Using the Kolmogorov–Smirnoff test[19] for checking the similarity of the density of the distributions of the SC measures depending on the size of the input table (the degree of aggregation, which reflects the number of regions and sectors used) or parameters used (different data scale and variation), while keeping the same sampling scheme. This is to verify the existence of the MAUP (*H1*), the impact of scale and variation of the data (*H2*), as well as inter-regional and inter-sectoral shifts for fixed global volume *(H3)*. The Kolmogorov–Smirnoff test for a similarity of the distributions of $t = 1000$ measures is run for different sets: H1 is verified by comparing the schemes: (1,7,13), (2,8,14), (3,9,15), (4,10,16), (5,11,17), (6,12,18), thus to control the effect of aggregation while keeping the parameters of the distribution and sampling design fixed; *H2* is checked by comparisons of schemes (1,4), (2,5), and consequently (7,10), (8,11), and (13,16), (14,17), thus to control the effect of the changing parameters of the distribution while keeping the aggregation level and sampling design fixed; *H3*, which supplements *H2*, is tested by performing comparisons of the remaining schemes for

random aggregation, proving that the means, as well as the regression and correlation coefficients stay unchanged, while the variation increases with the aggregation and smaller number of units.

16 There was also a trial to aggregate the regions and sectors in a regular way (equal groups) but this approach makes the aggregated "cells" very similar, which (a) is far from reality, (b) limits the possibility of testing the sixth hypothesis on the independence of the measures and region's size. This also proves the sensitivity of these measures for zoning.

17 Despite being listed in Section 1, the analysis does not include the Lilien, Herfindahl, Absolute Diversity, Ellison–Glaeser, Maurell–Sedillot, Bergstrand and Guillain–LeGallo indices, and they contain new information in addition to what is given in Table 1 (as data for a different period, a firm's size distribution, distances among regions and a spatial allocation), and thus, they require a different analytical approach.

18 Menon (2012) uses the permutation and scaling procedures in the Monte Carlo simulation to filter-out the industry-specific "noise" from the industrial concentration measures, which captures "lumpiness" and geographical bias, respectively. However, he uses data on a number of plants and, thus studies the Ellison–Glaeser and Herfindahl as well as the Krugman and locational Gini indices.

19 The Kolmogorov–Smirnoff test compares two distributions of variables by setting one as a benchmark and dividing its values into intervals (equal length or equal size). The second variable is checked to have the same distribution in the selected k intervals. The test statistic (product of the supremum of the differences of the cumulated probabilities $F_{ni}(x)$ and the weighted number of observations n_1, n_2) is $Chi^2(k\text{-}1)$.

$$\chi^2\,(k-1) = \sup|F_{n1}\,(x) - F_{n2}(x)| \cdot \sqrt{\frac{n_1 \cdot n_2}{(n_1 + n_2)}}.$$

ARTICLE IN PRESS

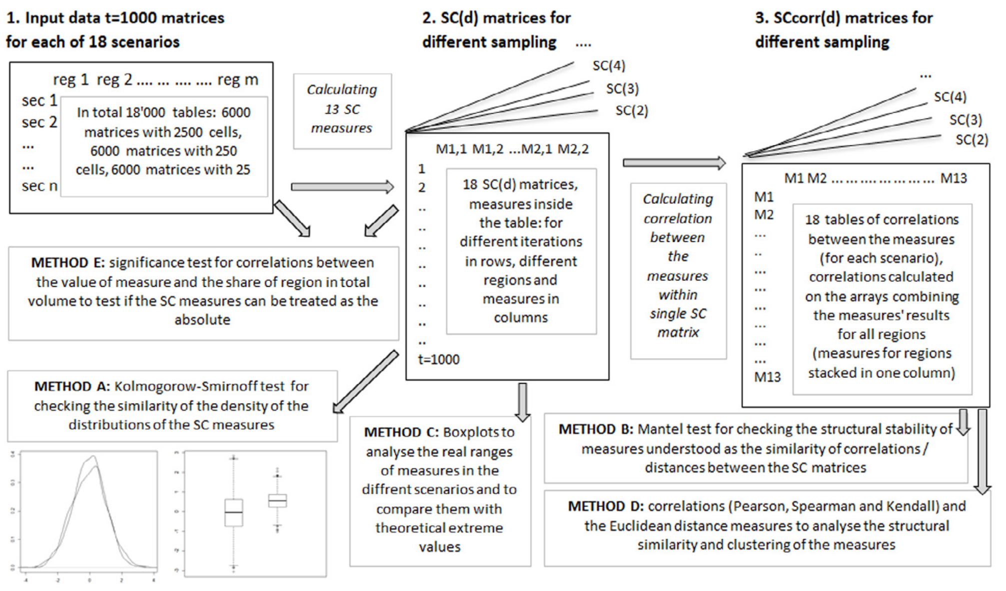


Fig. 1. The design of the study for sectoral and geographical concentration measures. Figure notation refers to SC measures. The symmetric solution is applied to the GC measures.
Source: Own concept.

the same size of tables (3,6), (9,12) and (15,18), which is to extract single differentiating factors — structural changes, and avoid the cross-influence of aggregation or sampling design. In all of the tests, it is assumed that when the distributions are similar, the differentiating factor is not significant. In the literature, Holt et al. (1996) control for the MAUP (aggregation effect) by measuring the ratio of the variances for different scales of aggregation. Flowerdew (2011) hypothesised that the scale of the data could impact the MAUP, but no impact was identified.[20]

METHOD B: Using the Mantel test for checking the structural stability of the measures understood as the similarity of correlations/distances between the SC matrices. This approach is to track the structural changes between the SC measures under different sampling schemes on the basis of a two-step procedure. First, the correlations/Euclidean distances between *13* measure results (stacked for all regions) for a given SC_d matrix are computed (on the basis of $t = 1000$ arrays times m regions), which gives the $SCcorr''\text{-}_d$ matrix. For a single SC_d matrix, there are $(13)^2$ individual correlation/distances results, which are denoted as $cor^{measurek}_{matrixd} = cor^k_d$. Second, the similarity of the correlation/distances matrices $SCcorr_d$ is assessed. For the hypothesis that the matrices $SCcorr_{d1}$ and $SCcorr_{d2}$ are similar, one can apply the Mantel test,[21] and for the hypothesis that the matrices $SCcorr_{d1}$ and $SCcorr_{d2}$ and $SCcorr_{d3}$ are similar, one can apply the partial Mantel test. This method is applied to test *H1* by comparing the schemes: (1,7,13), (2,8,14), (3,9,15), (4,10,16), (5,11,17), (6,12,18) with the partial Mantel test; to test *H2* by comparing the pairs (1,4), (2,5), and consequently (7,10), (8,11), and (13,16), (14,17) with the Mantel test, and to test *H3* by comparing the schemes (3,6), (9,12) and (15,18).

20 *"It could be argued that the variables with relatively high values might be more likely to be stable in comparison to those representing rare occurrences. It might also be expected that variables with a high standard deviation might be prone to the MAUP, because it could be that the pattern picked out at one scale would be lost through aggregation".*

21 The Mantel test is to compute a correlation between square matrices ($n \times n$). It is mostly applied to matrices of distances or other measures of similarity. The H_0 is that the relationship between the two matrices can result from a random process of allocation (no relationship in location/distance/similarity). The Mantel statistic is calculated as a sum of the cross products of the two tested similarity matrices. The Mantel test, in contrast to Pearson, addresses the autocorrelation within the distance measures, and it avoids bias.

Please cite this article in press as: Kopczewska K., Cluster-based measures of regional concentration. Critical overview. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.07.008.

METHOD C: Using boxplots to analyse the real ranges of measures. This approach is to check whether the real intervals of the measures' values are far from their theoretical extremes (*H4*). This results in a comparison of the given measure under all 18 schemes, controlling for the sizes of the tables. The assumption is that for very diverse input data (as in scenarios 1–18), the SC measures should reach their upper and lower extremes. Practically achievable limits strongly impact the interpretation of the results.

METHOD D: Using the correlation (Pearson, Spearman and Kendall) and the Euclidean distance measures to analyse the structural similarity and clustering of the measures (*H5*).[22] This analysis will be performed for all SC_d matrices to find the robust grouping — information clusters, where the measures behave within them. In the literature, Briant et al. (2010) use rank correlations to test the impacts of the zoning effect. The correlations are calculated between the measures, for the arrays that include all the stacked results (by regions) in a given scenario.

METHOD E: Using the significance test for the correlations to test if the SC measures can be treated as absolute *(H6)*. The hypothesis of independence of other factors is verified by correlating the values of the measures with the other characteristics of the regions, such as the share of the region in the overall volume (region size). The significance of the correlation (Pearson, Spearman and Kendall) between the all of the values of the measures and a share of the region in the total volume will be tested.

3. Results

The descriptive analysis of the results reveals the range of the measures' variability (*H4*) (see Figs. 2A and 2B and Tables 4A and 4B). Many empirical realisations of the measures are far from their extreme values (see Tables 4A and 4B), but in addition, for many, there are no theoretical maximum values established. The closest to the theoretical range are the entropy measures and locational Gini, while the values of the classic Gini and Hallet indices are placed in a narrow corridor that is far from the extreme theoretical benchmarks. These statistics prove a diversified variation of the measure's values. The coefficient of variation (*sd/mean*) divides the SC measures into three groups: low variance (NAI, Shannon's H, Relative H, Refined Diversification, Hachman, RSI (max.LQ)), middle variance (Ogive, Theil's H, Gini) and high variance (Krugman, Hallet, RDI, KLD). All of the GC measures present high variance. The higher the variation, the greater the potential for significant value changes and sensitivity to pattern changes in the data. Low-variance measures behave very similarly under data changes and in fact react poorly for this test.

There is also a visible bias (following Deltas (2003) methodology) or MAUP sensitivity (following this paper's hypotheses). The values of the measures predominantly decrease (or increase) in line with the aggregation, and the changes are highly monotonic. In most of the cases, the measures' values are a few times higher for the larger tables (less aggregated) than for the smaller ones. This MAUP reaction refers to the central tendency (mean) values as well as to the diversity of measures. The boxplots in Figs. 2A & 2B prove that the measures' values depend on the aggregation scheme. In most cases, the deep aggregation and usage of the small table mostly changes the range of the measures. Referring to Delta's (2003) findings on a small-sample bias of the Gini index, one should distinguish between a small sample and a small number of units (e.g., because of aggregation). This simulation shows that for the same sample size, but a diminishing number of units (due to the dataset aggregation), the Gini also decreases. Even if the mechanism of occurring bias is different, the conclusions are similar on the limited comparability of measures calculated on the different data. The correction of the two-dimensional Gini, as proposed by Deltas (2003) for the one-dimensional Gini, is hardly feasible, because of a further and more complicated relationship of the two-dimensional dataset parameters.

To verify *H5* on the information clusters and group behaviour, the Pearson correlation matrices in a group of the SC and GC measures (see Tables 5A and 5B) were studied. They reveal the

[22] The analytical work suggested the analysis of the Pearson correlations only, even if other methods were implemented. Principal Component Analysis was considered, but it gave an unstable classification, because of having very high correlations. The Spearman and Kendall correlation statistics were calculated, but they gave very similar results to the Pearson coefficient.

Table 4A
Ranges of the SC measures for three types of aggregation: large $n = 100$/middle $n = 25$/small table $n = 5$ (all scenarios. included)

Source: Own calculations.

Measure	Theoretical minimum	Theoretical maximum	Empirical minimum	Empirical maximum	Average value	Standard deviation	Coefficient of variation
Ogive	0/0/0	Not defined	0.23/0.40/0.33	0.86/1.33/0.92	0.48/0.71/0.58	0.07/0.10/0.07	0.15/0.14/0.11
Krugman Dissim.	0/0/0	Not defined	0.39/0.35/0.00	0.79/0.70/0.18	0.58/0.47/0.05	0.05/0.05/0.02	0.08/0.10/0.52
NAI	0/0/0	Not defined	1.19/0.88/0.65	1.79/1.22/0.85	1.44/0.96/0.73	0.07/0.05/0.02	0.05/0.05/0.03
Hallet	0/0/0	0.5/0.5/0.5	0.20/0.18/0.00	0.39/0.35/0.09	0.29/0.23/0.02	0.02/0.02/0.01	0.08/0.10/0.52
RDI (1/Krugman)	0.5	$+\infty$	1.27/1.44/5.54	2.56/2.84/221.10	1.74/2.15/29.60	0.14/0.21/19.00	0.08/0.10/0.64
Shannon	0/0/0	ln(100) / ln(25) / ln(5)	4.18/2.66/1.25	4.47/3.03/1.47	4.34/2.90/1.37	0.04/0.04/0.02	0.01/0.01/0.02
RelativeH	0/0/0	1/1/1	0.91/0.83/0.77	0.97/0.94/0.91	0.94/0.90/0.85	0.01/0.01/0.02	0.01/0.01/0.02
Theil	0/0/0	ln(100) / ln(25) / ln(5)	0.14/0.19/0.14	0.42/0.56/0.36	0.26/0.32/0.24	0.04/0.04/0.02	0.13/0.13/0.11
KLD	0/0/0	Not defined	0.04/0.00/0.00	0.41/1.16/0.66	0.23/0.54/0.25	0.05/0.27/0.18	0.21/0.49/0.70
Refined Divers.	0/0/0	Not defined	1.45/1.26/1.19	1.65/1.40/1.28	1.54/1.34/1.23	0.03/0.02/0.01	0.02/0.01/0.01
Hachman	0/0/0	1/1/1	0.51/0.75/0.96	0.82/0.99/1.00	0.67/0.95/1.00	0.04/0.04/0.00	0.06/0.04/0.00
RSI (max LQ)	0/0/0	Not defined	2.02/1.14/1.01	8.24/4.67/1.45	3.21/1.73/1.10	0.53/0.38/0.06	0.17/0.22/0.06
Gini	0/0/0	1/1/1	0.39/0.06/0.16	0.69/0.58/0.40	0.55/0.24/0.23	0.04/0.09/0.02	0.07/0.38/0.08

Please cite this article in press as: Kopczewska K., Cluster-based measures of regional concentration. Critical overview. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.07.008.

Table 4B
Ranges of the GC measures for three types of aggregation: large $m = 25$/middle $m = 10$/small table $m = 5$ (all scenarios included).

Source: Own calculations.

Measure	Theoretical minimum	Theoretical maximum	Empirical minimum	Empirical maximum	Average value	Standard deviation	Coefficient of variation
Gini	0/0/0	1/1/1	0.27/0.05/0.01	0.85/0.80/0.28	0.57/0.29/0.07	0.08/0.12/0.04	0.14/0.41/0.53
Krugman Concen.	0/0/0	Not defined	0.27/0.05/0.00	0.87/0.78/0.17	0.57/0.29/0.05	0.08/0.12/0.03	0.14/0.41/0.52
Bruelhart–Traeger	0/0/0	Not defined	0.26/0.03/0.00	0.84/0.65/0.01	0.57/0.16/0.00	0.08/0.08/0.00	0.14/0.46/0.95
Locational Gini	0/0/0	1/1/1	0.24/0.04/0.09	0.88/0.67/0.20	0.57/0.23/0.14	0.08/0.09/0.01	0.14/0.41/0.09
AgglomerationV	0/0/0	Not defined	0.25/0.03/0.01	0.93/0.66/0.51	0.57/0.17/0.14	0.09/0.10/0.07	0.16/0.59/0.52
Shannon	0/0/0	ln(25) / ln(10) / ln(5)	0.24/0.00/1.41	0.99/0.14/1.56	0.57/0.02/1.49	0.09/0.02/0.02	0.16/0.81/0.01
RelativeH	0/0/0	1/1/1	0.25/0.00/0.88	0.95/0.14/0.97	0.57/0.02/0.92	0.09/0.02/0.01	0.16/0.81/0.01
Theil	0/0/0	ln(25) / ln(10) / ln(5)	0.23/0.00/0.05	0.94/0.06/0.19	0.57/0.02/0.12	0.09/0.01/0.02	0.16/0.84/0.14
KLD	0/0/0	Not defined	0.02/0.01/0.05	0.25/0.06/0.19	0.11/0.03/0.12	0.03/0.01/0.02	0.27/0.19/0.14

Please cite this article in press as: Kopczewska K., Cluster-based measures of regional concentration. Critical overview. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.07.008.

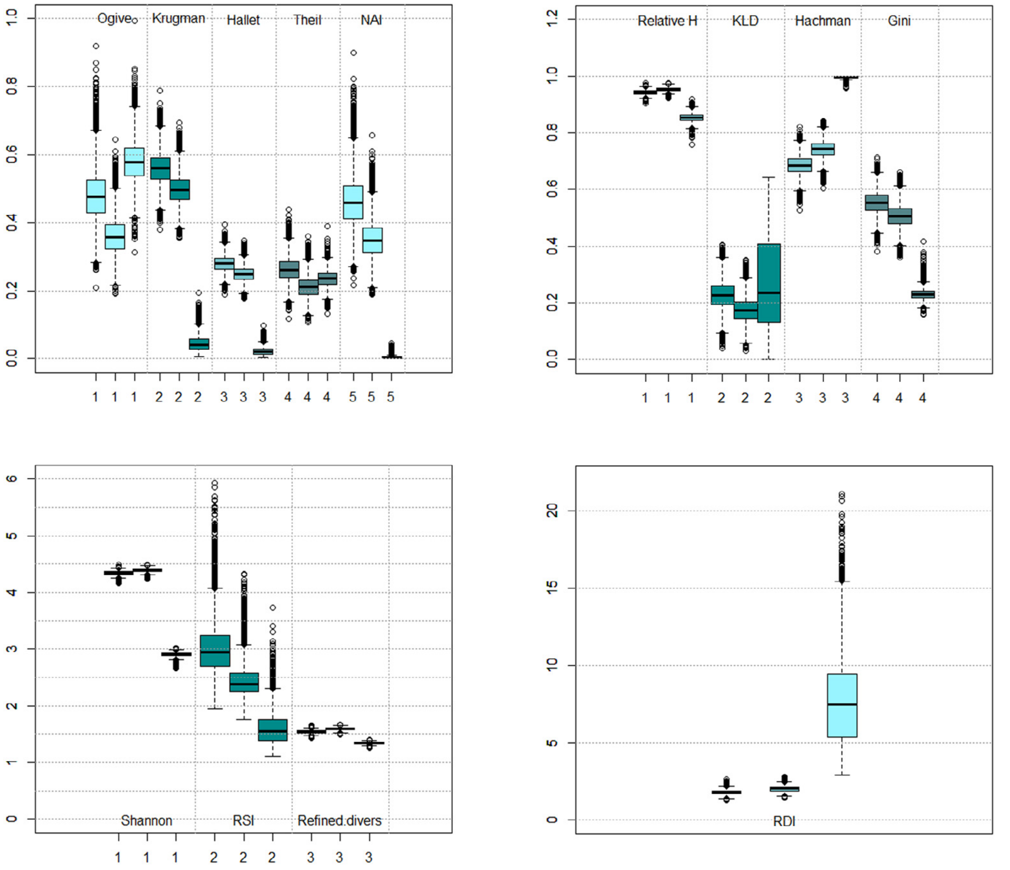


Fig. 2A. Boxplots of the SC measures for large, middle and small tables (all scenarios included).
Source: Own analysis in the R software.

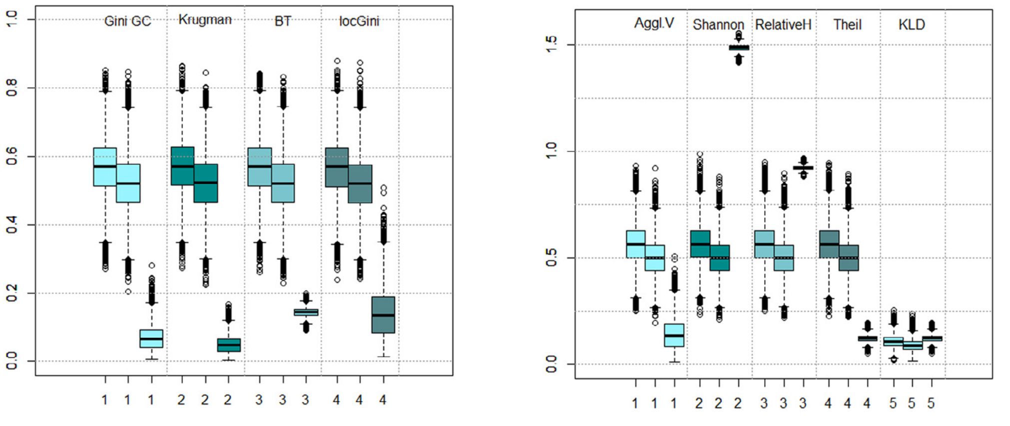


Fig. 2B. Boxplots of the GC measures for large, middle and small tables (all scenarios included).
Source: Own analysis in the R software.

pattern of behaviour of the measures depending on the drawing schemes and aggregation levels as well as their joint relations. The set of SC measures is mostly insensitive for the data scale (a slight difference between the distributions N1 and N2 — right and left columns), but they react strongly to the aggregation process (upper, middle and lower rows) (MAUP discussed in *H1*). The measures graphed in Tables 5A and 5B were sorted and clustered by the groups of the reference distribution (no 1: theoretical one, no 2: empirical one, no 3: no regular distribution, as discussed in Section 1).

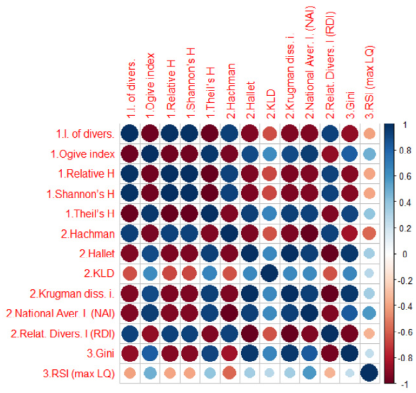
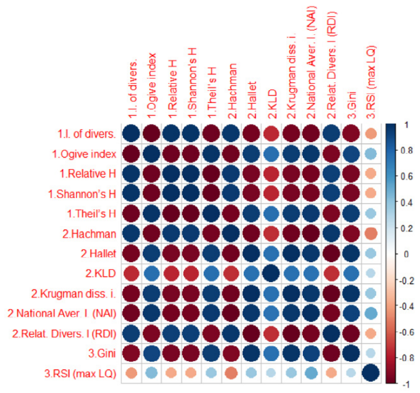
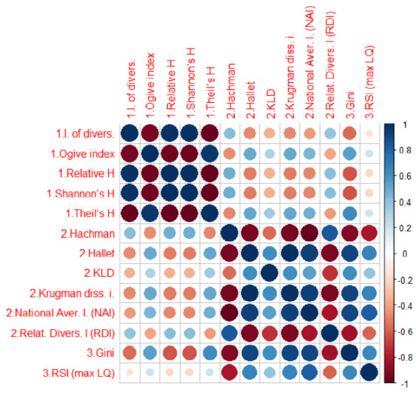
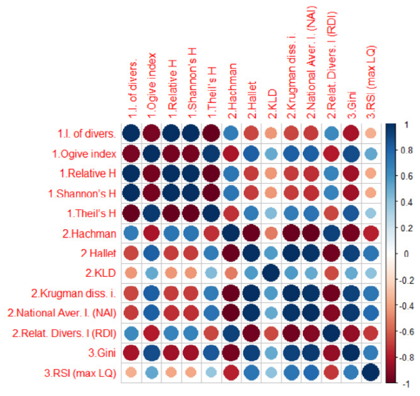
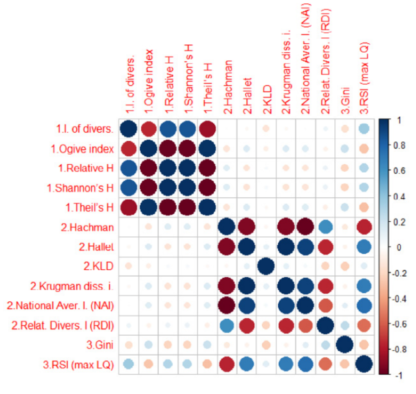
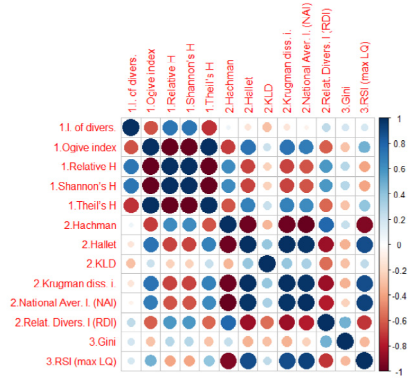
Among the SC measures (Table 5A), for the data generated randomly (large tables) the correlations are strong between all of the measures, except for the RSI (max. LQ), which provides different information. Strong fragmentation of data, which masks the specifics of the concentration patterns, makes all of the measures become insensitive and, thus, similar to one another, which can explain the strong correlations. The aggregation of data in an irregular manner (regions and industries clustered in random size groups) proves that the stronger the skewness of the data is after the aggregation, the stronger the pattern of dependence between the measures. Entropy measures (Shannon's H, Theil's H, Relative H) together with a Refined Diversification and the Ogive indices build a cluster (group no 1) based on the theoretical reference distribution, which presents a strong and significant correlation in every scenario. The second group (no 2), which consists of the RDI, NAI, KLD, Hallet, Hachman and Krugman measures is based on an empirical distribution and reveals the inner correlation, which weakens together with the aggregation. Measures such as Gini, KLD and RSI (max.LQ) have their own patterns of variation, as their construction is less straightforward than in the others. Among the GC measures (Table 5B), the correlations in the large tables are rather rare, while the stronger aggregation enhances the relations, which become very visible in the pre-defined groups (because of the underlying distribution) in the small tables. This finding confirms *H5* on the similarity of behaviour in the groups of measures. The conclusion is that for highly aggregated data, studies should be based on measures from all of the groups to ensure a comprehensive interpretation. For both the SC and GC measures, the stronger the aggregation is, the more visible the relations between the measures. The coherent results when using representative measures can be obtained for the highest degree of dataset aggregation.

These findings also refer to the literature on Pearson's correlation depending on the aggregation level. Following the study by Reynolds (1998), Gehlke and Biehl (1934), Robinson (1950) and Clark and Avery (1976) claim that Pearson is conditioned upon the number of units and the correlations are higher for a higher number of spatial units. However, the aforementioned studies are based on empirical data and study the correlations between observations, not measures, and do not consider the significance of Pearson's *r*. This study (see Tables 5A and 5B) shows that the relation of correlation is not obvious. The GC measures correlate in a different manner than the SC measures, and both *r* and its significance change as a function of the aggregation. These results are nevertheless in line with previous findings, which indicates that correlations for the same data but differently aggregated are incomparable.

The findings from the graphical analysis of the Pearson correlation matrices between the measures were statistically confirmed. The Mantel test for both the Pearson correlation coefficient as well as the Euclidean distance between the measures was performed in groups of three (large, middle and small) and two matrices (the extreme ones: large and small only) (see Tables 6A and 6B). The Mantel tests for the Pearson coefficients[23] for two and three matrices are statistically significant, which suggests that all SC correlation matrices are similar regardless of the aggregation level. Conversely, the GC measures behave differently. However, there is a visible trend of lowering the Mantel statistics for the extreme pairs, which indicates that the links weaken after aggregation. The Euclidean distances are more sensitive, and the Mantel tests are statistically significant at $\alpha = 0.1$ for triples and insignificant for pairs. This finding proves that the progressive data aggregation gradually weakens the links. *Ipso facto*, this proves the *H1* hypothesis about the sensitivity of measures to the MAUP. Thus, one cannot expect that the same data on a different aggregation level would give the same results.

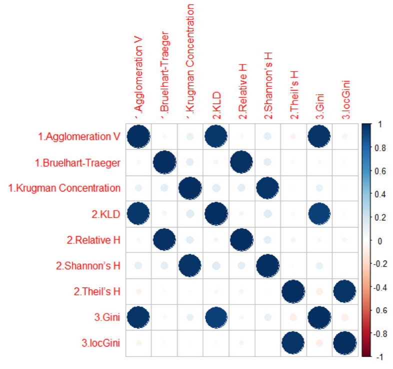
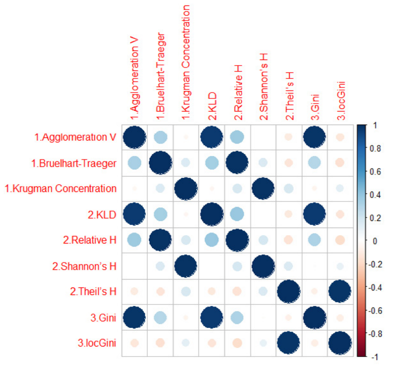
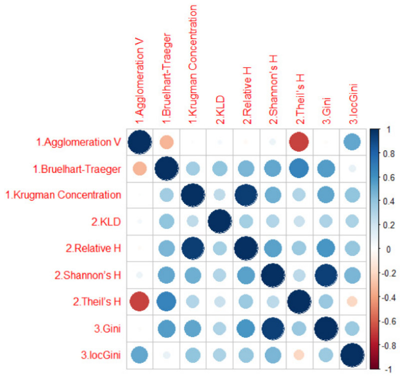
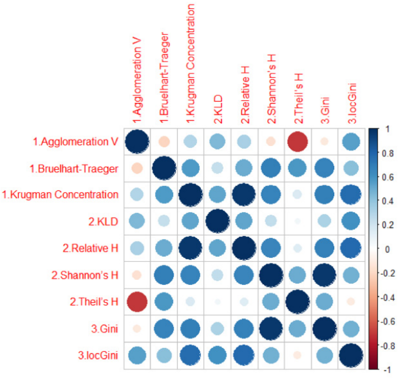
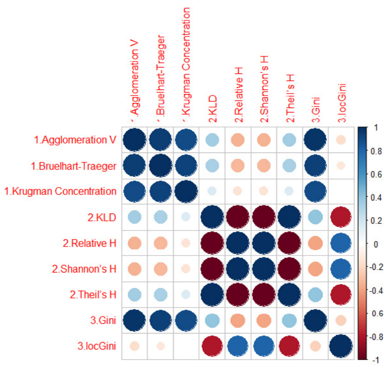
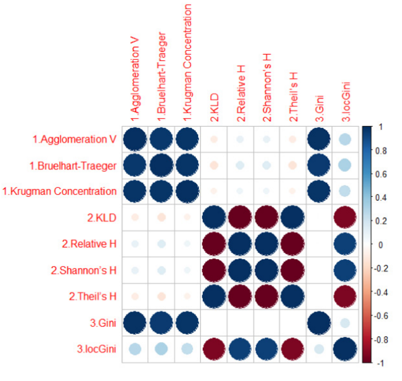
[23] In the Mantel test, H_0 assumes that the compared matrices are random and that they are different, while H_1 assumes the non-random pattern, which is interpreted as a similarity of the matrices.

Table 5A
Correlation of the SC measures — results for single randomised matrices (scenarios 2,5,8,11,14,17).

	N1	N2
Large table		
Middle table		
Small table		

Groups of measures: no 1 – theoretical benchmark distribution, no 2 – empirical benchmark distribution, no 3 – transformed empirical distribution or no benchmark distribution; Size of dots: strength of correlation (the larger the dot is, the stronger the relation); colour of dots: according to the bar scale, strength and direction of correlation.
Source: Own analysis in the R software using the *corrplot* package (Wei & Simko, 2016).

Table 5B
Correlations of the GC measures — results for single randomised matrices (scenarios 2,5,8,11,14,17).

<table>
<tr><th></th><th>N1</th><th>N2</th></tr>
<tr><td>Large table</td><td>
</td><td>
</td></tr>
<tr><td>Middle table</td><td>
</td><td>
</td></tr>
<tr><td>Small table</td><td>
</td><td>
</td></tr>
</table>

Groups of measures: no 1 – theoretical benchmark distribution, no 2 – empirical benchmark distribution, no 3 – transformed empirical distribution or no benchmark distribution; Size of dots: strength of correlation (the larger the dot is, the stronger the relation); colour of dots: according to the bar scale, strength and direction of correlation.

Source: Own analysis in the R software with the *corrplot* package (Wei & Simko, 2016).

Table 6A
Mantel test for the Pearson correlation and Euclidean distance – testing the aggregation factor (drawing scenario and data scale the same) – SC measures.
Source: Own calculations in the R software with the *vegan* package (Oksanen et al., 2017).

Hypothesis	Triple comparison of matrices[a] (large, middle, small) for the same drawing scheme	Partial Mantel test (and p-value) for the Pearson correlation	Partial Mantel test (and p-value) for the Euclidean distance	Dual comparison of extreme matrices (large, small)	Mantel test (and p-value) for the Pearson correlation	Mantel test (and p-value) for the Euclidean distance
H1 MAUP	N1.DR.(1,7,13)	0.945 (0.001)	0.560 (0.085)	N1.DR.(1,13)	0.574 (0.002)	−0.015 (0.150)
H1 MAUP	N1.SR.(2,8,14)	0.947 (0.001)	0.566 (0.087)	N1.SR.(2,14)	0.648 (0.001)	−0.022 (0.146)
H1 MAUP	N1.P.(3,9,15)	0.945 (0.001)	0.576 (0.086)	N1.P.(3,15)	0.614 (0.001)	−0.019 (0.153)
H1 MAUP	N2.DR.(4,10,16)	0.944 (0.001)	0.396 (0.096)	N2.DR.(4,16)	0.546 (0.001)	0.037 (0.155)
H1 MAUP	N2.SR.(5,11,17)	0.788 (0.001)	0.373 (0.101)	N2.SR.(5,17)	0.248 (0.029)	0.055 (0.158)
H1 MAUP	N2.P.(6,12,18)	0.944 (0.001)	0.397 (0.086)	N2.P.(6,18)	0.543 (0.001)	0.028 (0.159)

[a] Symbols in matrices: Large ($m = 25$ regions, $n = 100$ sectors), Middle ($m = 10$ regions, $n = 25$ sectors), Small ($m = 5$ regions, $n = 5$ sectors), DR-Double Randomisation, SR-Single Randomisation, P-Permutation, $N_1(\mu \in (0, 50), \sigma \in (0, 25))$, $N_2(\mu \in (0, 500), \sigma \in (0, 100))$.

Table 6B
Mantel test for the Pearson correlation and Euclidean distance – testing the aggregation factor (drawing scenario and data scale the same) – GC measures.
Source: Own calculations in the R software with the *vegan* package (Oksanen et al., 2017).

Hypothesis	Triple comparison of matrices[a] (large, middle, small) for the same drawing scheme	Partial Mantel test (and p-value) for Pearson correlation	Partial Mantel test (and p-value) for Euclidean distance	Dual comparison of extreme matrices (large, small)	Mantel test (and p-value) for Pearson correlation	Mantel test (and p-value) for Euclidean distance
H1 MAUP	N1.DR.(1,7,13)	−0.087 (0.740)	−0.17195 (0.743)	N1.DR.(1,13)	−0.0003 (0.429)	−0.210 (0.984)
H1 MAUP	N1.SR.(2,8,14)	−0.147 (0.846)	−0.18085 (0.859)	N1.SR.(2,14)	−0.003 (0.431)	−0.220 (0.995)
H1 MAUP	N1.P.(3,9,15)	−0.089 (0.719)	−0.17225 (0.765)	N1.P.(3,15)	−0.006 (0.424)	−0.210 (0.986)
H1 MAUP	N2.DR.(4,10,16)	−0.093 (0.736)	−0.18122 (0.901)	N2.DR.(4,16)	0.0003 (0.424)	−0.206 (0.936)
H1 MAUP	N2.SR.(5,11,17)	−0.114 (0.772)	−0.19262 (0.981)	N2.SR.(5,17)	−0.057 (0.573)	−0.186 (0.913)
H1 MAUP	N2.P.(6,12,18)	−0.087 (0.704)	−0.18233 (0.924)	N2.P.(6,18)	0.003 (0.431)	−0.205 (0.943)

[a] Symbols in matrices: Large ($m = 25$ regions, $n = 100$ sectors), Middle ($m = 10$ regions, $n = 25$ sectors), Small ($m = 5$ regions, $n = 5$ sectors), DR-Double Randomisation, SR-Single Randomisation, P-Permutation, $N_1(\mu \in (0, 50), \sigma \in (0, 25))$, $N_2(\mu \in (0, 500), \sigma \in (0, 100))$.

The Mantel test was also used to test *H2* about the sensitivity of the scale and the variation of the input data and *H3* about the location-invariant measures. The pair comparisons of the SC (and separately GC) matrices in groups of size and type of drawing, with different data scales (N1 and N2) (Tables 7A and 7B) prove that the Pearson correlations and the Euclidean distances are structurally the same, and thus, the measures do not react to changes in the volume and differentiation. The

ARTICLE IN PRESS

Table 7A
Mantel test for the Pearson correlation and Euclidean distance –testing the H2 scale data factor (drawing scenario and aggregation the same) and H3 permutation factor (drawing scenario and aggregation the same) – SC measures.
Source: Own calculations in R software with *vegan* package (Oksanen et al., 2017).

Hypothesis	Dual comparison of matrices[a]	Mantel test (and *p*-value) for Pearson correlation between SC measures	Mantel test (and *p*-value) for Euclidean distance between SC measures
H2 scale of data	Big.DR.N1-N2(1,4)	0.999 (0.001)	0.970 (0.001)
H2 scale of data	Big.SR.N1-N2(2,5)	0.998 (0.001)	0.954 (0.001)
H3 location invariant	Big.P.N1-N2(3,6)	0.999 (0.001)	0.965 (0.001)
H2 scale of data	Middle.DR.N1-N2(7,10)	0.999 (0.001)	0.999 (0.001)
H2 scale of data	Middle.SR.N1-N2(8,11)	0.888 (0.001)	0.999 (0.001)
H3 location invariant	Middle.P.N1-N2(9,12)	0.999 (0.001)	0.999 (0.001)
H2 scale of data	Small.DR.N1-N2(13,16)	0.996 (0.001)	0.999 (0.001)
H2 scale of data	Small.SR.N1-N2.(14,17)	0.657 (0.001)	0.999 (0.001)
H3 location invariant	Small.P.N1-N2.(15,18)	0.994 (0.001)	0.999 (0.001)

[a] Symbols in matrices: Large ($m = 25$ regions, $n = 100$ sectors), Middle ($m = 10$ regions, $n = 25$ sectors), Small ($m = 5$ regions, $n = 5$ sectors), DR-Double Randomisation, SR-Single Randomisation, P-Permutation, $N_1(\mu \in (0, 50), \sigma \in (0, 25))$, $N_2(\mu \in (0, 500), \sigma \in (0, 100))$.

Table 7B
Mantel test for the Pearson correlation and Euclidean distance –testing the H2 scale data factor (drawing scenario and aggregation the same) and H3 permutation factor (drawing scenario and aggregation the same) – GC measures.
Source: Own calculations in R software with *vegan* package (Oksanen et al., 2017).

Hypothesis	Dual comparison of matrices[a]	Mantel test (and *p*-value) for Pearson correlation between GC measures	Mantel test (and *p*-value) for Euclidean distance between GC measures
H2 scale of data	Big.DR.N1-N2(1,4)	0.999 (0.001)	0.999 (0.001)
H2 scale of data	Big.SR.N1-N2(2,5)	0.923 (0.001)	0.923 (0.001)
H3 location invariant	Big.P.N1-N2(3,6)	0.999 (0.002)	0.999 (0.001)
H2 scale of data	Middle.DR.N1.N2(7,10)	0.999 (0.001)	0.999 (0.001)
H2 scale of data	Middle.SR.N1-N2(8,11)	0.866 (0.001)	0.866 (0.001)
H3 location invariant	Middle.P.N1-N2(9,12)	0.999 (0.001)	0.999 (0.001)
H2 scale of data	Small.DR.N1-N2(13,16)	0.999 (0.001)	0.999 (0.001)
H2 scale of data	Small.SR.N1-N2.(14,17)	0.899 (0.001)	0.899 (0.001)
H3 location invariant	Small.P.N1-N2.(15,18)	0.999 (0.001)	0.999 (0.001)

[a] Symbols in matrices: Large ($m = 25$ regions, $n = 100$ sectors), Middle ($m = 10$ regions, $n = 25$ sectors), Small ($m = 5$ regions, $n = 5$ sectors), DR-Double Randomisation, SR-Single Randomisation, P-Permutation, $N_1(\mu \in (0, 50), \sigma \in (0, 25))$, $N_2(\mu \in (0, 500), \sigma \in (0, 100))$.

permutation of the once drawn random numbers confirms that the same data in different locations finally result in similar measures.

The simulation proved that the values of the SC and GC measures are sensitive to the share of the total regional volume in the overall volume (e.g., employment) (see Figs. 3A & 3B). Even if the SC measures explore the sectoral structure inside the region, they are not independent of the region's position among other regions. In other words, two regions with the same internal industrial structure, but with a different share in overall volume will have a different sectoral concentration performance. The same applies to the GC measures — when comparing two industries, with the same inter-regional distribution, the GC measures will differ because of the sizes of the industries. This result is also dependent on the scale of aggregation. In the case of strong disaggregation (detailed data for many regions/sectors), almost all of the SC and GC measures (except for RSI) strongly depend on the position of the region/sector among the others (the rank of the region/sector). Strong aggregation (averaged data in small table) weakens this dependence. Among the SC measures, the most vulnerable are the Krugman dissimilarity, Hallet, Relative Diversity, Hachman, RSI and Gini, while the entropy measures, Ogive, NAI and diversification become independent. Among the GC measures, the entropy and locational Gini also lose this dependence. This finding verifies *H6* negatively, proving the lack of independence of the measures from other factors. Thus, the measures cannot be absolutely compared.

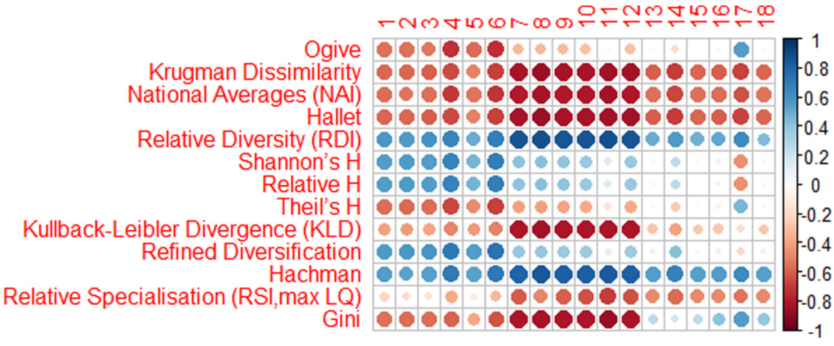


Fig. 3A. Pearson correlation between the SC measures and the share of the region in the overall volume. Columns (1–18) define drawing schemes and the aggregation degree; Size of dots: strength of correlation (the larger the dot is, the stronger the relation); Colour of dots: according to the bar scale, strength and direction of correlation.
Source: Own analysis in the R software with *corrplot* package (Wei and Simko, 2016).

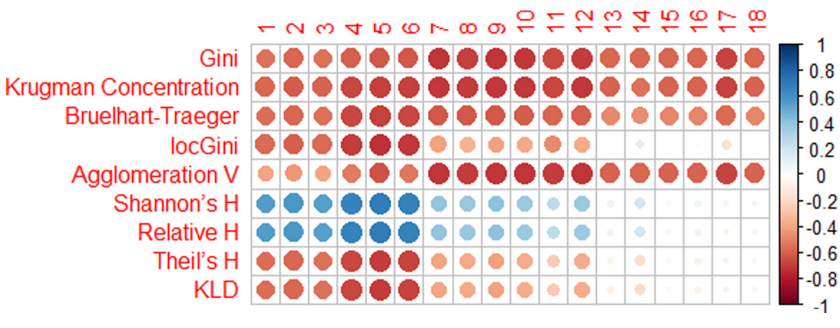


Fig. 3B. Pearson correlation between the GC measures, degree of aggregation and the share of the sector in the overall volume. Columns (1–18) define drawing schemes and aggregation degree; Size of dots: strength of correlation (the larger the dot is, the stronger the relation); Colour of dots: according to the bar scale, strength and direction of correlation.
Source: Own analysis in the R software with *corrplot* package (Wei and Simko, 2016).

The examples in Figs. 4–6 show the relation between the share of the region in the overall volume (x) and the estimates of the given SC and GC measure (the Krugman dissimilarity, SC Gini and GC Theil's H indices) (y). Because of the aggregation, the region's shares in the "large table" (out of 25 regions) are approximately 5 times less than in the "small" table (out of 5 regions). The difference between the possible results of the measures is not to be neglected, because it reaches up to 20% of the measure's range. Measures distributions for different drawing schemes and aggregations (Fig. 7) prove that they are aggregation dependent and MAUP sensitive. This finding verifies *H1* negatively, proving that the comparisons of the concentration measures should refer to the aggregation level. These results are in line with the empirical findings of Dewhurst and McCann (2007, p. 217), who evidenced the decreasing Gini with an increase in the log of the region's size expressed with the total employment, which is very similar to Fig. 5A. Deltas (2003) also finds an approximately 15% bias.

4. Empirical analysis

The results of the above simulations were verified on the empirical data. The hierarchical dataset for a number of the business entities in 2017 in Poland was collected from Polish Statistical Office.[24] There are two levels of branch classification (*sector*, $n_{\mathrm{LARGE}} = 86$ and $n_{\mathrm{SMALL}} = 20$) and two levels

24 A full dataset from the Local Data Bank at www.stat.gov.pl is available online.

Please cite this article in press as: Kopczewska K., Cluster-based measures of regional concentration. Critical overview. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.07.008.

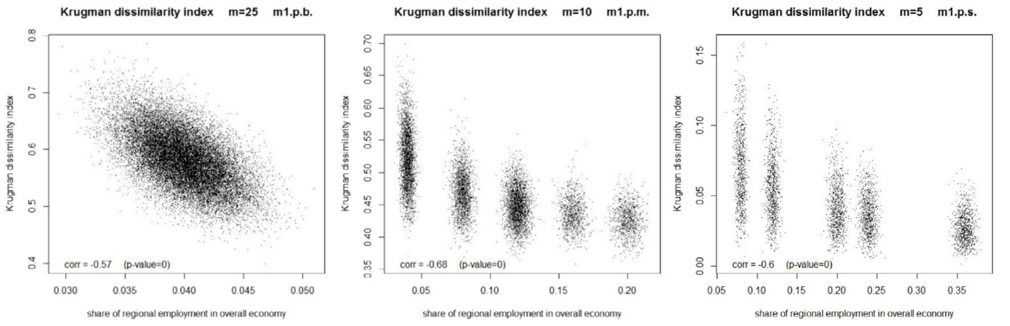


Fig. 4. Individual correlation of the SC Krugman dissimilarity index with the region's share in the overall volume in the (a) large; (b) middle, and (c) small table (schemes 3,9,15).
Source: Own analysis in the R software.

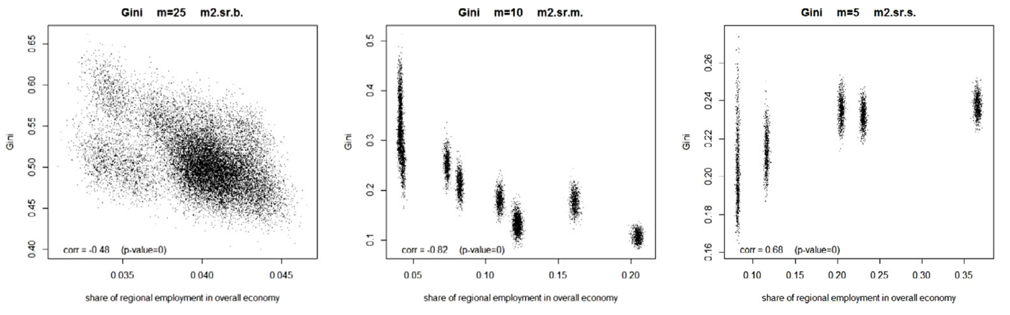


Fig. 5. Individual correlation of the SC Gini measure with the region's share in the overall volume in the (a) large; (b) middle, and (c) small table (schemes 5,11,17).
Source: Own analysis in the R software.

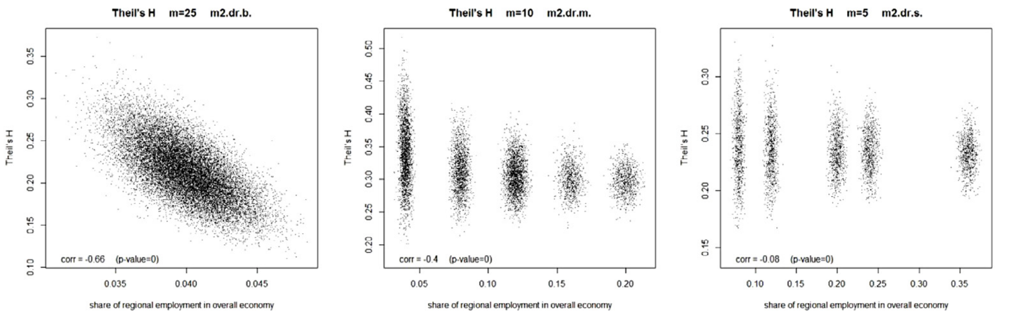


Fig. 6. Individual correlation of the GC Theil's H measure with the sectoral share in the overall volume in the (a) large; (b) middle, and (c) small table (schemes 4,10,16).
Source: Own analysis in the R software.

of administrative division (*region*, $m_{\text{LARGE}} = 380$ and $n_{\text{SMALL}} = 16$). Data on 4231504 business units were aggregated into four tables: *lrls* ($n = 86$, $m = 380$) [*lrls=large regions, large sectors*], *lrss* ($n = 20$, $m = 380$), *srls* ($n = 86$, m=16), *srss* ($n = 20$, $m = 16$).

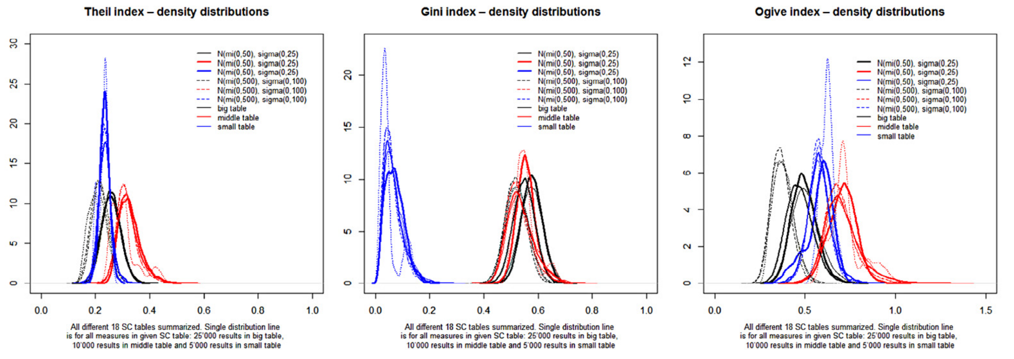


Fig. 7. Distributions of the SC measures for different schemes: (a) Theil, (b) Gini, (c) Ogive (all schemes).
Source: Own analysis in the R software.

The boxplots in Fig. 8 (in the order mentioned above) confirm the previous findings, that deeper aggregation mostly lowers the values of the SC measures. For example, Gini, Krugman, NAI, Hallet, RSI in *lrls*($n = 86, m = 380$) are a few times higher than in *srss*($n = 20, m = 16$) and any shrinking of the dataset (both n as well as m) gradually lowers their result. In some SC measures (Ogive, Shannon's H, KLD, Relative H, Theil's H, Refined Diversification) the effect of shrinking n sectors is much stronger than when shrinking m regions. The findings also confirm that measures reach narrower values than expected theoretically. The same holds for the geographical concentration measures (GC) (see Fig. 9), where the aggregation mostly lowers the measures. The dependence on the size of the region cannot be tested (and confirmed) because the administrative division of Poland by assumption equalises the higher level units in terms of the magnitude of their economic performance, and the required diversity of the size of units is not observed.

5. Conclusions and recommendations

The goal of the paper, except for summarising the majority of the available sectoral and geographical concentration measures, was to test in simulation its statistical behaviour. There were few hypotheses and questions to verify and answer, which are recapitulated below.

First, the sensitivity to the MAUP was confirmed, and this finding impacts the decisions about the aggregation scale as well as the zoning of the dataset. The paper shows that the size of the data table (the number of sectors and regions) does matter for the Sectoral Concentration (SC) and Geographical Concentration (GC) measure values. In a large dataset with a low aggregation level (many regions and sectors), the results of the measures could be similar, and the measures become much more sensitive in smaller (more aggregated) samples. This finding confirms the existence of a sample-size/number of units bias, proved by Deltas (2003).

Second, both the SC and GC measures are insensitive to the scale and the variation of the input data. Thus, the comparison of large and small countries could be legitimate and robust. This result is in line with the findings by Flowerdew (2011). The simulated datasets included "almost-zero" cells as well as extreme concentration, and these activity distributions did not impact the measures.

Third, the measures proved to be location-invariant, and the allocation (mathematical permutation) of the values does not matter for the overall result. This confirms its poor reaction for structural and regional shifts in business. This finding also shows, that those measures would be a poor instrument for tracking very slight changes or errors in the data.

Fourth, the "real" intervals of the values were shown and compared with the theoretically achievable extreme values of the measures. The paper clearly shows that one should neither assume that the measure will reach its extreme value (min or max) nor that the interpretation intervals must be adjusted to the real variability. Much narrower intervals of empirical values of measures make the

Please cite this article in press as: Kopczewska K., Cluster-based measures of regional concentration. Critical overview. Spatial Statistics (2018), https://doi.org/10.1016/j.spasta.2018.07.008.

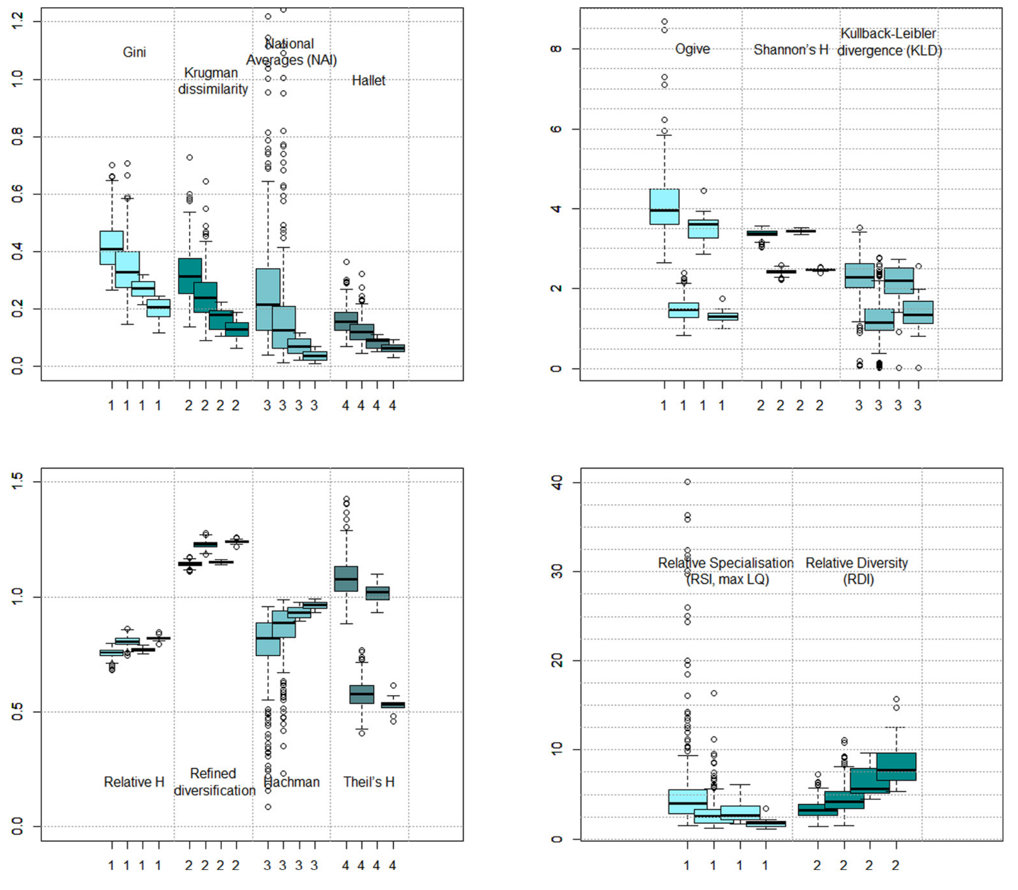


Fig. 8. Sectoral concentration of business units in Poland in 2017. Order of boxplots within the group: $lrls(n = 86, m = 380)$, $lrss(n = 20, m = 380)$, $srls(n = 86, m = 16)$, $srss(n = 20, m = 16)$.
Source: Own analysis in the R software on Central Statistical Office data.

interpretation more restrictive. Additionally, it sets the benchmark values, which can serve as typical results values.

Fifth, this paper proves with detailed correlations studies and similarity tests that the information capacity of the measures is similar within some groups. Thus, a more comprehensive summary of the empirical data analysed requires using the measures from the different information clusters. One should use one measure out of the group by a reference distribution, because the others are well correlated. These groups are (i) based on the empirical benchmark distribution, (ii) based on the uniform benchmark distribution, (iii) based on the transformed empirical distribution, or (iv) based on no marginal distribution of the employment table.

Sixth, the analysis indicates that one cannot treat the measure values as the absolute ones, because they are aggregation-dependent (matrix size dependent) and externally correlated with the position of the analysed region. Thus, they are not fully comparable between cases and they are not independent of the other factors and characteristics of the regions/sectors, such as the share in the region/sector in the overall volume. This finding limits the possibility of the inter-temporal and inter-regional/inter-national comparisons. The strongest dependence was observed for the disaggregated data, while in the smaller tables this relationship decreases.

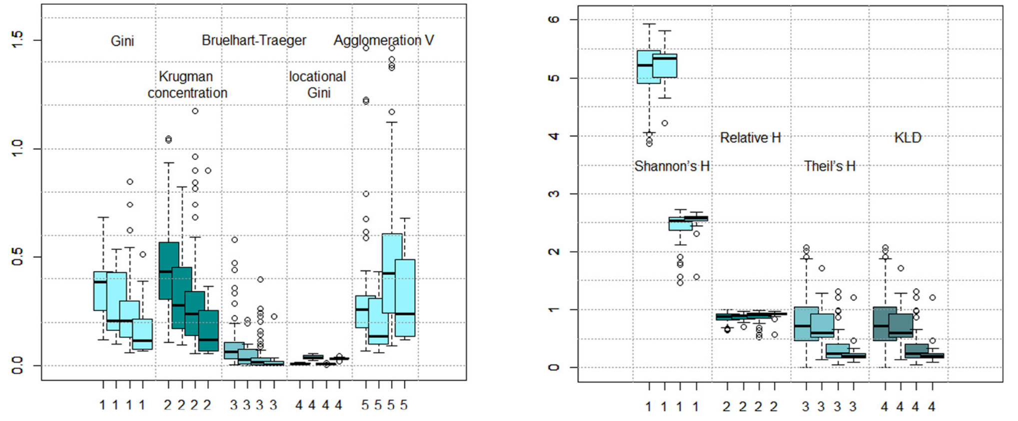


Fig. 9. Geographical concentration of business units in Poland in 2017. Order of boxplots within the group: $lrls(n = 86, m = 380)$, $srls(n = 86, m = 16)$, $lrss(n = 20, m = 380)$, $srss(n = 20, m = 16)$.
Source: Own analysis in the R software on Central Statistical Office data.

Lastly, one should not interpret them in a manner outside the predetermined scope as outlined in Section 1. These measures never look inside the cell because they give nothing more than the degree of concentration (sectoral concentration in a region or geographical concentration in a sector). Following the rules of interpretation in Table 2, the sectoral concentration (SC) measures can select the industries that have the largest share in the economy. The geographical concentration (GC) measures can test the inter-regional allocation of business and which region could attract more/less of the industry than the other regions.

This paper analysed the first generation concentration measures (Menon, 2012), which uses as the input the contingency tables by sectors and regions only. This simulation can be expanded to the second generation concentration measures, such as Ellison and Glaeser's index or Maurel and Sedillot index, which additionally include the component based on the Herfindahl index. There are also distance-based agglomeration measures, based on Ripley's K (Duranton and Overman, 2005) or SPAG (Kopczewska, 2017b), to answer the question of what occurs inside the region in terms of the spatial allocation of the business, and it can test the density of the allocation.

In conclusion, the simulation proves that the rule of thumb for the applied researchers is to use as highly aggregated data as possible (if there is a choice) and select a representative measure from each group for comparison purposes, with middle or high variance (e.g. Ogive/Theil's H + Hallet/Krugmann + Gini). The information content within a group is similar, and thus, the value added from analysing all of the measures from the same group is moderate. One cannot expect that disaggregated datasets will provide more reliable information. The measures should not be compared for different levels of aggregation and, consequently, for significantly different numbers of regions and sectors. Slight structural changes are difficult to track, and measures can be insensitive. Measures with high variance react more strongly for any divergences than those with low variance. In the interpretation, one should not expect the values to be close to the lower or upper limits, as they will mostly be from the moderate inside of the theoretically expected interval.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spasta.2018.07.008.

References

Amrhein, C.G., 1995. Searching for the elusive aggregation effect: evidence from statistical simulations. Environ. Plann. A 27 (1), 105–119.

Arbia, G., 1989. Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems. Kluwer.
Arbia, G., Bee, M., Espa, G., 2013. Testing isotropy in spatial econometric models. Spat. Econ. Anal. 8 (3), 228–240.
Bailey, D., Boyle, S.E., 1971. The optimal measure of concentration. J. Amer. Statist. Assoc. 66 (336), 702–706.
Briant, A., Combes, P.P., Lafourcade, M., 2010. Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations. J. Urban Econ. 67 (3), 287–302.
Cassey, A.J., Smith, B.O., 2014. Simulating confidence for the Ellison–Glaeser index. J. Urban Econ. 81, 85–103.
Churski, P., Ochojski, A., Polko, K., 2018. Rethinking regional specialization (forthcoming).
Clark, W.A., Avery, K.L., 1976. The effects of data aggregation in statistical analysis. Geogr. Anal. 8 (4), 428–438.
Deltas, G., 2003. The small-sample bias of the gini coefficient: results and implications for empirical research. Rev. Econ. Stat. 85 (1), 226–234.
Dewhurst, J., McCann, Ph., 2002. A comparison of measures of industrial specialization for travel-to-work areas in Great Britain, 1981–1997. Reg. Stud. 36 (5), 541–551.
Dewhurst, J., McCann, P., 2007. Specialization and regional size. In: Fingleton, B. New Directions in Economic Geography. Edward Elgar, Cheltenham, UK, pp. 204–229.
Duranton, G., Overman, H.G., 2005. Testing for localization using micro- geographic data. Rev. Econom. Stud. 72 (4), 1077–1106.
Eisenberg, B., 2015. The multivariate Gini ratio. Statist. Probab. Lett. 96 (2015), 292–298.
Ellison, G., Glaeser, E., 1997. Geographic concentration in US manufacturing industries: A dartboard approach. J. Polit. Econ. 105 (5), 889–927.
Fabrizi, E., Trivisano, C., 2016. Small area estimation of the gini concentration coefficient. Comput. Statist. Data Anal. 99 (2016), 223–234.
Flowerdew, R., 2011. How serious is the modifiable areal unit problem for analysis of English census data. Popul. Trends (2011), 145. Autumn.
Gehlke, C., Biehl, K., 1934. Certain effects on grouping upon the size of the correlation coefficient in census tract material. J. Amer. Statist. Assoc. 29 (185), 169–170.
Hart, P.E., 1971. Entropy and other measures of concentration. J. R. Stat. Soc. Ser. A 73–85.
Holt, D., Steel, D.G., Tranmer, M., 1996. Area homogeneity and the modifiable areal unit problem. Geogr. Syst. 3, 181–200.
Kopczewska, K., 2017a. Cluster-based measurement of agglomeration, concentration and specialisation. In: Kopczewska, K., Churski, P., Ochojski, A., Polko, A. (Eds.), 2017, Measuring Regional Specialisation –A New Approach. Palgrave MacMillen, London.
Kopczewska, K., 2017b. Distance-based measurement of agglomeration, concentration and specialisation. In: Kopczewska, K., Churski, P., Ochojski, A., Polko, A. (Eds.), 2017, Measuring Regional Specialisation –A New Approach. Palgrave MacMillen, London.
Krzanowski, W., 2000. Principles of Multivariate Analysis. OUP Oxford.
Kwoka, J.E., 1981. Does the choice on concentration measure really matter? J. Ind. Econ. 445–453.
Latreille, P.L., Mackley, J., 2011. Using excel to illustrate Hannah and Kay's concentration axioms. Int. Rev. Econ. Educ. 10 (1), 117–127.
López, F.A., Angulo, A., Artal, A., 2012. Capturing spatial clusters of activity in the Spanish Mediterranean axis. In: Defining the Spatial Scale in Modern Regional Analysis. Springer, Berlin, Heidelberg, pp. 255–280.
Manley, D., 2014. Scale, aggregation, and the modifiable areal unit problem. In: Fischer, M.M., Nijkamp, P. (Eds.), Handbook of Regional Science. Springer-Verlag, Berlin, pp. 1157–1171.
Menon, C., 2012. The bright side of MAUP: Defining new measures of industrial agglomeration. Papers Reg. Sci. 91 (1), 3–28.
Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos., P., Henry, M., Stevens, H., Szoecs, E., Wagner, H., 2017. Vegan: Community Ecology Package. R, Package Version 2.4-3. https://CRAN.R-project.org/package=vegan.
Openshaw, S., 1978. An optimal zoning approach to the study of spatially aggregated data. In: Masser, I., Brown, P.J.B. (Eds.), Spatial Representation and Spatial Interaction. Martinus Nijhoff, Leiden, pp. 93–113.
Openshaw, S., 1984. The Modifiable Areal Unit Problem, CATMOG 38. GeoBooks, Norwich.
Openshaw, S., Baxter, R.S., 1977. Algorithm 3: a procedure to generate pseudo-random aggregations of N zones into M zones, where M is less than N. Environ. Plann. A 9 (12), 1423–1428.
Palan, N., 2010. Measurement of Specialization –The Choice of Indices, FIW Working Paper No. 62, December 2010.
Reynolds, H.D., 1998. The Modifiable Areal Unit Problem: Empirical Analysis by Statistical Simulation (Ph.D. thesis), University of Toronto, Department of Geography.
Robinson, W.S., 1950. Ecological correlations and the behavior of individuals. Amer. Sociol. Rev. 15, 351–357.
Rosenbluth, G., 1955. Measures of concentration. In: Business Concentration and Price Policy. Princeton University Press, 57-99.
Santi, F., Arbia, G., Bee, M., Espa, G., 2017. A frequency domain test for isotropy in spatial data models. Spat. Stat. 21, 262–278.
Scherer, F.M., 1970. Industrial Market Structure and Economic Performance, Chicago.
Schmalensee, R., 1977. Using the H-index of concentration with published data. Rev. Econ. Stat. 186–193.
Sleuwaegen, L., Dehandschutter, W., 1986. The critical choice between the concentration ratio and the H-index in assessing industry performance. J. Ind. Econ. 193–208.
Stillwell, J., Daras, K., Bell, M., 2017. Spatial Aggregation Methods for Investigating the MAUP Effects in Migration Analysis, Working Paper.
Wei, T., Simko, V., 2016. Corrplot: Visualization of a Correlation Matrix. R Package Version 0.77. https://CRAN.R-project.org/package=corrplot.
Wong, D., 2009. The modifiable areal unit problem (MAUP). In: The Sage Handbook of Spatial Analysis. pp. 105–123.

Yitzhaki, S., Schechtman, E., 2012. The Gini Methodology: A Primer on a Statistical Methodology, vol. 272. Springer Science & Business Media.

Further Reading

Alonso-Villar, O., Chamorro-Rivas, J.M., González-Cerdeira, X., 2004. Agglomeration economies in manufacturing industries: the case of Spain. Appl. Econ. 36 (18), 2103–2116.
Bergstrand, J.H., 1985. The gravity equation in international trade: Some microeconomic foundations and empirical evidence. Rev. Econ. Stat. 67 (3), 474–481.
Bruelhart, M., Traeger, R., 2005. An account of geographic concentration patterns in Europe. Reg. Sci. Urban Econ. 35, 597–624.
Duranton, G., Puga, D., 2000. Diversity and specialisation in cities: Why, where and when does it matter? Urban Stud. 37 (3), 533–555.
Franceschi, F., Mussoni, M., Pelloni, G., 2009. A Note on New Measures of Agglomeration and Specialization. University of Bologna, Unpublished, http://amsacta.cib.unibo.it/2683/.
Gini, C., 1909. Concentration and dependency ratios. Riv. Polit. Econ. 87 (1997), 769–789. English Translation (in Italian).
Gini, C., 1936. On the Measure of Concentration with Special Reference to Income and Statistics. In: General Series, vol. 208, Colorado College Publication, pp. 73–79.
Guillain, R., Le Gallo, J., 2010. Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis. Environ. Plann. B 37, 961–981.
Hallet, M., 2000. Regional specialisation and concentration in the EU. In: Economic Papers, vol. 141. European Commission, Brussels.
Hoover, E.M., 1936. The measurement of industrial localization. Rev. Econ. Stat. 162–171.
Horowitz, A., Horowitz, I., 1968. Entropy, Markov processes and competition in the brewing industry. J. Ind. Econ. 196–211.
Isard, W., 1960. Methods of Regional Analysis: An Introduction to Regional Science. John Wiley & Sons Inc., New York.
Kim, Y., Barkley, D.L., Henry, M.S., 2000. Industry characteristics linked to establishment concentrations in nonmetropolitan areas. J. Reg. Sci. 40, 231–259.
Krugman, P., 1991. Geography and Trade. MIT Press.
Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86.
Lilien, D.M., 1982. Sectoral shifts and cyclical unemployment. J. Polit. Econ. 777–793.
Maurel, F., Sedillot, B., 1999. A measure for geographical concentration of French manufacturing industries. Reg. Sci. Urban Econ. 29 (5), 575–604.
Mori, T., Nishikimi, K., Smith, T.E., 2005. A divergence statistic for industrial localization. Rev. Econ. Stat. 87 (4), 635–651.
OECD Regions at a Glance, 2009. http://www.oecd-ilibrary.org/urban-rural-and-regional-development/oecd-regions-at-a-glance-2013_reg_glance-2013-en.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423 and 623–656. Mathematical Reviews (MathSciNet): MR10, 133e.
Tress, R.C., 1938. Unemployment and Diversification of Industry, vol. 9. The Manchester School, pp. 140–152.
Wundt, B.D., 1992. Reevaluating alternative measures of industrial diversity as indicators of regional cyclical variations. Rev. Reg. Stud. 22 (1), 59.