**Chapter 2**

**Cluster-based measurement of agglomeration,**

**concentration and specialisation**

***Katarzyna Kopczewska***

***Summary:***

***Cluster-based measures of concenatrtion and specialisation compare aggregated regional and sectoral activity, without looking inside the regional allocation of firms. Plenty of available measures give similar information, as the indices are based on the same two-dimensional data matrix and also have the same underlying benchmark distributions, the empirical or theoretical one. This chapter reviews in encyclopedic way the existing cluster-based measures, popularly called concentration and specialization indices, calculating them on the same data table. As the analysis of their results show, the complex indicators, compared with simple and more popular ones, perform similarly and the conclusions from simpler and advanced indicators are mainly coherent.***

***Abstract:***

***This chapter reviews in encyclopedic way the existing cluster-based measures, popularly called concentration and specialization indices, calculating them on the same data table.Using simple example it gives the way of its calculation and interpretation. Chapter compares the measures in terms of their construction, ways of calculation and results obtained. Last part of the chapter compares the results, and proves that even though there are plenty of indices, in fact one can use representative indicators for given groups, as the other measures converge.***

T

he literature of last century is very rich in coefficients, indicators, indices etc., which are to measure phenomena of agglomeration, concentration and specialisation. To our knowledge, there is no paper or book which systematises them all. Most of works arbitrarily select few of them and compare. This selection is usually based on the popularity of a given measure or compares a new measure to the existing one, and is treated as “classical”.

The second issue is what they really measure. As shown in Chapter 1, regionalists treat the terms: agglomeration, concentration and specialization in a very light manner. Sometimes they are distinguished, at times treated interchangeably, and other times interpreted oppositely. In this book, we propose to clarify them. As discussed earlier, the phenomenon of business locations is considered in relation to geographical space, as well as to other companies. Comparing the density of the location of companies in the area (in the region), one speaks about the agglomeration. The spatial distribution of companies can take various forms, from being concentrated in one location, to a random or uniform distribution throughout the area. Evaluation of spatial agglomeration requires the determination of the precise point geo-location and to compare this location with other companies. Comparing the importance of companies in the sector in a given area (eg. the share of employment) to the importance of companies from different sectors in a given area, one can talk about sectoral concentration, and comparing them with their meaning in a different region, one can talk about geographical concentration. With this measurement it is possible to carry on spatial aggregates (e.g. the sum of the employed in the region in the sector), and *de facto* abstracts of geographical location. The saturation of the region with a given sector (in relation to other industries or other regions) is independent of the fact that companies in the industry in the area are located at one point or evenly distributed over the territory. Thus, concentration might be twofold: sectoral concentration (that one sector dominates over others in given region, usually this is called “specialisation”) and geographical concentration (that one region concentrates most of the employment in a given sector). Agglomeration is the spatial density of location of single firms inside one region (and possibly one sector). Specialisation is the outcome of concentration and agglomeration and is interpreted in terms of both these measures. Therefore, concentration understood in this way (sectoral and geographical) uses a *cluster-based* measures (based on data aggregated regionally and sectoral) and agglomeration requires the measures based on the original data points, which use the distance between locations (*distance-based*).

Third issue is on what data they are based. We assume here two possible types of datasets: individual georeferenced data for business units (firms) giving information on the size of the company and sector, or a two-dimensional table with aggregated employment within the region and sector. The first type of dataset allows for calculating the agglomeration measure, and after aggregation also concentration measures. The second type of dataset, because of reduced information on absolute location inside the region, allows for concentration measures only. Following the types of data, we divide measures into two groups: cluster-based (on aggregated data) and distance-based (on individual point data). The typology of cluster-based measures given in Table 2.1. All those measures will be described and compared in this Chapter. Distance-based measures will be analysed in Chapter 3.

From this perspective, the measures based on the aggregates or points relate to the concentration (geographical or sectoral) or spatial agglomeration. What’s important is that these are not the measures of specialisation, as they are often said to be. Neither sectoral or geographical overrepresentation nor spatial clustering prejudges regional specialisation. Regional specialisation is the result of these two factors, and may occur regardless of the type of combination of these two phenomena. Both low and high concentration and agglomeration can lead to specialisation in the region.

Most of the literature of the past few decades looked for innovations for cluster-based measures. Starting with the simple measures, that bind only sectoral and regional employment, throughout the measures taking into account measures of firm size, the area of the region, and neighborhood relations. However, it does not change the fact that such measures are not able to assess the spatial behavior of firms in the region, and because of aggregation this information is not available. It is only to assess regions compared with the macro-region (country, continent). The list of these measures is in the below table 2.1.

**Table 2.1: Typology and properties of cluster-based methods**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Index[[1]](#footnote-1) | n x k matrix | W matrix | Distance | Area | Size of firms |
| Gini index  Location Quotient (Hoover-Balassa coefficient)  Ogive index  Diversification index (Isard Index, refined index of diversification)  Krugman Dissimilarity index  Index of Agglomeration V  Hallet index  Lillen indicator  Theil’s Entropy  Shannon Entropy  Entropy index of overall localization (Cutrini, 2009)  Entropy measure (Bruelhart & Traeger, 2005)  Disproportionality Measures (Bickenbach & Bode, 2008)  Hachman Index  NAI (National averages index)  KLD (Kullback-Leibler divergence)  RSI (Relative Specialization Index) | V | --- | --- | -- | --- |
| Herfindahl index  Absolute and relative Theil index (Bickenbach, Bode & Krieger-Bode, 2012)  Relative Diversity Index (Duranton & Puga, 2001) | --- | --- | --- | --- | V |
| Ellison & Glaeser index  Excessive concentration (Ellison & Glaeser, 1997)  Maurel & Sedillot, 1999 | V | --- | --- | --- | V |
| Clustering index (Bergstrand, 1985) | V | --- | V | --- | --- |
| Concentration index (Spiezia, 2002),  Regional Industrial Mass and Regional Industrial Concentration (Franceschi, Mussoni & Pelloni, 2009) | V | --- | --- | V | --- |
| Gini with ESDA (Guillain & Le Gallo 2010),  Using Gini together with Moran’s I and Getis-Ord (Arbia, 2001b),  Spatial Concentration Measure (Arbia & Piras, 2009),  Relative Industrial Relevance (Carlei & Nuccio, 2014),  Inflation factor as correction of other measures (Guimaraes, Figueiredo & Woodward, 2011),  Spatial distribution (Sohn, 2014) | V | V | V | --- | --- |

*Source*: Own synthesis

We assume that the starting point for most of these measures is the two-dimensional table (see Table 2.2) in which *empi,j*is employment in sector *i* (there are *n* sectors, *i=1,2,…, n*) and in region *j* (and here are *m* regions, *j=1,2,.., m*).

**Table 2.2: Two-dimensional table as the basis for most of cluster-based measures**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***n* sectors**  ***i=1,2,..,n*** |  | ***m* regions*, j=1,2,.., m*** | | | | | |
|  | **Territory**  **A** | **Territory**  **B** | **Territory**  **C** | **Territory**  **D** | **…** | **Total** |
| **Industry 1** | emp(*ij)* | emp(*ij)* | emp(*ij)* | emp(*ij)* |  | **Σemp(ind1)=** |
| **Industry 2** | emp(*ij)* | emp(*ij)* | emp(*ij)* | emp(*ij)* |  | **Σemp(ind2)=** |
| **Industry 3** | emp(*ij)* | emp(*ij)* | emp(*ij)* | emp(*ij)* |  | **Σemp(ind3)=** |
| **Industry 4** | emp(*ij)* | emp(*ij)* | emp(*ij)* | emp(*ij)* |  | **Σemp(ind4)=** |
| **…** |  |  |  |  |  | **…** |
| **Total** | **Σemp(terrA)=** | **Σemp(terrB)=** | **Σemp(terrC)=** | **Σemp(terrD)=** | **…** | **ΣΣemp =** |

*Source*: Own concept

Table 2.3 is a starting point for most cluster-based measures, and cluster is understood here as total employment in some territory in some industry (*emp(ij)*). Starting from Aiginger (1999), literature is used as a base for the classification that specialisation reflects over-representation in vertical cross-section, compared with shares of other sectors in given region. Concentration was thought to be also over-representation but horizontally, referring sectoral employment in the region studied to employment in other regions. This approach, was criticised in Chapter 1, where we explain the motivation for this. Finally, **agglomeration** is understood as the spatial coverage of a region with firms, the spatial density of business. It is to distinguish spatial patterns of business locations, even or agglomerated over space.

To enable simple comparisons, we assume for the whole chapter an empirical simple example matrix (see Table 2.3 and 2.4), for which the measures will be calculated. Values of the employment in given sector and region are given as below.

|  |  |
| --- | --- |
| **Table 2.3: Two-dimensional example table of employment** | **Figure 2.1: Example regional map** |
| |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | |  | **region**  **1** | **region**  **2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** | | **Industry 1** | 1 | 11 | 21 | 70 | 10 | 6 | **119** | | **Industry 2** | 1 | 40 | 24 | 40 | 5 | 11 | **121** | | **Industry 3** | 5 | 13 | 21 | 30 | 35 | 1 | **105** | | **Industry 4** | 9 | 14 | 14 | 11 | 3 | 17 | **68** | | **Total** | **16** | **78** | **80** | **151** | **53** | **35** | **413** | |  |

*Source*: Own concept *Source*: Own graphics

**Table 2.4: Matrix of distances between regions and their areas**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| **region 1** | 0.00 |  |  |  |  |  |
| **region 2** | 0.78 | 0.00 |  |  |  |  |
| **region 3** | 0.40 | 0.85 | 0.00 |  |  |  |
| **region 4** | 0.43 | 0.54 | 0.32 | 0.00 |  |  |
| **region 5** | 0.54 | 0.29 | 0.73 | 0.46 | 0.00 |  |
| **region 6** | 0.55 | 0.53 | 0.86 | 0.66 | 0.27 | 0.00 |
|  |  |  |  |  |  |  |
| **area (in units)** | **0.182** | **0.184** | **0.132** | **0.156** | **0.123** | **0.177** |

*Source*: Own calculations

**2.1 Cluster-based measures depending only on *n x m* matrix**

**2.1.1 Gini index**

This oldest index of concentration was introduced by Gini in the early 20th century (1909, 1936). There are at least two versions of Gini index: the sectoral concentration index and the geographical concentration index. The first one, sectoral concentration index, takes as the subject of study the region (*j*) and analyses all *i* sectors inside it, referring to the average sectoral structure in the whole economy. In terms of the technical analysis on the basis of the table 2.3, it is vertical analysis (in columns). The second, concentration index, operates oppositely, horizontally in rows. It analyses one *i* sector in all *j* regions, linking the sectoral inter-regional structure to the inter-regional structure of whole economy.

Traditional Gini sectoral concentration index for the *j* region is then as follows:

where n is number of industries, is ratio of employment, is the average of employment ratio (by industries), is the rank of the industry’s position in descending order of Ri, is the ratio of employment in sector *i* in region *j* to total employment (all *i* sectors) in the region j, is the ratio of employment in sector *i* in all *j* regions to total employment (all *i* sectors in all *j* regions). This Gini index is formulated for single regions and all sectors, and the reference points are all sectors in all regions. Thus it measures the average under/over representation of sectors in a given region in comparison with the benchmark given by all regions (a kind of average regional structure).

Oppositely, Gini “concentration” index for *i* sector is then as follows:

where m is number of regions, is ratio of employment, is the average of employment ratio (by regions), is the rank of the region’s position in descending order of Cj, is the ratio of employment in sector *i* in region *j* to total employment (all *j* regions) in sector *i*, is the ratio of employment in region *j* in all *i* sectors to total employment (all *i* sectors in all *j* regions). This Gini index is formulated for single industries and all regions, and the reference point are all industries in one region. Thus it measures the average under/over representation of one sector in all regions in comparison with the benchmark given by all sectors in all regions (a kind of sectoral average structure).

Interpretation of the traditional Gini index is straightforward. It can take values from 0 to 1. Gini = 0 means uniform distribution of activity among sectors/regions, thus studied and benchmarked distributions are equal. Gini = 1 is in case of full concentration (whole sectoral employment in one region only / full employment of region in one sector only). The higher the value of the Gini index the less similarity between industries and regions.

Apart from the traditional Gini index, there exists a **locational Gini** index (Kim et al., 2000, Guillain & Le Gallo, 2010). Its calculation is simpler than the in traditional Gini, and its values for concentration in *n* sectors can be compared. It is expressed as follows:

and

which is the total of matrix of absolute values of differences in share’s proportion for all pairs of regions, and

which is the proportion of shares: sector in region and region in country, and

what is the average proportion xj, where *i,j* are the subscripts of regions, *m* is the number of regions, and *n* is the number of sectors (industries). This index takes values from 0 to 0.5, and Gloc=0 is for equal distributions (between regions) of activity in the sector and whole economy, and Gloc=0.5 indicates extreme concentration of full activity in a single region only.

To operationalise the above formulas, we present traditional and locational Gini for data from Table 2.3.

**Table 2.5: Middle-steps for values of traditional Gini** *(for region 1 in rows and for industry 1 in columns)*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** |  |  |  |  |  |
| **Industry 1** | 1 | 11 | 21 | 70 | 10 | 6 | **119** | ss1,1=1/16  =0.06 | s1=119/413  =0.29 | R1=0.06/0.29  =0.22 | 3 | |0.22-1.27|⋅3  =3.16 |
| **Industry 2** | 1 | 40 | 24 | 40 | 5 | 11 | **121** | ss2,1=1/16  =0.06 | s2=121/413  =0.29 | R2=0.06/0.29  =0.21 | 4 | |0.21-1.27|⋅4  =4.22 |
| **Industry 3** | 5 | 13 | 21 | 30 | 35 | 1 | **105** | ss3,1=5/16  =0.31 | s3=105/413  =0.25 | R3=0.31/0.25  =1.23 | 2 | |1.23-1.27|⋅2  =0.08 |
| **Industry 4** | 9 | 14 | 14 | 11 | 3 | 17 | **68** | ss4,1=9/16  =0.56 | s4=68/413  =0.16 | R4=0.56/0.16  =3.42 | 1 | |3.42-1.27|⋅1  =2.15 |
| **Total** | **16** | **78** | **80** | **151** | **53** | **35** | **413** | =1 | =1 | =(0.22+0.21+ 1.23+3.42)/4=1.27 | --- | =  (3.16+4.22+0.08+  +2.15)=9.61 |
|  | =1/119  =0.01 | =11/119  =0.09 | =21/119  =0.18 | =70/119  =0.59 | =10/119  =0.08 | =6/119  =0.05 | =1 | = 0.95  = 0.49 | | | | |
|  | =16/413  =0.04 | =78/413  =0.19 | =80/413  =0.19 | =151/413  =0.37 | =53/413  =0.13 | =35/413  =0.08 | =1 |
|  | C1=0.01/0.04  =0.22 | C2=0.09/0.19  =0.49 | C3=0.18/0.19  =0.91 | C4=0.59/0.37  =1.61 | C5=0.08/0.13  =0.65 | C6=0.05/0.08  =0.59 | =(0.22+0.49+  +0.91+1.61+0.65+0.59)/6=0.75 |
|  | 6 | 5 | 2 | 1 | 3 | 4 | --- |
|  | |0.22-0.75|⋅6  =3.17 | |0.49-0.75|⋅5  =1.28 | |0.91-0.75|⋅2  =0.33 | |1.61-0.75|⋅1  =0.86 | |0.65-0.75|⋅3  =0.27 | |0.59-0.75|⋅4  =0.60 | =  (3.17+1.28+0.33+  0.86+0.27+0.60)=  =6.53 |

*Source*: Own calculations

**Table 2.6: Middle-steps for values of locational Gini**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **x** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
|
| **Industry 1** | (1/16)/(16/413)  =1.61 | (11/78)/(78/413)  =0.75 | (21/80)/(80/413)=1.36 | (70/151)/(151/413)  =1.27 | (10/53)/(53/413)  =1.47 | (6/35)/(35/413)  =2.02 |
| **Industry 2** | (1/16)/(16/413)  =1.61 | (40/78)/(78/413)  =2.72 | (24/80)/(80/413)  =1.55 | (40/151)/(151/413)  =0.72 | (5/53)/(53/413)  =0.74 | (11/35)/(35/413)  =3.71 |
| **Industry 3** | (5/16)/(16/413)  =8.07 | (13/78)/(78/413)  =0.88 | (21/80)/(80/413)  =1.36 | (30/151)/(151/413)  =0.54 | (35/53)/(53/413)  =5.15 | (1/35)/(35/413)  =0.34 |
| **Industry 4** | (9/16)/(16/413)  =14.52 | (14/78)/(78/413)  =0.95 | (14/80)/(80/413)  =0.90 | (11/151)/(151/413)  =0.20 | (3/53)/(53/413)  =0.44 | (17/35)/(35/413)  =5.73 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***for industry 1***  ***|*** | | 1.61 | 0.75 | 1.36 | 1.27 | 1.47 | 2.02 |
| **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| 1.61 | **region 1** | |1.61-1.61|=  0.000 | |1.61-0.75|=  0.867 | |1.61-1.36|=  0.258 | |1.61-1.27|=  0.345 | |1.61-1.47|=  0.143 | |1.61-2.02|=  0.410 |
| 0.75 | **region 2** | |0.75-1.61|=  0.867 | |0.75-0.75|=  0.000 | |0.75-1.36|=  0.608 | |0.75-1.27|=  0.521 | |0.75-1.47|=  0.724 | |0.75-2.02|=  1.276 |
| 1.36 | **region 3** | |1.36-1.61|=  0.258 | |1.3—0.75|=  0.608 | |1.36-1.36|=  0.000 | |1.36-1.27|=  0.087 | |1.36-1.47|=  0.115 | |1.36-2.02|=  0.668 |
| 1.27 | **region 4** | |1.27-1.61|=  0.345 | |1.27-0.75|=  0.521 | |1.27-1.36|=  0.087 | |1.27-1.27|=  0.000 | |1.27-1.47|=  0.202 | |1.27-2.02|=  0.755 |
| 1.47 | **region 5** | |1.47-1.61|=  0.143 | |1.47-0.75|=  0.724 | |1.47-1.36|=  0.115 | |1.47-1.27|=  0.202 | |1.47-1.47|=  0.000 | |1.47-2.02|=  0.553 |
| 2.02 | **region 6** | |2.02-1.61|=  0.410 | |2.02-0.75|=  1.276 | |2.02-1.36|=  0.668 | |2.02-1.27|=  0.755 | |2.02-1.47|=  0.553 | |2.02-2.02|=  0.000 |

|  |  |
| --- | --- |
| **Total of *|*** | 0+0.867+0.258+0.345+0.143+0.410+0.867+0+0.608+0.521+0.724+1.276+  0.258+0.608+0+0.087+0.115+0.668+0.345+0.521+0.087+0+0.202+0.755+  0.143+0.724+0.115+0.202+0+0.553+0.410+1.276+0.668+0.755+0.553+0=  =15.064 |
| **m(m-1)** | 6⋅(6-1)=  =30 |
|  | (1.61+0.75+1.36+1.27+1.47+2.02)/6=  =1.413 |
| **G** | (15.064/30)/(4\*1.413)=  =0.089 |

*Source*: Own calculations

Interpretation is as follows:

- The traditional Gini index of sectoral concentration for region 1 (G=0.95) shows that overrepresentation of a single sector in this region (compared with benchmark sectoral distribution for whole economy) is strong and the region does “specialise” in one of the sectors (industry 4).

- The traditional Gini index of concentration for industry 1 (G=0.49) proves that employment in this sector is rather less than equally distributed over regions and some of the regions slightly concentrate the employment in sector 1. As one can see, the Gini result are the aggregated indices for a given sector or region. For n sectors and m regions one gets *n+m* results.

- Location Gini for industry 1 Gloc,1=0,089 is closer to the lower border (G=0), which means that the distribution of activity in sector 1 is more diversified among regions.

It was commonly used in many studies as in e.g. Malmberg & Maskell (1997) etc. There are also modifications – an example can be the method of decomposing Gini as a disproportionality measure to flexibly measure sectoral and geographical concentration (usually called specialisation) (Bickenbach & Bode, 2008).

* + 1. **Location Quotient (Hoover-Balassa coefficient)**

One of the most popular and simplest measures to calculate is the Location Quotient (LQ), called also *Hoover-Balassa coefficient* or *specialisation index*, introduced by Hoover (1936). It is the measure of relative regional employment, comparing the distributions of employment by industry, having as the reference area all regions together (country). It is calculated as the relation of local sectoral employment to regional sectoral employment, with the formula:

where the counter is the ratio of employment in sector *i* in region *j* to total employment (all *i* sectors) in region j, and denominator is the ratio of employment in sector *i* in all *j* regions to total employment (all *i* sectors in all *j* regions). Its construction is similar to Gini’s index components. Sectoral LQ’s summed up are often called a measure of *specialisation*, even if they measure sectoral concentration only.

To operationalise the above formula, we present LQ for data from Table 2.3.

**Table 2.7: Regional-sectoral LQ indices**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *emp­i.j* | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
|
| **Industry 1** | (1/16)/  (119/413)  = 0.22 | (11/78)/  (119/413)  = 0.49 | (21/80)/  (119/413)  = 0.91 | (70/151)/  (119/413)  = 1.61 | (10/53)/  (119/413)  = 0.65 | (6/35)/  (/413)  = 0.59 |
| **Industry 2** | (1/16)/  (121/413)  = 0.21 | (40/78)/  (121/413)  = 1.75 | (24/80)/  (121/413)  = 1.02 | (40/151)/  (121/413)  = 0.90 | (5/53)/  (121/413)  = 0.32 | (11/35)/  (121/413)  = 1.07 |
| **Industry 3** | (5/16)/  (105/413)  = 1.23 | (13/78)/  (105/413)  = 0.66 | (21/80)/  (105/413)  = 1.03 | (30/151)/  (105/413)  = 0.78 | (35/53)/  (105/413)  = 2.60 | (1/35)/  (105/413)  = 0.11 |
| **Industry 4** | (9/16)/  (68/413)  = 3.42 | (14/78)/  (68/413)  = 1.09 | (14/80)/  (68/413)  = 1.06 | (11/151)/  (68/413)  = 0.44 | (3/53)/  (68/413)  = 0.34 | (17/35)/  (68/413)  = 2.95 |

*Source*: Own calculations

The example values for given regions and industries were calculated as follows:

As one can see, LQ result are the individual separate values for each “cell” – region and industry. For n sectors and m regions one gets *n⋅m* results. LQregion1, industry1=0.22 means a strong underrepresentation of branch 1 in employment of a given region (1). Oppositely, LQregion5, industry3=2.60 indicates the strong concentration of employment in industry 3 in region 5. From absolute values one can see it is the region with domination of this sector, both from a sectoral perspective (among all regions), as well as an inside-regional perspective (among all sectors).

The basic idea for LQ was a part of bigger concept, mainly the **Economic base analysis** (Haig, 1928). There, the economy is divided into two categories: *basic*, the activity which build the wealth of a region by exporting, and *nonbasic*, which supports the basic activity (mainly service). LQ can indicate if the given economic activity is basic or non-basic. Basic local activity is one above the national / regional average. LQ can be calculated on trade data, but if not available on employment data. Thus, the overrepresentation (dominance) of employment in a given sector or section is thought of as a factor which opens the economy for trade. In traditional interpretation, this sectoral concentration of employment is interpreted as the excessive supply over evenly and uniformly distributed demand, thus the region is an exporter then. The rule of thumb is that when LQ>1.25, the region can be classified as a potential exporter. Sectors with LQ<1 are potential importers. This is built, however, on comparative advantage theory, and is weakness is in too simplified conditions. This trade approach was widely criticised. The main reasons for the critique are the sensitivity to data aggregation (by industries and regions) and global linkages in production and consumption. The technical issue is in the production of intermediate goods, which when produced and used locally to real export production are “invisible” for the LQ in determining the export potential.

Because of the critique, the location quotient is currently used rather as a measure of local industry composition. This is often associated with specialisation, which, as we show in previous chapters, is also not right. Interpretation of the overrepresentation of employment should be also done with care. Values slightly above or under 1 can result from data statistical effect. Thus many researchers suggest defining a higher threshold for the cut-off point for significant concentration of activity. This approach is to indicate strengths and weaknesses of the regions, and it can also help with branding regions, when looking for their main activities. In this approach, the problems appear when analysing very small territorial units, here treated as autonomous economy. Outliers that cause high a LQ in a given industry do not really represent the competitive concentration. A single big company, making the LQ high, may not be enough to become an engine for development strategy.

The second issue is to analyse jointly the value of LQ and trends in employment. The two-dimensional approach, with employment growth and LQ on the axes, can help classify the regions as follows:

* *low employment growth and low LQ* – regions with poor potential in a given industry, least-promising goals for strategic development policy.
* *low employment growth and high LQ –* possible target of policy to support important elements of local economy.
* *high employment growth and low LQ* – high potential sectors for local economies
* *high employment growth and high LQ* – most successful regional economy drivers

The location quotient also has its application in employment trends forecasting. Comparing LQ over time one can easily find the tendencies in a given sector in a given region, relativised with overall tendencies in employment, both sectoral and general. This easy tool is very helpful to policy analysts, but also quite dangerous when used blindly.

It is worth underlining that in fact LQ assumes a uniform distribution of sectoral economic activity, at least in aggregated data, with reference to a given sector. It does not look inside the region and spatial distribution of this activity. It also does not assume that some sectors should be concentrated geographically, in a single region only.

A discussion on the interpretation of the characteristics of this index can be found in a book by Kuroiwa (2012). Following Benedictis and Tamberi (2004), one can note that an increase in the index is possible because of a drop in sectoral share in the overall economy or an increase in the sectoral share in the regional economy. Their second recommendation is to use the LQ together with other methods like input-output values. LQ is also often analysed because of its statistical properties. Guimaraes et al. (2009) construct the dartboard test for LQ and Billings & Johnson (2012) test its statistical properties and accuracy of test for LQ. It is commonly used in many studies. The applications of the LQ index can be found in e.g. Raj Sharma (2008).

**2.1.3 Hachman index of economic diversification**

The Location Quotient is the main component of the Hachman index of economic diversity. It is the inverse total of sectoral LQs, weighted with the share of regional-sectoral employment. It is expressed as follows:

where is the ratio of employment in sector *i* in region *j* to total employment (all *i* sectors) in the region *j*, is the ratio of employment in sector *i* in all *j* regions to total employment (all *i* sectors in all *j* regions). It is the measure of similarity of regional and national industrial structures. It is limited between 0 (when the region has a completely different structure than the country) and 1 (for exactly the same industrial structure on a regional and national level).

To operationalize the above formula, we present the Hachman Index we base on the regional – sectoral LQ for data from Table 2.3.

**Table 2.8: Hachman index on the basis of LQ and shares of industry in region**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
|
| **Industry 1** | (1/16)⋅0.22  =0.01 | (11/78)⋅0.49  =0.07 | …=0.24 | …=0.75 | …=0.12 | …=0.10 |
| **Industry 2** | (1/16)⋅0.21  =0.01 | (40/78)⋅1.75  =0.90 | …=0.31 | …=0.24 | …=0.03 | …=0.34 |
| **Industry 3** | (5/16)⋅1.23  =0.38 | (13/78)⋅0.66  =0.11 | …=0.27 | …=0.16 | …=1.72 | …=0.00 |
| **Industry 4** | (9/16)⋅3.42  =1.92 | (14/78)⋅1.09  =0.20 | …=0.19 | …=0.03 | …=0.02 | …=1.43 |
| **Total** | 0.01+0.01  +0.38+1.92=  =2.33 | 0.07+0.90  +0.11+0.20=  =1.27 | …=1.003 | …=1.17 | …=1.89 | …=1.88 |
| **Hachman index** | 1/2.33=0.429 | 1/1.27=0.786 | 1/1.00=0.997 | 1/1.17=0.853 | 1/189=0.529 | 1/1.88=0.533 |

*Source*: Own calculations

From these results one can see that region 3 has almost the same economic structure as the country (all 6 regions on average), as the HI is almost 1 (HI=0,997). The least similar is region 1 (the Hachman index is the lowest, HI=0,429).

An application of the Hachman index can be found in e.g. Raj Sharma (2008), Shuai (2013) as well the OCED Territorial Reviews.

* + 1. **Ogive index**

The Ogive index was introduced by Tress (1938) to measure industrial diversity. It is mainly a measure of export structure, but it is sometimes applied to the regional production structure. It is based on the uniform (equal) distribution of export shares treated as a benchmark, and it captures the deviations from it. In the regional version, the equal distribution of activity (employment) is a benchmark. The formula for the regional Ogive index is expressed as follow:

|  |  |
| --- | --- |
|  | (4) |

where is the ratio of employment in sector *i* in region *j* to total employment (all *i* sectors) in region *j* (empirical share of employment in given sector), and is the ideal share of employment in a given industry, resulting from equal distribution. Values of the Ogive index are for the whole region, and define diversification or sectoral concentration of activity in the region analysed. When employment shares among sectors are equal, the ogive index is 0, which is interpreted as perfect diversity. The more diversified (different, unequal) values, the higher the Ogive measure.

To operationalise the above formula, we present the Ogive index for data from Table 2.3.

**Table 2.9: Components of the Ogive index**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *emp­i.j* | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
|
| **Industry 1** | ((1/16)-(1/4))2/  (1/4)=  =0.14 | ((11/78)-(1/4))2/  (1/4)=  =0.05 | ((21/80)-(1/4))2/  (1/4)=  =0.00 | ((70/151)-(1/4))2/  (1/4)=  =0.18 | ((10/53)-(1/4))2/  (1/4)=  =0.02 | ((6/35)-(1/4))2/  (1/4)=  =0.02 |
| **Industry 2** | ((1/16)-(1/4))2/  (1/4)=  =0.14 | ((40/78)-(1/4))2/  (1/4)=  =0.28 | ((24/80)-(1/4))2/  (1/4)=  =0.01 | ((40/151)-(1/4))2/  (1/4)=  =0.00 | ((5/53)-(1/4))2/  (1/4)=  =0.10 | ((11/35)-(1/4))2/  (1/4)=  =0.02 |
| **Industry 3** | ((5/16)-(1/4))2/  (1/4)=  =0.02 | ((13/78)-(1/4))2/  (1/4)=  =0.03 | ((21/80)-(1/4))2/  (1/4)=  =0.00 | ((30/151)-(1/4))2/  (1/4)=  =0.01 | ((35/53)-(1/4))2/  (1/4)=  =0.67 | ((1/35)-(1/4))2/  (1/4)=  =0.20 |
| **Industry 4** | ((9/16)-(1/4))2/  (1/4)=  =0.39 | ((14/78)-(1/4))2/  (1/4)=  =0.02 | ((14/80)-(1/4))2/  (1/4)=  =0.02 | ((11/151)-(1/4))2/  (1/4)=  =0.13 | ((3/53)-(1/4))2/  (1/4)=  =0.15 | ((17/35)-(1/4))2/  (1/4)=  =0.22 |
| **Total** | **0.14+0.14+**  **0.02+0.39=**  **=0.69** | **0.05+0.28+**  **0.03+0.02=**  **=0.37** | **0.00+0.01+**  **0.00+0.02=**  **=0.03** | **0.18+0.00+**  **0.01+0.13=**  **=0.32** | **0.02+0.10+**  **0.67+0.15=**  **=0.94** | **0.02+0.02+**  **0.20+0.22=**  **=0.46** |

*Source*: Own calculations

The Ogive index for region 1 (with 4 sectors) is as follows:

In the above example, the most diversified region (with the least uniformly distributed employment) is region 5, and region 3 has the least diversified economical structure.

Although the position of the Ogive index was for long well-grounded, it received a lot of criticism (Bahl et al., 1971, Wasylenko & Erickson, 1978). The points raised against the Ogive index are: a) that absolute diversity measured by this index is not an adequate comparative norm, as there is no reason why all sectors should have the same share in employment (when different productivity, demand as well organisational and institutional structures are being considered); b) that the inherent nature of sectors is reflected in stability and resistance against crises, thus those results are not very informative; c) time shifts in economic structure cause the comparison of sector shares over periods and regions to be misleading, especially when a region with some lag to the country is changing its structure. Also, the Ogive index was compared with an entropy measure for its usefulness in measuring the diversification of economy (Wasylenko & Erickson, 1978). As the authors claim, that the Ogive index behaves the same as the entropy measure. This measure is also sensitive to the number of sectors in a region.

* + 1. **Diversification index**

Diversification index was one the earliest popular measures of economic diversity of regions, introduced by Tress (1938). Isard (1960) defines the “*crude index of diversification*” as follows: one starts with shares of employment in all *k* sections in a given region (noted as x (%)), ranks them decreasingly and sums up cumulatively (each crude share is associated with the cumulative one, which is the sum of all bigger shares and itself). The crude index of the diversification is the sum of all those cumulative sums. If all activity is concentrated in one sector, diversification index has a value of *k*\*100. The extreme values for least and greatest diversity are shown in the table 2.10 below. This is an absolute measure, with no reference to neighbours and other regions. It gives one single value for all sectors in one region.

Rodgers (1957) introduced the refined index of diversification. This refinement requires the inclusion of other regions as benchmarks and makes this measure relative. The reference here is the whole big area (a few regions, the country, etc.) by industry. The refined index of diversification (also *crude index of diversity*) is as follows: (*crude index for region* – *crude index for all regions together*)/(*crude index for least diversity case* – *crude index for all regions together*). A value of 0 means that the given region has the same diversification as the fully diversified benchmark area (equal shares), while a value of 1 means complete non-diversification. A negative value may appear if the region’s distribution is more equal than in the reference system. Thus the refined index of diversification is the measure of deviation of a region from the diversification pattern in all regions.

Below we operationalise the crude and refined index of diversification on the data from Table 2.3.

**Table 2.10: Components of crude and refined index of diversification**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | … | **Total** | real shares  region 1  (%) | real shares by sectors all regions (%) |  | hypothetic equal shares (any region) (%) | cumulative sums for equal shares (greatest diversity) |  | hypothetic full concentration shares (any region) (%) | cumulative sums for full concentration shares (least diversity) |
| **Industry 1** | 1 | … | **119** | 1/16= 6,25 | 119/413= 28,81 |  | 25 | 25 |  | 100 | 100 |
| **Industry 2** | 1 | … | **121** | 1/16= 6,25 | 121/413= 29,30 |  | 25 | 50 |  | 0 | 100 |
| **Industry 3** | 5 | … | **105** | 5/16= 31,25 | 105/413= 25,42 |  | 25 | 75 |  | 0 | 100 |
| **Industry 4** | 9 | … | **68** | 9/16= 56,25 | 68/413= 16,46 |  | 25 | 100 |  | 0 | 100 |
| **Total** | **16** | … | **413** | 100,00 | 100,00 |  | 100 | **∑=250** |  | 100 | **∑=400** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Industries (sorted)** | shares in region 1  (%) (decreasingly) | cumulative sums |  | **Industries (sorted)** | shares by sectors all regions (%) (decreasingly) | cumulative sums |
| **Industry 4** | 56.25 | 56.25 |  | **Industry 2** | 29.30 | 29.30 |
| **Industry 3** | 31.25 | 56.25+31.25 = 87.5 |  | **Industry 1** | 28.81 | 29.30+28.81 = 58.11 |
| **Industry 1** | 6.25 | 87.5+6.25 = 93.75 |  | **Industry 3** | 25.42 | 58.11+25.42 = 83.54 |
| **Industry 2** | 6.25 | 93.75+6.25 = 100 |  | **Industry 4** | 16.46 | 83.54+16.46 = 100.00 |
| **Total** | 100.00 | **∑=337.5** |  | **Total** | 100.00 | **∑=270.94** |

*Source*: Own calculations

Border values for this four-industry regions are 250 for greatest diversity (equal shares) and 400 for least diversity (full concentration). Region 1 with both measures has a middle level of sectoral diversity, as the crude index=337.5 which is between border values of 250 and 400, and the refined index=0.52, which is also between 0 and 1. The overall coverage of the economy with employment is relatively close to uniform (equal) distribution, as the crude index=270.94 is close to the border value 250.

This index is commonly used in many studies. It was well described in a book by Isard (1960).

* + 1. **Krugman dissimilarity and concentration index**

The Krugman dissimilarity index, called also the Krugman specialisation index, introduced by Krugman (1991a, p.76), is based on the standard error concept and measures the standard error of industry shares. As a sectoral dissimilarity index, it is expressed as follows:

which means it is the total by *n* industries, summing up the differences between the share of employment in a given industry *i* in a given region *j* and the share of employment in a given industry in all regions (or reference area). The total is for the absolute values (| |) of differences. The minimum value is 0 (regional structure fully consistent with the reference one), the maximum is . The higher the Krugman dissimilarity Kj index value, the stronger the deviation of regional economic structure from the average reference structure. This overrepresentation of employment in a given industry is often treated as a specialisation, but in fact this is only the regional structure of industries.

This index is also known as the Krugman concentration index Ki, which analyses employment but in another dimension. The concentration index measures industrial structure by regions, thus compares shares of employment in a given industry across regions. Then, as a concentration index, it is expressed as follows:

which means it is total by *m* regions, summing up the differences between the share of employment in a given region *j* in a given industry *i* and the share of employment in a given region in all industries (or reference industry). The minimum value is 0 (industrial structure fully consistent with the reference one), the maximum is .

Below we operationalise both Krugman indices, Kj dissimilarity index for comparing regional structure with the national structure (Table 2.11), as well as the Ki concentration index for testing the equal distribution of firms among regions for a given industry (Table 2.12).

**Table 2.11: Calculations of Krugman Kj­ dissimilarity index on sample data**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **abs(share-share\*)** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| **Industry 1** | |(1/16)-(119/413)| = 0.226 | |(11/78)-(119/413)| = 0.147 | |(21/80)-(119/413)| = 0.026 | |(70/151)-(119/413)| = 0.175 | |(10/53)-(119/413)| = 0.099 | |(6/35)-(119/413)| = 0.117 |
| **Industry 2** | |(1/16)-(121/413)| = 0.230 | |(40/78)-(121/413)| = 0.220 | |(24/80)-(121/413)| = 0.007 | |(40/151)-(121/413)| = 0.028 | |(5/53)-(121/413)| = 0.199 | |(11/35)-(121/413)| = 0.021 |
| **Industry 3** | |(5/16)-(105/413)| = 0.058 | |(13/78)-(105/413)| = 0.088 | |(21/80)-(105/413)| = 0.008 | |(30/151)-(105/413)| = 0.056 | |(35/53)-(105/413)| = 0.406 | |(1/35)-(105/413)| = 0.226 |
| **Industry 4** | |(9/16)-(68/413)| = 0.398 | |(14/78)-(68/413)| = 0.015 | |(14/80)-(68/413)| = 0.010 | |(11/151)-(68/413)| = 0.092 | |(3/53)-(68/413)| = 0.108 | |(17/35)-(68/413)| = 0.321 |
| **Total - Krugman index** | **0.226+0.230**  **+0.058+0.398= 0.912** | **0.147+0.220**  **+0.088+0.015= 0.469** | **0.026+0.007**  **+0.008+0.010= 0.051** | **0.175+0.028**  **+0.056+0.092= 0.351** | **0.099+0.199**  **+0.406+0.108= 0.812** | **0.117+0.021**  **+0.226+0.321= 0.685** |

*Source*: Own calculations

The Krugman Kj dissimilarity index is calculated for every single region. The minimum value in this case is 0 (as always), and the maximum is (2\*(4-1))/4 = 1.5. In the case of the data above, region 3 is closest to the reference distribution (Kj=0.051), and region 1 is the least similar (Kj­=0.912). The degree of dissimilarity is from 3.4% in region 3 (0.051/1.5) up to 60.8% in region 1 (0.912/1.5).

**Table 2.12: Calculations of Krugman Ki­ concentration index on sample data**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **abs(share-share\*)** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total - Krugman index** |
| **Industry 1** | |(1/119)-(16/413)| = 0.030 | |(11/119)-(78/413)| =0.096 | |(21/119)-(80/413)| =0.017 | |(70/119)-(151/413)| =0.223 | |(10/119)-(53/413)| =0.044 | |(6/119)-(35/413)| =0.034 | **0.030+0.096+0.017+0.223**  **+0.044+0.034**  **= 0.445** |
| **Industry 2** | |(1/121)-(16/413)| =0.030 | |(40/121)-(78/413)| =0.142 | |(24/121)-(80/413)| =0.005 | |(40/121)-(151/413)| =0.035 | |(5/121)-(53/413)| =0.087 | |(11/121)-(35/413)| =0.006 | **0.030+0.142+0.005+0.035**  **+0.087+0.006**  **= 0.305** |
| **Industry 3** | |(5/105)-(16/413)| =0.009 | |(13/105)-(78/413)| =0.065 | |(21/105)-(80/413)| =0.006 | |(30/105)-(151/413)| =0.080 | |(35/105)-(53/413)| =0.205 | |(1/105)-(35/413)| =0.075 | **0.009+0.065+0.006+0.080**  **+0.205+0.075**  **= 0.440** |
| **Industry 4** | |(9/68)-(16/413)| =0.094 | |(14/68)-(78/413)| =0.017 | |(14/68)-(80/413)| =0.012 | |(11/68)-(151/413)| =0.204 | |(3/68)-(53/413)| =0.084 | |(17/68)-(35/413)| =0.165 | **0.094+0.017+0.012+0.204**  **+0.084+0.165**  **= 0.576** |

*Source*: Own calculations

The Krugman Ki concentration index is expected to have values in the range between 0 in the case where firms from a given industry were allocated to regions proportionally to the region’s size and (2\*(6-1))/6 = 1.67 when firms from a given industry were allocated to a single region. The above results for Ki evidence that all the industries are similar in terms of disproportions in allocations of firms (and employment) to regions. The degree of concentration equals 0% in the case of Ki=0 and 100% for Ki=1.67. In this case the degree of concentration is from 18% in industry 2 (0.305/1.67) to 34% in industry 4 (0.576/1.67).

As with most of the cluster-based measures, the Krugman index is sensitive to aggregation and disaggregation of sectors. It can be extended by a method of decomposing the Krugman index as a disproportionality measure to flexibly measure sectoral and geographical concentration (usually called specialisation) (Bickenbach & Bode, 2008). It is also programmed - one can find the STATA code for the Krugman dissimilarity index, prepared by Ansari (2013). It is commonly used in many studies as e.g Longhi et al. (2014).

* + 1. **Hallet index**

The Hallet index (2000) in fact resembles the Krugman’s dissimilarity index after slight modifications. Hallet’s index of industrial concentration is as follows:

where *yi,j* is originally the share of Gross Value Added (GVA) in a given region in given sector (and may be substituted with employment) and is the national average summed over all sectors. This indicator compares the absolute difference between the shares in GVA or employment delivered in region in sector, to the over-regional average of this value in the sector, summed up over all sectors. The minimum value is 0, and appears when GVA or employment structures in the region are the same as in the over-regional distribution (like in a country or macro-region). The maximum value is 0.5 when the structures differ significantly (or even completely). The higher the value, the stronger the sectoral concentration. A zero value means equal shares of industries (no over-representation). The unit value represents the extreme, single sector concentration.

The index is mainly known from the analysis for the European Commission, where the Eurostat data from 1980-1995 for 119 regions of EU15 and 17 NACE branches were applied. Its popularity stems from the fact, that this analysis initiated the regional dimension of industry studies. The analysis was accompanied by a concentration index where the clustering measure is based on distances between core cities, and the centrality measure based on market potential, which follows the analysis by Midelfart-Knarvik et al. (2000). The supporting studies on centrality and peripherality were based on papers by Keeble et al. (1988) and Copus (1999).

Operationalisation of Hallet’s sectoral concentration index follows the rules of Krugman’s dissimilarity index. There is only a correction of multiplying the Krugman’s index by ½. This correction keeps the index between 0 and 1.

**Table 2.13: Calculations of Hallet specialisation index**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| **Krugman dissimilarity index** | **0.912** | **0.469** | **0.051** | **0.351** | **0.812** | **0.685** |
| **Hallet sectoral concentration index** | **0.912/2=**  **45.6%** | **0.469/2=**  **23.5%** | **0.051/2=**  **2.6%** | **0.351/2=**  **17.5%** | **0.812/2=**  **40.6%** | **0.685/2=**  **34.2%** |

*Source*: Own calculations

Interpretation of the results is the same as in the case of the Krugman Index. It differs with the changed scale.

* + 1. **Geographic concentration index**

On the same basis as the Krugman index and the Hallet index, one can build a geographic concentration index (OECD, 2009), which is to compare a concentration as a share of any process/activity with a share of territory (area). It is expressed as follows:

where yj is the region’s share in total activity measured, and aj is the share of a region’s area in the whole territory. Its values are as in the case of the Krugman and Hallet indices: 0 in case of no concentration (full diversification) and 1 (100%) in case of full concentration.

**Table 2.14: Calculations of geographical concentration index**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** |
|
| **Industry 1** | 1 | 11 | 21 | 70 | 10 | 6 | **119** |
| **Industry 2** | 1 | 40 | 24 | 40 | 5 | 11 | **121** |
| **Industry 3** | 5 | 13 | 21 | 30 | 35 | 1 | **105** |
| **Industry 4** | 9 | 14 | 14 | 11 | 3 | 17 | **68** |
| **Total** | **16** | **78** | **80** | **151** | **53** | **35** | **413** |
|  |  |  |  |  |  |  |  |
| **share of employment in region** | **16/413=**  **=3.87%** | **78/413=**  **=18.89%** | **80/413=**  **=19.37%** | **151/413=**  **=36.56%** | **53/413=**  **=12.83%** | **35/413=**  **=8.47%** | **1** |
| **area (in units)** | **0.182** | **0.184** | **0.132** | **0.156** | **0.123** | **0.177** | **0.953** |
| **absolute difference** | |0.039-0.182|  =0.143 | |0.189-0.184|  =0.005 | |0.194-0.132|  =0.062 | |0.366-0.156|  =0.210 | |0128-0.123|  =0.005 | |0.085-0.177|  =0.092 | 0.143+0.005+  0.062+0.210+  0.005+0.092=  =0.517 |
| **Geographic concentration index** |  |  |  |  |  |  | 0.517/2=  **0.258** |

*Source*: Own calculations

The above result proves that for this dataset one can observe medium-low geographical concentration, as the index GC=0.517/2=0.258, which in fact is close to geographical diversification. The index is much more informative in international comparisons (even when it is sensitive to the size of a country or a region) or in temporal comparison, when observing the dynamics of inter-regional and inter-sectoral shifts.

* + 1. **Lilien Indicator**

The Lilien indicator (Lilien, 1982) measures the dynamics, the speed of sectoral reallocation (i.e. of employment). It is expressed as follows:

where empijt is the employment in industry i in region j in period t, is the total employment in region j in period t (total of i), is the first-difference operator. It compares the changes of sectoral regional employment to full regional employment over time overs sectors. High values of this index indicate a relatively strong structural shifts between industries. Zero means structural stability over time.

**Table 2.15: Dynamic data for Lilien index**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Periods** | **region 1** | ***Change in region***  ***1*** | **region 2** | ***Change in region***  ***2*** | **region 3** | ***Change in region***  ***3*** | **region 4** | ***Change in region***  ***4*** | **region 5** | ***Change in region***  ***5*** | **region 6** | ***Change in region***  ***6*** |
| **Industry 1** | **Time 0** | 1 | *5/1*  *=5* | 11 | *17/11*  *=1.5* | 21 | *25/21*  *=1.2* | 70 | *88/70*  *=1.3* | 10 | *12/10*  *=1.2* | 6 | *8/6*  *=1.3* |
| **Time 1** | 5 | 17 | 25 | 88 | 12 | 8 |
| **Industry 2** | **Time 0** | 1 | *3/1*  *=3* | 40 | *51/40*  *=1.3* | 24 | *35/24*  *=1.5* | 40 | *49/40*  *=1.2* | 5 | *24/5*  *=4.8* | 11 | *12/11*  *=1.1* |
| **Time 1** | 3 | 51 | 35 | 49 | 24 | 12 |
| **Industry 3** | **Time 0** | 5 | *23/5*  *=4.6* | 13 | *16/13*  *=1.2* | 21 | *30/21*  *=1.4* | 30 | *32/30*  *=1.1* | 35 | *53/35*  *=1.5* | 1 | *3/1*  *=3.0* |
| **Time 1** | 23 | 16 | 30 | 32 | 53 | 3 |
| **Industry 4** | **Time 0** | 9 | *22/9*  *=2.4* | 14 | *27/14*  *=1.9* | 14 | *16/14*  *=1.1* | 11 | *13/11*  *=1.2* | 3 | *6/3*  *=2.0* | 17 | *22/17*  *=1.3* |
| **Time 1** | 22 | 27 | 16 | 13 | 6 | 22 |
| **Total** | **Time 0** | **16** | ***53/16***  ***=3.3*** | **78** | ***111/78***  ***=1.4*** | **80** | ***106/80***  ***=1.3*** | **151** | ***182/151***  ***=1.2*** | **53** | ***95/53***  ***=1.8*** | **35** | ***45/35***  ***=1.3*** |
| **Time 1** | **53** | **111** | **106** | **182** | **95** | **45** |

*Source*: Own calculations

**Table 2.16: Components of Lilien’s index**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 1** |  |  |  | **ab** |
| **time 0** | **time 1** |
|  |  |  | **a** |  | b |  |
| **Industry 1** | 1 | 5 | **5/53=0.094** | log(5)-log(1)=  =0.699 | (0.699-0.52)2=  =0.1792=0.032 | 0.0940.032=0.003 |
| **Industry 2** | 1 | 3 | **3/53=0.057** | log(3)-log(1)=  =0.477 | (0.477-0.52)2=  =-0.0432=0.002 | 0.0570.002=0.000 |
| **Industry 3** | 5 | 23 | **23/53=0.434** | log(23)-log(5)=  =0.663 | (0.663-0.52)2=  =0.1432=0.020 | 0.4340.020=0.009 |
| **Industry 4** | 9 | 22 | **22/53=0.415** | log(22)-log(9)=  =0.388 | (0.388-0.52)2=  =-0.1322=0.017 | 0.4150.017=0.007 |
| **Total** | **16** | **53** | **1.000** | **log(53)-log(16)=**  **=0.520** | **---** | 0.003+0.000+0.009+0.007=0.019 |
|  |  |  |  |  |  | **0.0190.5=0.138** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| **Lilien index** | 0.138 | 0.078 | 0.045 | 0.026 | 0.231 | 0.102 |

*Source*: Own calculations

Comparison of Lilien indices for all regions in periods 0 and 1 reveals that the strongest shifts were in region 5 (σ5=0,231) due to a change in industry 2 by 480% (the strongest individual change), and then in region 1 (σ1=0,138) because of a shift in industry 3 by 460% (second strongest individual shift in a sample). The weakest changes were in region 4 (σ4=0,026) as the individual shifts were also weak.

Following Sava (2016), the Lilien indicator is used officially by the National Bank of Slovakia (together with Krugman’s dissimilarity index, which is used also by the Central European Bank, and in contrast to OECD using the Hannah-Kay (1977) concentration index, which is rather rarely used.).

* + 1. **National averages index (NAI)**

A simple but efficient indicator of sectoral composition is the national averages index (NAI), which is based on a typical concept of squared difference, as it is as follows:

which means it is total by *n* industries, summing up squared differences between the share of employment in a given industry *i* in given region *j* () and share of employment in a given industry in all regions (or reference area) (). It is based on the same concept as Krugman’s dissimilarity index. When the economic structure of a region is the same as in the country (all regions), then NAI=0, which is interpreted as low disparity the between national and regional economy. The higher the disparity, the higher the value of NAI.

To operationalise the above formula, we present NAI for data from Table 2.3.

**Table 2.17: Components of NAI**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Shares by regions and sectors** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Share of industry in economy** |
|
| **Industry 1** | 1/16=  0,063 | 11/78=  0,141 | 21/80  =0,263 | 70/151  =0,464 | 10/53  =0,189 | 6/35  =0,171 | 119/413  =0,288 |
| **Industry 2** | 1/16=  0,063 | 40/78  =0,513 | 24/80  =0,300 | 40/151  =0,265 | 5/53  =0,094 | 11/35  =0,314 | 121/413  =0,293 |
| **Industry 3** | 5/16=  0,313 | 13/78  =0,167 | 21/80  =0,263 | 30/151  =0,199 | 35/53  =0,660 | 1/35  =0,029 | 105/413  =0,254 |
| **Industry 4** | 9/16=  0,563 | 14/78  =0,179 | 14/80  =0,175 | 11/151  =0,073 | 3/53  =0,057 | 17/35  =0,486 | 68/413  =0,165 |
| **Total** | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
|
| **Industry 1** | (0,063-0,288)2/0,288=  =0,177 | (0,141-0,288)2/0,288=  =0,075 | (0,263-0,288)2/0,288=  =0,002 | (0,464-0,288)2/0,288=  =0,107 | (0,189-0,288)2/0,288=  =0,034 | (0,171-0,288)2/0,288=  =0,047 |
| **Industry 2** | (0,063-0,293)2/0,293=  =0,181 | (0,513-0,293)2/0,293=  =0,165 | (0,300-0,293)2/0,293=  =0,000 | (0,265-0,293)2/0,293=  =0,003 | (0,094-0,293)2/0,293=  =0,135 | (0,314-0,293)2/0,293=  =0,002 |
| **Industry 3** | (0,313-0,254)2/0,254=  =0,013 | (0,167-0,254)2/0,254=  =0,030 | (0,263-0,254)2/0,254=  =0,000 | (0,199-0,254)2/0,254=  =0,012 | (0,660-0,254)2/0,254=  =0,649 | (0,029-0,254)2/0,254=  =0,200 |
| **Industry 4** | (0,563-0,165)2/0,165=  =0,961 | (0,179-0,165)2/0,165=  =0,001 | (0,175-0,165)2/0,165=  =0,001 | (0,073-0,165)2/0,165=  =0,051 | (0,057-0,165)2/0,165=  =0,071 | (0,486-0,165)2/0,165=  =0,626 |
| **Total** | **0,177+0,181+**  **0,013+0,961=**  **=1,333** | **0,075+0,165+**  **0,030+0,001=**  **=0,272** | **0,002+0,000+**  **0,000+0,001=**  **=0,003** | **0,107+0,003+**  **0,012+0,051=**  **=0,173** | **0,034+0,135+**  **0,649+0,071=**  **=0,889** | **0,047+0,002+**  **0,200+0,626=**  **=0,875** |

*Source*: Own calculations

Region 3 has the lowest NAI (NAI3=0,003), which confirms its almost identical regional structure with the national one. The highest NAI is in region 1 (NAI1=1,333) where the significant inter-sectoral shifts are visible. [[2]](#footnote-2)

The applications of the NAI index can be found in some studies on regional development e.g. Raj Sharma (2008).

* + 1. **Agglomeration index (industrial dispersion index)**

Franseschi, Mussoni and Pelloni (2009) indicate a popular index of agglomeration, which is based on a comparison of industrial dispersion within and between regions. Its construction is as follows:

where y is the share of employment (in sector *i* and/or region *j*), m is the number of regions. The coefficient Vi is calculated for each sector. In fact, the counter is the dispersion of regional sectoral shares compared with the average sectoral share, summed up by regions, and the denominator is the dispersion of a region’s share compared with the average region’s share, summed up by regions. Values of agglomeration index less than 1 (Vi<1) appear when differences in the sector are smaller than differences in the country, which indicates that the given sector is less geographically concentrated than the overall economy. On the contrary, values of Vi higher than 1 (Vi>1) are for bigger regional than national differences, which proves there is more geographical concentration.

To operationalise the above formula, we present Vi for data from Table 2.3.

**Table 2.18: Components of agglomeration Vi index**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Shares by regions and sectors** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Average share by sectors (average y)** |
|
| **Industry 1** | 1/16=  0.06 | 11/78=  0.14 | 21/80  =0.26 | 70/151  =0.46 | 10/53  =0.19 | 6/35  =0.17 | (0.06+0.14+0.26+0.46+0.19+0.17)/6=0.21 |
| **Industry 2** | 1/16=  0.06 | 40/78  =0.51 | 24/80  =0.30 | 40/151  =0.26 | 5/53  =0.09 | 11/35  =0.31 | (0.06+0.51+0.30+0.26+0.09+0.31)/6=0.26 |
| **Industry 3** | 5/16=  0.31 | 13/78  =0.17 | 21/80  =0.26 | 30/151  =0.20 | 35/53  =0.66 | 1/35  =0.03 | (0.31+0.17+0.26+0.20+0.66+0.03)/6=0.27 |
| **Industry 4** | 9/16=  0.56 | 14/78  =0.18 | 14/80  =0.18 | 11/151  =0.07 | 3/53  =0.06 | 17/35  =0.49 | (0.56+0.18+0.18+0.07+0.06+0.49)/6=0.26 |
| **Average share by regions** | 16/413  =0.04 | 78/413  =0.19 | 80/413  =0.19 | 151/413  =0.37 | 53/413  =0.13 | 35/413  =0.08 | (0.04+0.19+0.19+0.37+0.13+0.08)/6=0.17 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Squared diff. in counter** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** | Sq.root of total/n | Vi |
| **Industry 1** | (0.06-0.21)2  =0.023 | (0.14-0.21)2  =0.005 | (0.26-0.21)2  =0.002 | (0.46-0.21)2  =0.062 | (0.19-0.21)2  =0.001 | (0.17-0.21)2  =0.002 | (0.023+0.005+0.002+  0.062+0.001+0.002)  =0.095 | (0.095/6)0.5  =0.126 | (0.126/0.21)/  (0.104/0.17)  =**0.937** |
| **Industry 2** | (0.06-0.26)2  =0.038 | (0.51-0.26)2  =0.065 | (0.30-0.26)2  =0.002 | (0.26-0.26)2  =0.000 | (0.09-0.26)2  =0.027 | (0.31-0.26)2  =0.003 | (0.038+0.065+0.002+  0.000+0.027+0.003)  =0.135 | (0.135/6)0.5  =0.150 | (0.150/0.26)/  (0.104/0.17)  =**0.928** |
| **Industry 3** | (0.31-0.27)2  =0.002 | (0.17-0.27)2  =0.011 | (0.26-0.27)2  =0.000 | (0.20-0.27)2  =0.005 | (0.66-0.27)2  =0.151 | (0.03-0.27)2  =0.059 | (0.002+0.011+0.000+  0.005+0.151+0.059)  =0.228 | (0.028/6)0.5  =0.195 | (0.195/0.27)/  (0.104/0.17)  =**1.147** |
| **Industry 4** | (0.56-0.26)2  =0.094 | (0.18-0.26)2  =0.006 | (0.18-0.26)2  =0.006 | (0.07-0.26)2  =0.033 | (0.06-0.26)2  =0.040 | (0.49-0.26)2  =0.053 | (0.094+0.006+0.006+  0.033+0.040+0.053)  =0.232 | (0.232/6)0.5  =0.197 | (0.197/0.26)/  (0.104/0.17)  =**1.231** |
|  |  |  |  |  |  |  |  |  |  |
| **Squared diff. in denomi-nator** | (0.04-0.17)2  =0.016 | (0.19-0.17)2  =0.000 | (0.19-0.17)2  =0.001 | (0.37-0.17)2  =0.040 | (0.13-0.17)2  =0.001 | (0.08-0.17)2  =0.007 | (0.016+0.000+0.001+  0.040+0.001+0.007)  =0.065 | (0.065/6)0.5=  0.104 |  |

*Source*: Own calculations

In this case industries 1 and 2 (with V1=0,937 and V2=0,928) are less agglomerated than the whole economy, and sectors 3 and 4 more (with V3=1,147 and V4=1,231). The border value Vi=1 represents the situation where sectoral agglomeration is of the same degree as in the whole economy.

* + 1. **Shannon’s, Theil’s and relative entropy**

Entropy is to measure the deviation of the analysed distribution from full concentration (minimum of entropy) or from full dispersion (maximum of entropy). Full dispersion is mostly given with the uniform distribution, where probabilities of all events are equal. Similarly to the Ogive index, it refers the empirical distribution to the uniform benchmark distribution. Entropy is called a “*measure of the disorder of a system*”, the “*measure of unpredictability of information content*” as well “*the uncertainty associated with a random variable*”. In information technology it is understood as the expected value of the information in the message. In terms of predictability, the lower the entropy, the lower the risk and higher predictability.

In regional and industry analysis, Horowitz and Horowitz (1968) developed the entropy measure of competition H. It is based on Shannon entropy (Shannon, 1948). It is expressed as:

where *s* is the probability of point (discrete) event, and *N=1,2,…,n* is the number of events. In regional studies in the regional concentration measurement, s is the share of employment in a given sector in a given region with reference to full regional employment () and the number of events *n* is the number of industries inside the region. The maximum value of Shannon’s H is for equal probabilities of all events (uniform distribution), is and takes the value . The minimum value of Shannon’s H (Hmin­=0) is available only in the case of full concentration of employment in one single industry, . If there are two equally likely events with s=1/2 then the H=1. One should note that entropy is directly log-proportional to the number of industries, the higher the number of sectors, the higher the entropy measure.

It can be easily transformed from information theory to competitiveness. When there are many firms in the sector or many sectors in a region, the uncertainty grows and the entropy increases. The market with equal share of all firms has the highest degree of competitiveness, and oppositely the more diversified shares of company (with dominating firms) the lower the competitiveness and the lower the risk of operating. An extreme point of a single firm in the industry is a monopoly which operates without competition and risk. High entropy is then for high competitiveness with high *n[[3]](#footnote-3)*. In regional applications, *N* is the number of industry classes and *s* is the share (proportion) of each industry (i.e in the employment). Then the maximum H is obtained at full industrial diversification (equal shares of all industries) and full sectoral concentration (single industry in a region) for a minimum entropy H[[4]](#footnote-4).

The formula above is transformed to give **relative entropy** R:

where H is the measured entropy and is the maximum entropy for finite *n* events. It gives the missing gap between observed and potential entropy, thus the degree of getting to the highest competitiveness, assuming a given number of sectors. The interpretation is as follows: R=1 for equal shares of industries within the region, R=0 is for full concentration of industry.

Theil’s entropy is a measure built on Shannon entropy. By relativising the input, it measures the disorder within the measure – for how much the Theil’s entropy deviates from the maximum Shannon entropy. It is expressed as:

where Hmax is Shannon’s maximum entropy (for equal distribution) and HTheil is Shannon’s entropy for observed data. Thus Theil’s entropy is the gap between observed and maximum entropy, and is called *redundancy*.

The above three measures of entropy, Shannon’s H, Relative H and Theil’s H, can be used in the assessment of geographical concentration of industries, in the cross-section for sectors. This approach would compare the empirical distribution of employment among regions for a given sector with the benchmark one, which assumes equal distribution of firms. The formulas would be the same as above, with the difference that *s* is the share of employment in a given sector in a given region with reference to full employment in the sector () and where number of events *n* is the number of regions in the sample. This could reveal the spatial pattern of business allocation. Shannon’s H would be 0 if all the firms from the sector are in one region, and maximum value if the firms were allocated equally to the regions. This geographical concentration of business in the sector may indicate mechanisms of over-regional agglomeration.

In general pure entropy measures relate only to the studied region with no reference to other regions / sectors or the whole economy. However, when Theil’s entropy is calculated for data for *m* regions and *n* sectors, one can apply the decomposition rule that overall Theil’s index (TTotal) is a weighted average of regional Theil’s indices (Tregional) plus Theil’s measure among the regions (Tinter-regional) - there are then *m+1* components[[5]](#footnote-5). The components can be written as:

*Overall Theil’s index as the difference of Shannon’s max H and Shannon’s empirical H*

where is the ratio of employment in sector *i* in region *j* to total employment (all *i=n* sectors in all *j=m* regions) (empirical share of employment in the given sector in the region to full national employment) (). Thus overall Theil is calculated on *m x n* data (all single cells of a two-dimensional table) as the difference between the maximum Shannon H and empirical Shannon H.

*Theil’s index for single region and all sectors*

*Regional Theil’s index as the weighted average of Theil’s indices*

where is the ratio of regional employment in the whole economy ().

*Inter-regional Theil’s index for shares of regional employment in the whole economy*

where is as above. Decomposition follows the rule that:

Below we operationalise Shannon’s and Theil’s entropy measures for the data from Table 2.3 – as a measures of sectoral concentration within the region (Table 2.19) and as the measure of geographical concentration within the sector (Table 2.20).

**Table 2.19: Entropy measures on sample data – sectoral concentration**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | | **region 2** | | **region 3** | | **region 4** | | **region 5** | | **region 6** | | **equal distrib.** | |
| **s** | **s\*ln s** | **s** | **s\*ln s** | **s** | **s\*ln s** | **s** | **s\*ln s** | **s** | **s\*ln s** | **s** | **s\*ln s** | **s** | **s\*ln s** |
| **Industry 1** | 0.06 | -0.17 | 0.14 | -0.28 | 0.26 | -0.35 | 0.46 | -0.36 | 0.19 | -0.31 | 0.17 | -0.30 | 0.25 | -0.35 |
| **Industry 2** | 0.06 | -0.17 | 0.51 | -0.34 | 0.30 | -0.36 | 0.26 | -0.35 | 0.09 | -0.22 | 0.31 | -0.36 | 0.25 | -0.35 |
| **Industry 3** | 0.31 | -0.36 | 0.17 | -0.30 | 0.26 | -0.35 | 0.20 | -0.32 | 0.66 | -0.27 | 0.03 | -0.10 | 0.25 | -0.35 |
| **Industry 4** | 0.56 | -0.32 | 0.18 | -0.31 | 0.18 | -0.31 | 0.07 | -0.19 | 0.06 | -0.16 | 0.49 | -0.35 | 0.25 | -0.35 |
| **Total /** | **1.00** | **-1.03** | **1.00** | **-1.23** | **1.00** | **-1.37** | **1.00** | **-1.22** | **1.00** | **-0.97** | **1.00** | **-1.12** | **1.00** | **-1.39** |
|  | | | | | | | | | | | | | | |
| **Shannon’s H** | **1.03** | | **1.23** | | **1.37** | | **1.22** | | **0.97** | | **1.12** | | **1.39** | |
| **Relative H** | **1.03/1.39=**  **0.75** | | **1.23/1.39=**  **0.88** | | **1.37/1.39=**  **0.99** | | **1.22/1.39=**  **0.88** | | **0.97/1.39=**  **0.70** | | **1.12/1.39=**  **0.81** | | **1.39/1.39=**  **1.00** | |
| **Theil’s H** | **1.39-1.03=**  **0.35** | | **1.39-0.12=**  **0.16** | | **1.39-1.37=**  **0.02** | | **1.39-1.22=**  **0.17** | | **1.39-0.97=**  **0.41** | | **1.39-1.12=**  **0.27** | | **1.3-1.39=**  **0** | |

*Source*: Own calculations

The Shanon entropy measure was calculated for every single region (Table 2.19). Component *s* is the share of industry in the region (i.e. for industry 1 in region 1 *s*=1/16=0.06), and component *s\*ln(s)* is counted then as 0.06\*ln(0.06)=0.06\*(-2.77)=-0.17. The total for the region is multiplied by (-1) to change the sign to positive, which gives Shannon’s entropy.

The last column (equal distribution) shows the maximum entropy (for full diversification) and Hmax=1.39. The most diversified is region 3 (Hregion\_3=1,39), and the least diversified is region 5 (Hregion\_3=0.97). The minimum entropy is Hmin=0 for full allocation to a single industry in a region. Relative entropy, understood as the degree of regional diversity, is from 0.7 (medium strong) in region 5 to 0.99 (very strong) in region 3.

The higher Theil’s entropy the higher the degree of uneven distribution, as the gap between equal and empirical distributions is high. A Theil index close to 0 is for equal (uniform) distribution.

**Table 2.20: Entropy measures on sample data – geographical concentration**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** |  | **Shannon H** | **Relative H** | **Theil's**  **H** |
| **Industry 1** | **s** | 1/119=0,01 | 11/119=0,09 | 21/119=0,18 | 70/119=0,59 | 10/119=0,08 | 6/119  =0,05 | **1,00** |  | 1,24 | 1,24/1,79=0,69 | 1,79-1,24  =0,55 |
| **s\*ln s** | -0,04 | -0,22 | -0,31 | -0,31 | -0,21 | -0,15 | **-1,24** |  |
| **Industry 2** | **s** | 1/121=0,01 | 40/121=0,33 | 24/121=0,20 | 40/121=0,33 | 5/121  =0,04 | 11/121=0,09 | **1,00** |  | 1,44 | 1,44/1,79=0,80 | 1,79-1,44  =0,35 |
| **s\*ln s** | -0,04 | -0,37 | -0,32 | -0,37 | -0,13 | -0,22 | **-1,44** |  |
| **Industry 3** | **s** | 5/105=0,05 | 13/105=0,12 | 21/105=0,20 | 30/105=0,29 | 35/105=0,33 | 1/105  =0,01 | **1,00** |  | 1,49 | 1,49/1,79=0,83 | 1,79-1,49  =0,30 |
| **s\*ln s** | -0,14 | -0,26 | -0,32 | -0,36 | -0,37 | -0,04 | **-1,49** |  |
| **Industry 4** | **s** | 9/68  =0,13 | 14/68=0,21 | 14/68  =0,21 | 11/68  =0,16 | 3/68  =0,04 | 17/68  =0,25 | **1,00** |  | 1,70 | 1,70/1,79=0,95 | 1,79-1,70  =0,09 |
| **s\*ln s** | -0,27 | -0,33 | -0,33 | -0,29 | -0,14 | -0,35 | **-1,70** |  |
| **equal dist** | **s** | 1/6  =0,17 | 1/6  =0,17 | 1/6  =0,17 | 1/6  =0,17 | 1/6  =0,17 | 1/6  =0,17 | **1,00** |  | 1,79 | 1,79/1,79=1,00 | 1,79-1,79  =0,00 |
| **s\*ln s** | -0,30 | -0,30 | -0,30 | -0,30 | -0,30 | -0,30 | **-1,79** |  |

*Source*: Own calculations

The Shanon entropy measure was calculated for every single sector (Table 2.20). Component *s* is the share of the region in the industry (i.e. for industry 1 in region 1 *s*=1/119=0.008), and component *s\*ln(s)* is counted then as 0.008\*ln(0.008) = 0.008\*(-4.779) = -0.040. The total for the industry is multiplied by (-1) to change the sign to positive, which gives Shannon’s entropy. The last row (equal distribution) shows the maximum entropy (for fully equal spatial allocation) and Hmax=1.792. The most equal distribution of business among regions is in sector 4 (H­sector 4=1,697), and the highest geographical concentration is in sector 1 (Hsector 1=1,237). The minimum entropy is Hmin=0 is for full allocation of sectoral activity to region. Relative entropy, understood as the degree of sectoral diversity, is from 0.69 (medium strong) in sector 1 to 0.95 (very strong) in industry 4. The higher the Theil’s entropy the higher degree of uneven distribution, as the gap between equal and empirical distributions is high. Theil index close to 0 is for equal (uniform) distribution. Sector 4 is very close to uniform distribution of business.

In the case of decomposition one can observe how much total equality of distribution results from inter-regional economy shifts (Tregional) as well as the size of regional economies (). The decomposition of Theil’s index can be done as follows:

**Table 2.21: Shares of region-industry employment in full national employment**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** |
|
| **Industry 1** | 1/413=  0.002 | 11/413=  0.027 | 21/413=  0.051 | 70/413=  0.169 | 10/413=  0.024 | 6/413=  0.015 |  |
| **Industry 2** | 1/413=  0.002 | 40/413=  0.097 | 24/413=  0.058 | 40/413=  0.097 | 5/413=  0.012 | 11/413=  0.027 |  |
| **Industry 3** | 5/413=  0.012 | 13/413=  0.031 | 21/413=  0.051 | 30/413=  0.073 | 35/413=  0.085 | 1/413=  0.002 |  |
| **Industry 4** | 9/413=  0.022 | 14/413=  0.034 | 14/413=  0.034 | 11/413=  0.027 | 3/413=  0.007 | 17/413=  0.041 |  |
| **Total** |  |  |  |  |  |  | **1.000** |
| **Share of region** | **16/413=**  **0.039** | **78/413=**  **0.189** | **80/413=**  **0.194** | **151/413=**  **0.366** | **53/413=**  **0.128** | **35/413=**  **0.085** | **1.000** |

*Source*: Own calculations

0.193

=0.184+0.193=0.376

This decomposition shows that in measurement of the total regional / sectoral inequality the inside-regional distribution inequalities are slightly weaker (ca.49%) than inter-regional differences in regional capacity (employment) (51%).

Entropy was introduced to economics by Horowitz & Horowitz (1968) in the analysis of the brewing industry[[6]](#footnote-6). It was to measure concentration, which was treated analogously to industry competition. The entropy measure can be used as a measure of competitiveness, explaining the performance of a sector. Nawrocki and Carter (2010) apply an entropy measure together with the Herfindahl index to determine the importance of competition on a company’s performance. There is a large amount of literature on the market share, concentration and competitiveness which is a-spatial by nature[[7]](#footnote-7). The applications of the entropy measure can be found in e.g. Raj Sharma (2008).

There is also a method of decomposing general and Theil Entropy as a disproportionality measure to flexibly measure sectoral and geographical concentration (usually called specialisation) (Bickenbach & Bode, 2008), as well the absolute and relative Theil index (Bickenbach, Bode & Krieger-Bode, 2012). Also (Cutrini, 2009) proposed a modification of entropy by decomposition and weighting to measure the entropy of overall localisation, which allows for tracking for sectoral and geographical concentration and the divergence in agglomeration patterns. The measure introduced by Cutrini (2009) stems from modern regional science needs. The starting point is that there is strong need to see the regional structure of industries in relation to national one, which scales the different spatial scales together.

* + 1. **Kullback-Leibler divergence (KLD)**

Entropy measures were developed also as relative entropy. This solution was proposed by Kullback and Leibler (1951), to assess two distributions and the direct divergence between them. In information theory it is understood as the information lost when A is approximating B. In regional science it was introduced recently (Mori et al., 2005) to compare two distributions of economic structure: regional and national ones. Kullback-Leibler divergence (KLD) can be expressed as follows:

where *qi* stands for the share of employment in sector *i* in the country and *sij* is defined as above as the share of employment in a given sector in a given region with reference to the full regional employment (). KLD is thus the sum by *n* sectors (*i=1,2,..,n)* for a given region *j*. It is to measure the difference in economic stucture on the regional and national level. Because of the construction, KLD is always non-negative. A minimum value of KLD=0 is for full similarity of distributions. The higher the divergence of two distributions, the higher the value of KLD. It might be indefinite for mono-industry and situations with an industry absent in a region. This is because *ln(0)* is indefinite. Thus the precondition of analysis with KLD would be that all national industries are represented in a region. Mori et al. (2005) apply the rule that *0⋅ln(0)=0*, which makes the KLD definite in all situations (see Mori et al., 2005 for more properties of KLD).

KLD, as with the Krugman dissimilarity index and Gini index can be calculated by regions or by industries. The one above is for a given region by industries. It can be also, as originally proposed by Mori et al. (2005), for a single industry by regions. It is then to measure the “*complete spatial dispersion*” or the “*degree of localization*”. The benchmark distrubution is then the uniform distribution of firms among regions, which gives an equal probability of a business location of a given industry in the regions analysed. Then KLD is expressed as follows:

where *qj* is the expected value of employment in region *j* in the industry analysed, and *sij* is the share of employment in a given sector in a given region with reference to full industrial employment (). The properties of index are the same as above. Interpretation is conducted as inter-regional comparison of a given industry. KLDi close to 0 means *complete spatial dispersion* of business. The higher the value of KLDi the higher the degree of localization, which can be understood as the degree of regional concentration. It is worth noting, that KLD with uniform benchmark as above gives the same result as Theil’s H.

Below we operationalise the KLD index for the data from Table 2.3. Table 2.22 presents a measure of industrial concentration KLDj for each region (by sectors), where the benchmark is the national economic structure. Table 2.23 presents a measure of regional concentration KLD­i for each industry (by regions), where the benchmark is the uniform (equal) distribution of firms in the industry between *m* regions.

**Table 2.22: Kullback-Leibler divergence for regions (KLDj)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
|
| **Industry 1** | (1/16)⋅ln((1/16)/ (119/413))=  = -0.096 | (11/78)⋅ln((11/ 78)/(119/413))=  = -0.101 | (21/80)⋅ln((21/ 80)/(119/413))=  = -0.024 | (70/151)⋅ln((70/151)/(119/413))  = 0.220 | (10/53)⋅ln((10/ 53)/(119/413))=  = -0.080 | (6/35)⋅ln((6/35)/(119/413))=  = -0.089 |
| **Industry 2** | (1/16)⋅ln((1/16)/ (121/413))=  = -0.097 | (40/78)⋅ln((40/ 78)/(121/413))=  = 0.287 | (24/80)⋅ln((24/ 80)/(121/413))=  = 0.007 | (40/151)⋅ln((40/151)/(121/413))= -0.027 | (5/53)⋅ln((5/53)/(121/413))=  = -0.107 | (11/35)⋅ln((11/ 35)/(121/413))=  = 0.022 |
| **Industry 3** | (5/16)⋅ln((5/16)/ (105/413))=  = 0.064 | (13/78)⋅ln((13/ 78)/(105/413))=  = -0.070 | (21/80)⋅ln((21/ 80)/(105/413))=  = 0.008 | (30/151)⋅ln((30/151)/(105/413))  = -0.049 | (35/53)⋅ln((35/ 53)/(105/413))=  = 0.630 | (1/35)⋅ln((1/35)/(105/413))=  = -0.062 |
| **Industry 4** | (9/16)⋅ln((9/16)/ (68/413))=  = 0.691 | (14/78)⋅ln((14/ 78)/(68/413))=  = 0.015 | (14/80)⋅ln((14/ 80)/(68/413))=  = 0.011 | (11/151)⋅ln((11/151)/(68/413))=  = -0.059 | (3/53)⋅ln((3/53)/(68/413))=  = -0.060 | (17/35)⋅ln((17/ 35)/(68/413))=  = 0.525 |
| **Total** | **-0.096+-0.097**  **+0.064+0.691**  **= 0.563** | **-0.101+0.287**  **-0.07+0.015**  **= 0.131** | **-0.024+0.007**  **+0.008+0.011**  **= 0.002** | **0.220-0.027**  **-0.049-0.059**  **= 0.085** | **-0.08-0.107**  **+0.63-0.06**  **= 0.383** | **-0.089+0.022**  **-0.062+0.525**  **= 0.396** |

*Source*: Own calculations

The interpretation of KLDj­ calculated above is that region 3 is most similar to the national distribution of activity, and can be treated as the “average region”, as the KLDj=3=0,002 is close to KLDj=benchmark=0. In other regions, the distribution of economic activity differs. The strongest structural differences are between country and region 1 (KLDj=1=0,563).

**Table 2.23: Kullback-Leibler divergence for industries (KLDi)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** |
| **Industry 1** | (1/119)⋅ln(  (1/119)/(1/6))  = -0,03 | (11/119)⋅ln(  (11/119)/ (1/6))  = -0,05 | (21/119)⋅ln(  (21/119)/ (1/6))  = 0,01 | (70/119)⋅ln(  (70/119)/ (1/6))  = 0,74 | (10/119)⋅ln(  (10/119)/ (1/6))  = -0,06 | (6/119)⋅ln(  (6/119)/(1/6))  = -0,06 | **-0,03-0,05+**  **0,01+0,74**  **-0,06-0,06**  **=0,55** |
| **Industry 2** | (1/121)⋅ln(  (1/121)/(1/6))  = -0,02 | (40/121)⋅ln(  (40/121)/ (1/6))  = 0,23 | (24/121)⋅ln(  (24/121) /(1/6))  = 0,03 | (40/121)⋅ln(  (40/121)/ (1/6))  = 0,23 | (5/121)⋅ln(  (5/121)/(1/6))  = -0,06 | (11/121)⋅ln(  (11/121)/ (1/6))  = -0,06 | **-0,02+0,23**  **+0,03+0,23**  **-0,06-0,06**  **=0,35** |
| **Industry 3** | (5/105)⋅ln(  (5/105)/(1/6))  = -0,06 | (13/105)⋅ln(  (13/105)/ (1/6))  = -0,04 | (21/105)⋅ln(  (21/105)/ (1/6))  = 0,04 | (30/105)⋅ln(  (30/105)/ (1/6))  = 0,15 | (35/105)⋅ln(  (35/105)/ (1/6))  = 0,23 | (1/105)⋅ln(  (1/105)/(1/6))  = -0,03 | **-0,06-0,04**  **+0,04+0,15**  **+0,23-0,03**  **=0,30** |
| **Industry 4** | (9/68)⋅ln(  (9/68)/(1/6))  = -0,03 | (14/68)⋅ln(  (14/68)/(1/6))  = 0,04 | (14/68)⋅ln(  (14/68)/(1/6))  = 0,04 | (11/68)⋅ln(  (11/68)/(1/6))  = 0,00 | (3/68)⋅ln(  (3/68)/(1/6))  = -0,06 | (17/68)⋅ln(  (17/68)/(1/6))  = 0,10 | **-0,03+0,04**  **+0,04+0,00**  **-0,06+0,10**  **=0,09** |

*Source*: Own calculations

The interpretation of KLDi­ calculated above is that industry 4 represents the highest spatial dispersion, and business is almost uniformely alocated among regions, as the KLDi=4=0,09 is closest to the KLDi=benchmark=0. In other industries, the distribution of economic activity differs from the equal one. The highest degree of localization appears in industry 1 (KLDi=1=0,55).

This indicator is used together with other measures of diversity and specialization, e.g. by Simonen et al. (2015).

* + 1. **Bruelhart’s & Traeger’s concentration entropy measure**

Bruelhart and Traeger (2005) proposed a measure which is based on the entropy concept, but is decomposable, and supported by the boostrap test. It allows for the between- and within-country comparisons as well quantifying the contribution of sectors to overall concentration. Starting from generalized entropy, they develop the basic entropy indices, GE as *Theil index* and CV as *coefficient of variation*, which are as follows:

and

where *nj* is the weighting variable (i.e. total regional employment, ), N is total national employment, is the share of regional (*j*) sectoral (*i*) employment in regional employment, (with ) is the share of sectoral employment in full national employment[[8]](#footnote-8). The GE index reflects the sectoral concentration, with regard to differences in regional activity (employment).

Bruelhart & Traeger (2005) introduce “*relative concentration*” to capture the degree to which sectors are concentrated relative to the geographic distribution of aggregate activity, and “*topographic concentration*” where they measure the degree to which sectors are concentrated relative to physical space. This difference is reflected in which can be unweighted for topographic concentration or weighted for relative concentration.

Below we operationalise the Bruelhart & Traeger’s entropy measures for the data from Table 2.3.

**Table 2.24: Components of GE Bruelhart & Traeger’s (2005) entropy measures**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Region 1** | **Region 2** | **Region 3** | **Region 4** | **Region 5** | **Region 6** |
|
| **Industry 1** | (1/16)/  (119/413)  =0.217 | (11/78) /  (119/413)  =0.489 | (21/80) /  (119/413)  =0.911 | (70/151) /  (119/413)  =1.609 | (10/53) /  (119/413)  =0.655 | (6/35) /  (119/413)  =0.595 |
| **Industry 2** | (1/16)/  (121/413)  =0.213 | (40/78) /  (121/413)  =1.750 | (24/80) /  (121/413)  =1.024 | (40/151) /  (121/413)  =0.904 | (5/53) /  (121/413)  =0.322 | (11/35) /  (121/413)  =1.073 |
| **Industry 3** | (5/16)/  (105/413)  =1.229 | (13/78) /  (105/413)  =0.656 | (21/80) /  (105/413)  =1.033 | (30/151) /  (105/413)  =0.781 | (35/53) /  (105/413)  =2.597 | (1/35) /  (105/413)  =0.112 |
| **Industry 4** | (9/16)/  (68/413)  =3.416 | (14/78) /  (68/413)  =1.090 | (14/80) /  (68/413)  =1.063 | (11/151) /  (68/413)  =0.442 | (3/53) /  (68/413)  =0.344 | (17/35) /  (68/413)  =2.950 |
|  |  |  |  |  |  |  |
| **n(j)/N** | 16/413=0.04 | 78/413=0.19 | 80/413=0.19 | 151/413=0.37 | 53/413=0.13 | 35/413=0.08 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total (by industries)** |
|
| **Industry 1** | 0.04⋅0.217⋅  log(0.217)  = -0.01 | 0.19⋅0.4897⋅  log(0.489)  = -0.03 | 0.19⋅0.911⋅  log(0.911)  = -0.01 | 0.37⋅1.609⋅  log(1.609)  = 0.121 | 0.13⋅0.655⋅  log(0.655)  = -0.02 | 0.08⋅0.595⋅  log(0.595)  = -0.01 | **(-0.01)+(-0.03)+(-0.01)+**  **0.121+(-0.02)+(-0.01)=**  **= -0.053** |
| **Industry 2** | 0.04⋅0.213⋅  log(0.213)  =-0.01 | 0.19⋅1.750⋅  log(1.750)  = 0.08 | 0.19⋅1.024⋅  log(1.024)  = 0.002 | 0.37⋅0.904⋅  log(0.904)  = -0.01 | 0.13⋅0.322⋅  log(0.322)  = -0.02 | 0.08⋅1.073⋅  log(1.073)  = 0.003 | **(-0.01)+0.08+0.002+**  **(-0.01)+(-0.02)+0.003=**  **=0.045** |
| **Industry 3** | 0.04⋅1.229⋅  log(1.229)  =0.004 | 0.19⋅0.656⋅  log(0.656)  = -0.02 | 0.191.033⋅  log(1.033)  = 0.003 | 0.37⋅0.781⋅  log(0.781)  = -0.03 | 0.13⋅2.597⋅  log(2.597)  = 0.138 | 0.08⋅0.112⋅  log(0.112)  = -0.01 | **0.004+(-0.02)+0.003+**  **(-0.03)+0.138+(-0.01)= =0.083** |
| **Industry 4** | 0.04⋅3.416⋅  log(3.416)  =0.071 | 0.19⋅1.090⋅  log(1.090)  = 0.008 | 0.19⋅1.063⋅  log(1.063)  = 0.005 | 0.37⋅0.442⋅  log(0.442)  = -0.06 | 0.13⋅0.344⋅  log(0.344)  = -0.02 | 0.08⋅2.950⋅  log(2.950)  = 0.117 | **0.071+0.008+0.005+**  **(-0.06)+(-0.02)+0.117= =0.123** |

*Source*: Own calculations

Following the above calculations, GE indices for sectors are: GEsector1=0.053, GEsector2=0.045, GEsector3=0.083, GEsector4=0.123. This means that Sector 4 is most concentrated (highest value of GE) and Sector 2 is most dispersed. For dynamic datasets (as panels) one can calculate GE indices for sectors and periods, and then compare changes over time to conclude on concentration patterns.

* 1. **Cluster-based measures depending only on size of companies**
     1. **Herfindahl index**

One of the best known indices is the Herfindal index (or sometimes Hirschman-Herfindahl index). It can be applied simply to single firms to measure the monopolistic position of firms and assess the market organization. In a regional context it is applied to the distribution of shares of industries within a region (or between regions) to cover economic diversity. It is expressed as follows:

where is the share of employment (A high value of H results from uneven distribution and thus indicates at high degree of concentration (in the case of a firm’s monopoly), while a low value of H is for even distribution and/or for a high level of competition. The H index may be between 0 (when many units are evenly distributed) and 1 (one significant share covering most of the activity). Decreasing values prove increasing diversification, and oppositely increasing values of H are for monopolisation or extreme concentration.

The Herfindahl index, similar to Ogive or entropy, examines the region only, and does not detect overall patterns. The extreme points for regional full diversification or concentration can be calculated theoretically, using the formula *(()2)⋅n* for *n* sectors. The upper bound of a mono-industry is always 1. The lower bound for n=10 sectors is 0.1 and for 100 000 sectors would be 0.00001.

Below we operationalize the Herfindahl index for the data from Table 2.3. For four sectors assumed, the Herfindahl index is between 25% *(1/4)2⋅4* for even distribution of sectors and 100% for a one-sector economy.

**Table 2.25: Components of Herfindahl index**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** |
|
| **Industry 1** | (1/16)2=  0.4% | (11/78)2=  2.0% | (21/80)2=  6.9% | (70/151)2=  21.5% | (10/53)2=  3.6% | (6/35)2=  2.9% | (119/413)2=  8.3% |
| **Industry 2** | (1/16)2=  0.4% | (40/78)2=  26.3% | (24/80)2=  9.0% | (40/151)2=  7.0% | (5/53)2=  0.9% | (11/35)2=  9.9% | (121/413)2=  8.6% |
| **Industry 3** | (5/16)2=  9.8% | (13/78)2=  2.8% | (21/80)2=  6.9% | (30/151)2=  3.9% | (35/53)2=  43.6% | (1/35)2=  0.1% | (105/413)2=  6.5% |
| **Industry 4** | (9/16)2=  31.6% | (14/78)2=  3.2% | (1/16)2=  3.1% | (11/151)2=  0.5% | (3/53)2=  0.3% | (17/35)2=  23.6% | (68/413)2=  2.7% |
| **Total** | **0.4%+0.4%+9.8%+31.6%=42.2%** | **2.0%+26.3%+2.8%+3.2%=34.3%** | **69.%+9.0%+6.9%+3.1%**  **=25.8%** | **21.5%+7.0%+3.9%+0.5%=33.0%** | **3.6%+0.9%+43.6%+0.3%=48.4%** | **2.9%+9.9%+0.1%+23.6%=36.5%** | **8.3%+8.6%+6.5%+2.7%**  **=26.1%** |

*Source*: Own calculations

The results prove that the general distribution of activity between sectors is close to the uniform one (H=26.1%), as well the distribution in region 3 (H=25.8%). The most asymmetric distribution (highest sectoral concentration) appears in region 5 (H=48.4%), because of a 66% share of industry 3.

Guimaraes, Figueiredo & Woodward (2011) show its insensitivity to spatial permutation of values. They also present a version of the Herfindhal index, which includes Moran’s I. The Herfindahl (or Herfindahl-Hirschman index) is continuously used as a good measure of sectoral concentration in many studies (see e.g. Rodriguez-Pose et al., 2013).

* + 1. **Relative and Absolute Diversity and Specialisation Indices**

The Relative Diversity Index (RDI) is used together with the Absolute Diversity Index (ADI) to assess diversity in the economy. They are confronted with simple indices: Relative Specialization Index (RSI) and Dissimilarity Index (DIS).

The Relative Diversity Index (RDI), introduced by Duranton and Puga (2000), is calculated as an inverse Dissimilarity Index DIS (Krugman dissimilarity index), while DIS compares the regional and national structure by summing up (by industries for one region) absolute values of differences between the regional share of industry and the national one.

where *Sij* is the share of employment in industry *i* in region *j* in total regional employment, while *Si* is the same share but on national level (national employment in sector to total national employment). DIS compares the structure of industries in a given region with the national structure. It can be between 0 and , while DIS=0 means the full similarity (zero dissimilarity), and the maximum value is the opposite – full dissimilarity. The maximum asymptotic value is 2 (e.g. for 1 000 000 sectors this limit is 1,9999999998), and for small *n* it is less (as 1,333n=3, 1,75n=8, 1,9n=20 etc.).

The RDI is counted as the inverse of DIS:

The more similar the regional and national economies, the smaller the DIS and the higher the RDI. Values of DIS close to 0 indicate similar structures of regional and national economies.

Similar to RDI is the Absolute Diversity Index (ADI), also introduced by Duranton and Puga (2000). The difference is that ADI takes the inverse of the Hirshman-Herfindahl HH index:

The higher the diversification the higher the HH, and the lower the ADI. As the HH is limited asymptotically to 0 and 1, thus the ADI is limited from 0 to infinity.

In a similar fashion one can build the sectoral concentration indices. Introduced also by Duranton and Puga (2000), the Relative Specialisation Index (RSI) for region *j*, is simply calculated as the maximum of LQij­ for this region (by sectors *i*). It can be written as:

while , where Sij and Si­ are as defined above. The higher the over-representation of sector in the region, the higher is the regional RSI.

Calculations for the example data are as below:

**Table 2.26: RDI, ADI and RSI indicators**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| **Krugman Index (DIS)** | 0.912 | 0.469 | 0.051 | 0.351 | 0.812 | 0.685 |
| **Relative Diversity Index (RDI) (inverse Krugman)** | 1.096 | 2.131 | 19.504 | 2.850 | 1.231 | 1.460 |
|  |  |  |  |  |  |  |
| **Herfindahl idnex (HH) for regions** | 0.422 | 0.343 | 0.258 | 0.330 | 0.484 | 0.365 |
| **Absolute Diversity Index (ADI) (inverse HH)** | 2.370 | 2.917 | 3.869 | 3.032 | 2.067 | 2.740 |
|  |  |  |  |  |  |  |
| **RSI (max LQ)** | 3.416 | 1.750 | 1.063 | 1.609 | 2.597 | 2.950 |

*Source*: Own calculations

Interpretation of these indices is very similar to that for the basic indicators. The main goal is to keep the scale similar for slightly different values and make the indicators more sensitive to outlier values.

These indices were calculated for example for Turkey (Peker, 2012), as well for Finland and its high-tech industries (Simonen et al., 2015), where they test relations between specialization and economic growth.

* 1. **Cluster-based measures depending on *n x m* matrix and size of companies**
     1. **Ellison & Glaeser index**

One of the most often used indices of spatial / geographical concentration of industry among the regions is the Ellison & Glaeser index (1997) (EG). It compounds the effect of natural advantages as well as industry spillovers (see Ellison & Glaeser, 1997, 1999; Kominers, 2008). It can be used to look for industrial clusters (Cassey & Smith, 2015). It is expressed as follows:

where *si* is the share of employment in the industry in the region, *xi* is the share of employment in the region, H is the industrial Herfindahl index on the plant level (for *X* firms of size *z*) in all regions (). In fact reflects the similarity of industrial and regional distribution (as a kind of taxonomy), is an inverse regional Herfindahl (for shares of regional employment) and H is business industrial Herfindahl (for size of companies). The Ellison & Glaeser index might be also expressed as:

where is called the spatial Gini index.

The Ellison-Glaeser index can take both negative and positive values. EG=0 is for a random distribution. Positive values prove that there is an industrial concentration. Most of the literature gives the critical values for EG interpretation: EG<0.02 is for low concentration, EG values between 0.02-0.05 are for intermediate concentration, EG>0.05 are for high concentration of a given sector between regions. Negative values are typical for uniform geographic coverages or distant co-locations, with significant spatial separation (Glaeser, 2010, p.154[[9]](#footnote-9)). The literature remains rather silent on the fact that when there are few big plants in the sector, then H becomes very large (close to 1) and EG might be less than -1 or even -2. Thus positive EG appears usually when H is very small, which happens when there are many small companies on the market and no big firms.

Below we operationalise the Ellison-Glaeser index for the data from Table 2.3. There is additional information on number and size of firms in each sector/industry, to cover the Herfindahl part of the index. We make two different assumptions on the distribution of firms’ size: a) that there exist some big firms (which in consequence makes the H big – close to 1 and EG very negative); b) that there are no big firms (which in consequence makes the H small 0 close to 0 and EG very positive).

**Table 2.27: Components of Ellison-Glaeser index**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** |
| **Industry 1** | [(1/119)-(16/413)]2=  =0.001 | [(11/119)-(78/413)]2=  0.009 | [(21/119)-(80/413)]2=  0.000 | [(70/119)-(151/413)]2=  0.050 | [(10/119)-(53/413)]2=  0.002 | [(6/119)-(35/413)]2=  0.001 | **0.001+0.009+**  **0.000+0.050+**  **0.002+0.001=**  **=0.063** |
| **Industry 2** | [(1/121)-(16/413)]2=  0.001 | [(40/121)-(78/413)]2=  0.020 | [(24/121)-(80/413)]2=  0.000 | [(40/121)-(151/413)]2=  0.001 | [(5/121)-(53/413)]2=  0.008 | [(11/121)-(35/413)]2=  0.000 | **0.001+0.020+**  **0.000+0.001+**  **0.008+0.000=**  **=0.030** |
| **Industry 3** | [(5/105)-(16/413)]2=  0.000 | [(13/105)-(78/413)]2=  0.004 | [(21/105)-(80/413)]2=  0.000 | [(30/105)-(151/413)]2=  0.006 | [(35/105)-(43/413)]2=  0.042 | [(1/105)-(35/413)]2=  0.006 | **0.000+0.004+**  **0.000+0.006+**  **0.042+0.006=**  **=0.058** |
| **Industry 4** | [(9/68)-(16/413)]2=  0.009 | [(14/68)-(78/413)]2=  0.000 | [(14/68)-(80/413)]2=  0.000 | [(11/68)-(151/413)]2=  0.042 | [(3/68)-(53/413)]2=  0.007 | [(17/68)-(35/413)]2=  0.027 | **0.009+0.000+**  **0.000+0.042+**  **0.007+0.027=**  **=0.085** |
|  | **[16/413]2**  **=0.002** | **[78/413]2**  **=0.036** | **[80/413]2**  **=0.038** | **[151/413]2**  **=0.134** | **[53/413]2**  **=0.016** | **[35/413]2**  **=0.007** | **0.002+0.036+**  **0.038+0.134+**  **0.016+0.007=**  **=0.232** |

*Source*: Own calculations

Assumption A on the distribution of firms’ size – few big firms (closer to monopoly)

One should assume an allocation of companies to cover the employment:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Size of companies** | **100** | **50** | **10** | **1** | **Total employment** |
| **Number of firms** | **Industry 1** | 0 | 1 | 6 | 9 | 0⋅100+1⋅50+6⋅10+9⋅1=119 |
| **Industry 2** | 0 | 0 | 11 | 11 | 0⋅100+0⋅50+11⋅10+11⋅1=121 |
| **Industry 3** | 0 | 0 | 9 | 15 | 0⋅100+0⋅50+9⋅10+15⋅1=105 |
| **Industry 4** | 0 | 0 | 4 | 28 | 0⋅100+0⋅50+4⋅10+28⋅1=68 |

*Source*: Own calculations

Then the Herfindahl index, which is the part of Ellison-Glaeser index, would be as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **100** | **50** | **10** | **1** | **Total H** |
| **industry1** | (100/119)2⋅0=  =0.00 | (50/119)2⋅1=  =0.177 | (10/119)2⋅6=  =0.042 | (1/119)2⋅9=  =0.001 | **0.00+0.18+0.04+0.00=**  **=0.220** |
| **industry2** | (100/121)2⋅0=  =0.00 | (50/121)2⋅0=  =0.00 | (10/121)2⋅11=  =0.075 | (1/121)2⋅11=  =0.001 | **0.00+0.00+0.08+0.00=**  **=0.076** |
| **industry3** | (100/105)2⋅0=  =0.00 | (50/105)2⋅0=  =0.00 | (10/105)2⋅9=  =0.082 | (1/105)2⋅15=  =0.001 | **0.00+0.00+0.07+0.00=**  **=0.083** |
| **industry4** | (100/68)2⋅0=  =0.00 | (50/68)2⋅0=  =0.00 | (10/68)2⋅4=  =0.087 | (1/68)2⋅28=  =0.006 | **0.00+0.00+0.11+0.00=**  **=0.093** |

*Source*: Own calculations

Thus, finally the Ellison-Glaeser index would be:

|  |  |  |  |
| --- | --- | --- | --- |
|  | counter | nominator | Ellison-Glaeser index |
| **Industry 1** | 0.063-(1-0.232)⋅0.22 = - 0.105 | (1-0.232)⋅(1-0.22) = 0.599 | -0.105/0.599 = - 0.176 |
| **Industry 2** | 0.030-(1-0.232)⋅0.076 = - 0.028 | (1-0.232)⋅(1-0.076) = 0.710 | -0.028/0.710 = - 0.040 |
| **Industry 3** | 0.058-(1-0.232)⋅0.083 = - 0.005 | (1-0.232)⋅(1-0.083) = 0.704 | -0.005/0.704 = - 0.008 |
| **Industry 4** | 0.085-(1-0.232)⋅0.093 = 0.014 | (1-0.232)⋅(1-0.093) = 0.697 | 0.014/0.697 = 0.020 |
| **Total** | --- | --- | --- |

*Source*: Own calculations

Assumption B on the distribution of firms’ size – many small firms (closer to competition)

One should assume an allocation of companies to cover the employment:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Size of companies** | **100** | **50** | **10** | **1** | **Total employment** |
| **Number of firms** | **Industry 1** | 0 | 0 | 5 | 69 | 0⋅100+0⋅50+5⋅10+69⋅1=119 |
| **Industry 2** | 0 | 0 | 4 | 81 | 0⋅100+0⋅50+4⋅10+81⋅1=121 |
| **Industry 3** | 0 | 0 | 1 | 95 | 0⋅100+0⋅50+1⋅10+95⋅1=105 |
| **Industry 4** | 0 | 0 | 1 | 58 | 0⋅100+0⋅50+1⋅10+58⋅1=68 |

Then the Herfindahl index, which is the part of Ellison-Glaeser index, would be as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **100** | **50** | **10** | **1** | **Total H** |
| **industry1** | (100/119)2⋅0=  =0.00 | (50/119)2⋅0=  =0.00 | (10/119)2⋅5=  =0.035 | (1/119)2⋅69=  =0.005 | **0.00+0.00+0.035+0.005=**  **=0.040** |
| **industry2** | (100/121)2⋅0=  =0.00 | (50/121)2⋅0=  =0.00 | (10/121)2⋅4=  =0.027 | (1/121)2⋅81=  =0.006 | **0.00+0.00+0.027+0.006=**  **=0.033** |
| **industry3** | (100/105)2⋅0=  =0.00 | (50/105)2⋅0=  =0.00 | (10/105)2⋅1=  =0.009 | (1/105)2⋅95=  =0.009 | **0.00+0.00+0.009+0.009=**  **=0.018** |
| **industry4** | (100/68)2⋅0=  =0.00 | (50/68)2⋅0=  =0.00 | (10/68)2⋅1=  =0.022 | (1/68)2⋅58=  =0.013 | **0.00+0.00+0.022+0.013=**  **=0.034** |

*Source*: Own calculations

Thus, finally the Ellison-Glaeser index would be:

|  |  |  |  |
| --- | --- | --- | --- |
|  | counter | nominator | Ellison-Glaeser index |
| **Industry 1** | 0.063-(1-0.232)⋅0.040 = 0.032 | (1-0.232)⋅(1-0.040) = 0.737 | 0.032/0.737= 0.044 |
| **Industry 2** | 0.030-(1-0.232)⋅0.033 = 0.005 | (1-0.232)⋅(1-0.033) = 0.743 | 0.005/0.743= 0.006 |
| **Industry 3** | 0.058-(1-0.232)⋅0.018 = 0.045 | (1-0.232)⋅(1-0.018) = 0.754 | 0.045/0.754= 0.059 |
| **Industry 4** | 0.085-(1-0.232)⋅0.034 = 0.059 | (1-0.232)⋅(1-0.034) = 0.742 | 0.059/0.742= 0.079 |
| **Total** | --- | --- | --- |

*Source*: Own calculations

The goal of the above simulation was to present the sensitivity of EG because of changes in firms’ size distribution (and consequently in H index). Scenario A assumes the biggest firms possible, and scenario B rather small ones. There can be additionally scenario C in which all the firms are small. Results are summarized below (Table 2.28):

**Table 2.28: EG index sensitivity to Herfindahl component**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Scenario A**  **Rather big firms** | | **Scenario B**  **Rather small firms** | | **Scenario C**  **Small firms only** | |
|  | **H** | **EG** | **H** | **EG** | **H** | **EG** |
| **Industry 1** | 0.220 | -0.176 | 0.040 | 0.044 | 0.008 | 0.075 |
| **Industry 2** | 0.076 | -0.040 | 0.033 | 0.006 | 0.008 | 0.031 |
| **Industry 3** | 0.083 | -0.080 | 0.018 | 0.059 | 0.010 | 0.067 |
| **Industry 4** | 0.093 | 0.020 | 0.034 | 0.079 | 0.015 | 0.098 |

*Source*: Own calculations

One can see that introducing the bigger firms dramatically changes the EG index (lowers it dramatically), without changing the distribution of location among regions. All three scenarios are technically feasible. With the bigger firms EG is negative and indicates low concentration, with small firms only, EG indicates high concentration. It seems that the conclusions might be opposite, as the many small firms more easily spread over territory than a few big ones. EG indicates that the more competitive market has a higher concentration than the less competitive one.

The literature indicates also the approximation when the size of companies is unknown, and only aggregated data are obtainable. Schmalensee (1977) proposed a method to estimate the Herfindahl component with surrogates, based on the geometric and simulation model.

In the literature there are many studies which use the EG index, mainly determining concentration patterns, but also searching for determinants of agglomeration (e.g. Rosenthal & Strange, 2001). There are also many extensions. Cassey and Smith (2014) propose confidence intervals for the Ellison & Glaeser index, which can play the role of a significance test. Guimaraes, Figueiredo & Woodward (2011) show its insensitivity to spatial permutation of values. They also present a version of the Ellison-Glaeser index, which includes Moran’s I to cover the effects of neighbourhood relations. They examine the Herfindahl index and Ellison-Glaeser (1997) index and propose a correction-component based on Moran’s I. This is to include information obtained from the statistics: the degree of spatial autocorrelation for a given neighbourhood pattern. For the Ellison-Gleaser (EG) index they apply an “*inflation factor*” which depends on Moran’s I applied to differences in shares.

The EG index is also available in software: Dubey (2015) programmed it for STATA and Cassey and Smith (2015) prepared the software to run statistical tests for EG. Ellison, Glaeser and Kerr (2010) proposed a co-agglomeration index, which is closely related to the basic Ellison-Glaeser index (1997), but omits the Herfinahl component. Details of its construction can be found in Ellison, Glaeser and Kerr (2009).

* + 1. **Maurel & Sedillot index of spatial concentration**

Maurel and Sedillot (1999) proposed an improvement to the Ellison-Glaeser index (EG) of spatial concentration, the so called Maurel-Sedillot index (MS). Their main criticism of EG was that it does not include spillovers, which appear because of proximity between firms. Following Alonso-Villar et al. (2004), the MS index can be written as:

and compared with EG:

where *si* is the share of employment in an industry in a region, *xi* is the share of employment in a region, *H* is industrial Herfindahl index. Both indices are calculated for industry, by summing up over regions. Interpretation of the MS index is as follows: is treated as an excess of raw geographic concentration on productive concentration (H), and allows for controlling of the size distribution of plants. It reaches a value of 0 if industry is located randomly across regions, without considering H. Negative values of the index (MS<0) appear when dispersion is a dominating force, and firms do not cluster. Positive values have the same threshold as the EG index: MS<0.02 is for low concentration, MS values between 0.02-0.05 are for intermediate concentration, MS>0.05 are for high concentration of a given sector between regions.

Both indices, MS and EG stem from probabilistic models of location and both measure the geographical concentration of firms/production which is beyond the concentration in selected (biggest) firms (given by 1-H). The main difference is in the counter of both. Even if they include the differences between territorial location of the sector and industrial aggregate, the EG index includes in location by location (, and the MS index takes it aggregated (). Alonso-Villar et al. (2004) present details of a comparison between EG and MS. Some extensions to the MS index were proposed by Maré and Timmins (2006).

Below we operationalize the Maurel-Sedillot index for the data from Table 2.3. Herfindahl indices are as in EG index.

**Table 2.29: Components of Maurel-Sedillot index**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** |
|
| **Industry 1** | (1/119)2=  =0.0001 | (11/119)2=  =0.0085 | (21/119)2=  =0.0311 | (70/119)2=  =0.3460 | (10/119)2=  =0.0071 | (6/119)2=  =0.0025 | **0.0001+0.0085+0.0311+**  **0.3460+0.0071+0.0025=**  **=0.395** |
| **Industry 2** | (1/121)2=  =0.0001 | (40/121)2=  =0.1093 | (24/121)2=  =0.0393 | (40/121)2=  =0.1093 | (5/121)2=  =0.0017 | (11/121)2=  =0.0083 | **0.0001+0.1093+0.0393+**  **0.1093+0.0017+0.0083=**  **=0.268** |
| **Industry 3** | (5/105)2=  =0.0023 | (13/105)2=  =0.0153 | (21/105)2=  =0.0400 | (30/105)2=  =0.0816 | (35/105)2=  =0.1111 | (1/105)2=  =0.0001 | **0.0023+0.0153+0.0400+**  **0.0816+0.1111+0.0001=**  **=0.250** |
| **Industry 4** | (9/68)2=  =0.0175 | (14/68)2=  =0.0424 | (14/68)2=  =0.0424 | (11/68)2=  =0.0262 | (3/68)2=  =0.0019 | (17/68)2=  =0.0625 | **0.0175+0.0424+0.0424+**  **0.0262+0.0019+0.0625**  **=0.193** |
|  |  |  |  |  |  |  | **Total** |
| **Total** | (16/413)2=  =0.0015 | (78/413)2=  =0.0357 | (80/413)2=  =0.0375 | (151/413)2=  =0.1337 | (53/413)2=  =0.0165 | (35/413)2=  =0.0072 | **0.0015+0.0357+0.0375+**  **0.1337+0.0165+0.0072=**  **=0.232** |

*Source*: Own calculations

For scenario A (as in EG index) – big firms

|  |  |  |  |
| --- | --- | --- | --- |
| **MSi** | **counter** | **nominator** | **MS** |
|
| **Industry 1** | ((0.395-0.232)/(1-0.232)) – 0.220 = -0.0068 | 1-0.220 = 0.7805 | -0.0068 / 0.7805 = -0.0087 |
| **Industry 2** | ((0.268-0.232)/(1-0.232)) – 0.076 = -0.0291 | 1-0.076 = 0.9241 | -0.0291 / 0.9241 = -0.0315 |
| **Industry 3** | ((0.250-0.232)/(1-0.232)) – 0.083 = -0.0590 | 1-0.083 = 0.9170 | -0.0590 / 0.9170 = -0.0644 |
| **Industry 4** | ((0.193-0.232)/(1-0.232)) – 0.093 = -0.1435 | 1-0.093 = 0.9074 | -0.1435 / 0.9074 = -0.1581 |

*Source*: Own calculations

For scenario B (as in EG index) – small firms

|  |  |  |  |
| --- | --- | --- | --- |
|  | **counter** | **nominator** | **MS** |
|
| **Industry 1** | ((0.395-0.232)/(1-0.232)) – 0.040 = 0.1725 | 1-0.040 = 0.9598 | 0.1725 / 0.9598 = 0.1798 |
| **Industry 2** | ((0.268-0.232)/(1-0.232)) - 0.033 = 0.0139 | 1-0.033 = 0.9671 | 0.0139 / 0.9671 = 0.0144 |
| **Industry 3** | ((0.250-0.232)/(1-0.232)) - 0.018 = 0.0063 | 1-0.018 = 0.9823 | 0.0063 / 0.9823 = 0.0064 |
| **Industry 4** | ((0.193-0.232)/(1-0.232)) – 0.034 = -0.0851 | 1-0.034 = 0.9658 | -0.0851 / 0.9658 = -0.0881 |

*Source*: Own calculations

The MS index calculated for scenarios with smaller and bigger firms show very divergent results, from concentration to dispersion. When compared with the EG index, results appear to be opposite, yielding negative correlations between these indices even with the same H.

* 1. **Cluster-based measures depending on *n x m* matrix and distance between regions**
     1. **Clustering index**

Franceschi, Mussoni & Pelloni (2009) define clustering index, following Bergstrand (1985). This measure relates the shares of employment in sector and region, and weights this with the distance between regions, in gravity model style. The clustering index is calculated for each sector *m*. It is expressed as follows:

where *m* is the sectors, *i* and *j* represent pairs of regions, *yin, yjn* are the sectoral regional shares of activity measured in the region’s total activity, *yi, yj* are the regional shares of activity measured in national total activity and *dij* is the distance between the regions. Values Cn=1 are in the case of similar distribution of activity in sector and in whole economy, weighted with the distance. High values of Cn suggest that neighbouring regions have a similar share of a given activity. As dividing by 0 is unavailable, one should correct the distances by epsilon, by adding a small value to all pair distances. This impacts strongly on the individual components, but totals stay relatively stable.

It can be operationalised as follows:

**Table 2.30: Matrix of distances between regions – corrected by epsilon**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| **region 1** | 0,001 | 0,782 | 0,404 | 0,431 | 0,541 | 0,550 |
| **region 2** | 0,782 | 0,001 | 0,854 | 0,538 | 0,290 | 0,528 |
| **region 3** | 0,404 | 0,854 | 0,001 | 0,319 | 0,728 | 0,859 |
| **region 4** | 0,431 | 0,538 | 0,319 | 0,001 | 0,460 | 0,658 |
| **region 5** | 0,541 | 0,290 | 0,728 | 0,460 | 0,001 | 0,267 |
| **region 6** | 0,550 | 0,528 | 0,859 | 0,658 | 0,267 | 0,001 |

*Source*: Own calculations

**Table 2.31: Matrix of components of clustering index - counter**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | sector 1 share | 1/16=0.06 | 11/78=0.14 | 21/80=0.26 | 70/151=0.46 | 10/53=0.19 | 6/35=0.17 |
| sector 1 share |  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| 1/16  =0.06 | **region 1** | (0.06\*0.06/  0.001)=3,91 | (0.06\*0.14/0.78)  =0.01 | (0.06\*0.26/0.40)  =0.04 | (0.06\*0.46/0.43)  =0.07 | (0.06\*0.19/0.54)  =0.02 | (0.06\*0.17/0.55)  =0.02 |
| 11/78  =0,14 | **region 2** | (0.06\*0.14/0.78)  =0.01 | (0.14\*0.14/  0.001) =19.89 | (0.14\*0.26/0.85)  =0.04 | (0.14\*0.46/0.54)  =0.12 | (0.14\*0.19/0.29)  =0.09 | (0.14\*0.17/0.53)  =0.05 |
| 21/80  =0,26 | **region 3** | (0.06\*0.26/0.40)  =0.04 | (0.26\*0.14/0.85)  =0.04 | (0.26\*0.26/  0.001)=68.91 | (0.26\*0.46/0.32)  =0.38 | (0.26\*0.19/0.73)  =0.07 | (0.26\*0.17/0.86)  =0.05 |
| 70/151  =0,46 | **region 4** | (0.06\*0.46/0.43)  =0.07 | (0.46\*0.14/0.54)  =0.12 | (0.46\*0.26/0.32)  =0.38 | (0.46\*0.46/  0.001)=214.90 | (0.46\*0.19/0.46)  =0.19 | (0.46\*0.17/0.66)  =0.12 |
| 10/53  =0,19 | **region 5** | (0.06\*0.19/0.54)  =0.02 | (0.19\*0.14/0.29)  =0.09 | (0.19\*0.26/0.73)  =0.07 | (0.19\*0.46/0.46)  =0.19 | (0.19\*0.19/  0.001)=35.60 | (0.19\*0.17/0.27)  =0.12 |
| 6/35  =0,17 | **region 6** | (0.06\*0.17/0.55)  =0.02 | (0.17\*0.14/0.53)  =0.05 | (0.17\*0.26/0.86)  =0.05 | (0.17\*0.46/0.66)  =0.12 | (0.17\*0.19/0.27)  =0.12 | (0.17\*0.17/  0.001)=29.39 |
| Total=375,38 | | | | | | | |

*Source*: Own calculations

**Table 2.32: Matrix of components of clustering index - nominator**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| total activity share | | 16/413=0.04 | 78/413=0.19 | 80/413=0.19 | 151/413=0.37 | 53/413=0.13 | 35/413=0.08 |
| **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| 16/413=0.04 | **region 1** | (0.04\*0.04/  0.001)=1.50 | (0.04\*0.19/  0.78)=0.01 | (0.04\*0.19/  0.40)=0.02 | (0.04\*0.37/  0.43)=0.03 | (0.04\*0.13/  0.54)=0.01 | (0.04\*0.08/  0.55)=0.01 |
| 78/413  =0.19 | **region 2** | (0.04\*0.19/  0.78)=0.01 | (0.19\*0.19/  0.001)=35.67 | (0.19\*0.19/  0.85)=0.04 | (0.19\*0.37/  0.54)=0.13 | (0.19\*0.13/  0.29)=0.08 | (0.19\*0.08/  0.53)=0.03 |
| 80/413  =0,19 | **region 3** | (0.04\*0.19/  0.40)=0.02 | (0.19\*0.19/  0.85)=0.04 | (0.19\*0.19/  0.001)=37.52 | (0.19\*0.37/  0.32)=0.22 | (0.19\*0.13/  0.73)=0.03 | (0.19\*0.08/  0.86)=0.02 |
| 151/413=0.37 | **region 4** | (0.04\*0.37/  0.43)=0.03 | (0.37\*0.19/  0.54)=0.13 | (0.37\*0.19/  0.32)=0.22 | (0.37\*0.37/  0.001)=133.68 | (0.37\*0.13/  0.46)=0.10 | (0.37\*0.08/  0.66)=0.05 |
| 53/413  =0,13 | **region 5** | (0.04\*0.13/  0.54)=0.01 | (0.13\*0.19/  0.29)=0.08 | (0.13\*0.19/  0.73)=0.03 | (0.13\*0.37/  0.46)=0.10 | (0.13\*0.13/  0.001)=16.47 | (0.13\*0.08/  0.27)=0.04 |
| 35/413  =0,08 | **region 6** | (0.04\*0.08/  0.55)=0.01 | (0.08\*0.19/  0.53)=0.03 | (0.08\*0.19/  0.86)=0.02 | (0.08\*0.37/  0.66)=0.05 | (0.08\*0.13/  0.27)=0.04 | (0.08\*0.08/  0.001)=7.18 |
| Total=233,67 | | | | | | | |

*Source*: Own calculations

For sector 1 the C1 = 375.38/233.67 = 1.61 which proves some clustering patterns of similar values in the neighbourhood. The index without weighting with distance would be 1,66, which proves that neighbourhood effects lowers the clustering pattern.

In the same paper, Franceschi, Mussoni & Pelloni (2009) define an index to measure agglomeration, concentration and specialization together, including information on volume, density and region dimensionality. Their Regional Industrial Mass and Regional Industrial Concentration indices stem from physics, that mass equals density times volume. On this basis they redefine it for regions.

* 1. **Cluster-based measures depending on *n x m* matrix and spatial relations**

In regional studies, the relative relation of regions matters as well as the spatial pattern of activity measured. The basic tool of spatial analysis is the spatial weights matrix W, which defines who is whose neighbor and for how much. There are few criteria for building the spatial relationship matrix, but the most common is the contiguity matrix. It assumes that regions which share a common border are neighbours, and the strength of this link depends on the number of neighbours (if region has 5 neighbours, each gets the weight wij=1/5=0.2).

In the case of the example used in this study, spatial neighborhood matrix and spatial weights matrix are as follows:

**Table 2.33: Spatial neighbourhood matrix and spatial weights matrix for map on Fig.4**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **neighbourhod** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** |
| **region 1** | 0 | 0 | 1 | 1 | 1 | 1 | **4** |
| **region 2** | 0 | 0 | 0 | 1 | 1 | 0 | **2** |
| **region 3** | 1 | 0 | 0 | 1 | 0 | 0 | **2** |
| **region 4** | 1 | 1 | 1 | 0 | 1 | 0 | **4** |
| **region 5** | 1 | 1 | 0 | 1 | 0 | 1 | **4** |
| **region 6** | 1 | 0 | 0 | 0 | 1 | 0 | **2** |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **weights wij** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total** |
| **region 1** | 0 | 0 | 0.25 | 0.25 | 0.25 | 0.25 | **1** |
| **region 2** | 0 | 0 | 0 | 0.5 | 0.5 | 0 | **1** |
| **region 3** | 0.5 | 0 | 0 | 0.5 | 0 | 0 | **1** |
| **region 4** | 0.25 | 0.25 | 0.25 | 0 | 0.25 | 0 | **1** |
| **region 5** | 0.25 | 0.25 | 0 | 0.25 | 0 | 0.25 | **1** |
| **region 6** | 0.5 | 0 | 0 | 0 | 0.5 | 0 | **1** |

*Source*: Own calculations

The spatial weights matrix is applied then to statistics and econometric models. There are a few basic statistics, which are commonly used to assess the spatial dependence. If similar regions (in terms of phenomena observed) are neighbors more often than would happen randomly, than it is called positive spatial autocorrelation. This clustering pattern appears often in socio-economic studies on a regional level. The most important measure is Moran’s I – the coefficient of spatial autocorrelation. For concentration studies purpose, it is expressed for sector *n* as follows:

where m are the regions, wij are the spatial weights from row-standardised matrix W, xi is the value in the studied region, is the average of values in the given sector. Values of Moran’s I which are greater than the expected value (*E(I)=-1/(m-1)*) indicate positive autocorrelation.

* + 1. **Gini with Moran’s I and Getis-Ord’s G**

Arbia (2001a) introduced spatial measures to the measurement of regional concentration. This method is based on well-known tools of spatial statistics, and by comparing traditional a-spatial with spatial measures, can give more insight to processes observed. The procedure is to compare the Gini index, Moran’s I and Getis-Ord G. The main issue is to see manufacturing (detailed) and industrial (general) employment on the commune (lower, NTS5[[10]](#footnote-10)) and province (higher, NTS4) level of aggregation.

In counting the Gini-type index it is to calculate firstly two indices, for manufacturing employment and total industrial employment: the communal share of provincial employment as the ratio of employment in each commune to employment in each province. Those pairs for communes are ordered ascendingly within provinces, and as cumulated values mapped to build a Lorenz curve (*x=cumulative share of manufacturing employment*, *y= cumulative share of industrial employment*). The area between the empirical line and 45 degrees line is multiplied by 2 and standardised in interval [0,1]. Moran’s I and Getis-Ord’s G are calculated on the basis of the ratio relating manufacturing employment in communes and province and standardized to keep comparability between provinces. Arbia (2001a) proves that those three measures capture different aspects of the same phenomenon and should be considered jointly.

* + 1. **Gini with ESDA (local and global Moran’s I)**

Guillain & Le Gallo (2010) use the locational Gini and global and local Moran’s I for LQ to conclude about the agglomeration patterns on a local level. In fact, as Arbia (2001), they apply the standard tool of spatial statistics, which involves the spatial weights matrix W defining the neighbourhood. This joint interpretation of Gini and Moran statistics is to provide different but complementary information on the spatial agglomeration. The main rationale is that the concentration measure does not refer to the spatial pattern of location and also does not assess the significance of clusters. It is worth it to underline that Guillain & Le Gallo (2010) apply territorially aggregated data, not the point data.

The approach by Guillain & Le Gallo (2010) is built on global and local indicators. The global indicator, single for each sector, allows for a global perspective on concentration and agglomeration. The local indicator, for each sector and region, assesses the position of each spatial unit, and therefore is more precise in the local context.

In the global approach, locational Gini and Moran’s I for LQ measure to some degree similar and different phenomena. Concentration given by Gini for each sector checks the overrepresentation of some activity in some regions (or its diversification) without looking at its location. Spatial clustering given by Moran’s I checks the spatial pattern of location and if similar values are located closer to other similar. In this approach those measures are calculated for each sector.

The joint interpretation of these measures is as follows:

* high Gini and low Moran’s I – apparent agglomeration does not sprawl over the territory and is just located in a single region
* high Gini and high Moran’s I – sectoral concentration appears and is present in neighboring 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

Guillain & Le Gallo (2010) test the results with a local approach. They apply two local measures. They map the Moran scatterplot for LQ and count the percentage distribution of sectors in quarters (HH, HL, LH, LL). They also count LISA (*Local Indicator of Spatial Association*) for LQ and run the significance test for LISA, to summarise the distribution of significant LISA in the same quarters. They compare the distributions of Moran Scatterplots and significant LISA in quarters and conclude on agglomeration patterns.

The table 2.34 below operationalises the global approach.

**Table 2.34: LQ for sectors**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **LQ** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Average LQ**  **in sectors** |
| **Industry 1** | 0.22 | 0.49 | 0.91 | 1.61 | 0.65 | 0.59 | **0.75** |
| **Industry 2** | 0.21 | 1.75 | 1.02 | 0.90 | 0.32 | 1.07 | **0.88** |
| **Industry 3** | 1.23 | 0.66 | 1.03 | 0.78 | 2.60 | 0.11 | **1.07** |
| **Industry 4** | 3.42 | 1.09 | 1.06 | 0.44 | 0.34 | 2.95 | **1.55** |

*Source*: Own calculations

**Table 2.35: Components of Moran’s I for sectors**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| LQ in sector 1🡪 | | 0.22 | 0.49 | 0.91 | 1.61 | 0.65 | 0.59 |
| LQ in sector 1↓ | **Average LQ in sector 1**  **=0.75** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| 0.22 | **region 1** | 0⋅(0.22-0.75)⋅  (0.22-0.75)=  =0.00 | 0⋅(0.49-0.75)⋅  (0.22-0.75)=  =0.00 | 0.25⋅  (0.91-0.75)⋅  (0.22-0.75)=  =-0.02 | 0.25⋅  (1.61-0.75)⋅  (0.22-0.75)=  =-0.11 | 0.25⋅  (0.65-0.75)⋅  (0.22-0.75)=  =0.01 | 0.25⋅  (0.59-0.75)⋅  (0.22-0.75)=  =0.02 |
| 0.49 | **region 2** | 0⋅(0.22-0.75)⋅  (0.49-0.75)=  =0.00 | 0⋅(0.49-0.75)⋅  ( 0.49-0.75)=  =0.00 | 0⋅(0.91-0.75)⋅  (0.49-0.75)=  =0.00 | 0.5⋅(1.61-0.75)⋅  (0.49-0.75)=  =-0.11 | 0.5⋅(0.65-0.75)⋅  (0.49-0.75)=  =0.01 | 0⋅(0.59-0.75)⋅  (0.49-0.75)=  =0.00 |
| 0.91 | **region 3** | 0.5⋅(0.22-0.75)⋅  (0.91-0.75)=  =-0.04 | 0⋅(0.49-0.75)⋅  ( 0.91-0.75)=  =0.00 | 0⋅(0.91-0.75)⋅  (0.91-0.75)=  =0.00 | 0.5⋅(1.61-0.75)⋅  (0.91-0.75)=  =0.07 | 0⋅(0.65-0.75)⋅  (0.91-0.75)=  =0.00 | 0⋅(0.59-0.75)⋅  (0.91-0.75)=  =0.00 |
| 1.61 | **region 4** | 0.25⋅  (0.22-0.75)⋅  (1.61-0.75)=  =-0.11 | 0.25⋅(0.49-0.75)⋅  ( 1.61-0.75)=  =-0.06 | 0.25⋅  (0.91-0.75)⋅  (1.61-0.75)=  =0.04 | 0⋅(1.61-0.75)⋅  (1.61-0.75)=  =0.00 | 0.25⋅(0.65-0.75)⋅  (1.61-0.75)=  =-0.02 | 0⋅(0.59-0.75)⋅  (1.61-0.75)=  =0.00 |
| 0.65 | **region 5** | 0.25⋅  (0.22-0.75)⋅  (0.65-0.75)=  =0.01 | 0.25⋅(0.49-0.75)⋅  ( 0.65-0.75)=  =0.01 | 0⋅(0.91-0.75)⋅  (0.65-0.75)=  =0.00 | 0.25⋅  (1.61-0.75)⋅  (0.65-0.75)=  =-0.02 | 0⋅(0.65-0.75)⋅  (0.65-0.75)=  =0.00 | 0.25⋅  (0.59-0.75)⋅  (0.65-0.75)=  =0.00 |
| 0.59 | **region 6** | 05⋅(0.22-0.75)⋅  (0.59-0.75)=  =0.04 | 0⋅(0.49-0.75)⋅  ( 0.59-0.75)=  =0.00 | 0⋅(0.91-0.75)⋅  (0.59-0.75)=  =0.00 | 0⋅(1.61-0.75)⋅  (0.59-0.75)=  =0.00 | 0.5⋅(0.65-0.75)⋅  (0.59-0.75)=  =0.01 | 0⋅(0.59-0.75)⋅  (0.59-0.75)=  =0.00 |
| **Total** | | -0.28 | | | | | |
|  |  |  |  |  |  |  |  |
| **squared difference of LQi and LQav** | | (0.22-0.75)2=  0.28 | (0,49-0,75)2=  0.07 | (0,91-0,75)2=  0.03 | (1,61-0,75)2=  0.74 | (0,65-0,75)2=  0.01 | (0,59-0,75)2=  0.02 |
| **Total** | | 0.28+0.07+0.03+0.74+0.01+0.02=1.15 | | | | | |
| **Moran’s I** | | -0.28/1.15 = **-0.24** | | | | | |

*Source*: Own calculations

**Table 2.36: Locational Gini and Moran’s I for sectors**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Locational Gini** | **Locational Gini ranking**  **(1=the highest value)** | **Moran's I for LQ** | **Moran's I for LQ ranking**  **(1=the highest value)** |
| **Industry 1** | 0.09 | 4 | -0.24 | 2 |
| **Industry 2** | 0.19 | 3 | -0.27 | 4 |
| **Industry 3** | 0.32 | 2 | -0.48 | 3 |
| **Industry 4** | 0.38 | 1 | -0.03 | 1 |

*Source*: Own calculations

This analysis shows a few facts. First of all, there is no spatial pattern of agglomeration, as for all sectors the Moran’s I are negative. This means that neighborhood links are rather with the different regions. This is not very typical for most of economies, but this example follows this pattern. It might also be biased by edge effect – a map with six regions only, where all regions are borders “to nothing” may also cause this effect of negative relationship. Secondly, the first position in both rankings goes to industry 4, which will mean that in both dimensions, sectoral and geographical, it reveals the strongest of all agglomeration. The other sectors are less agglomerated.

In a somehow similar way Sohn (2014) analyses the EG index in a spatial context, by applying both Moran’s I, as well as factor analysis and looking for between-industry and within-industry pattern.

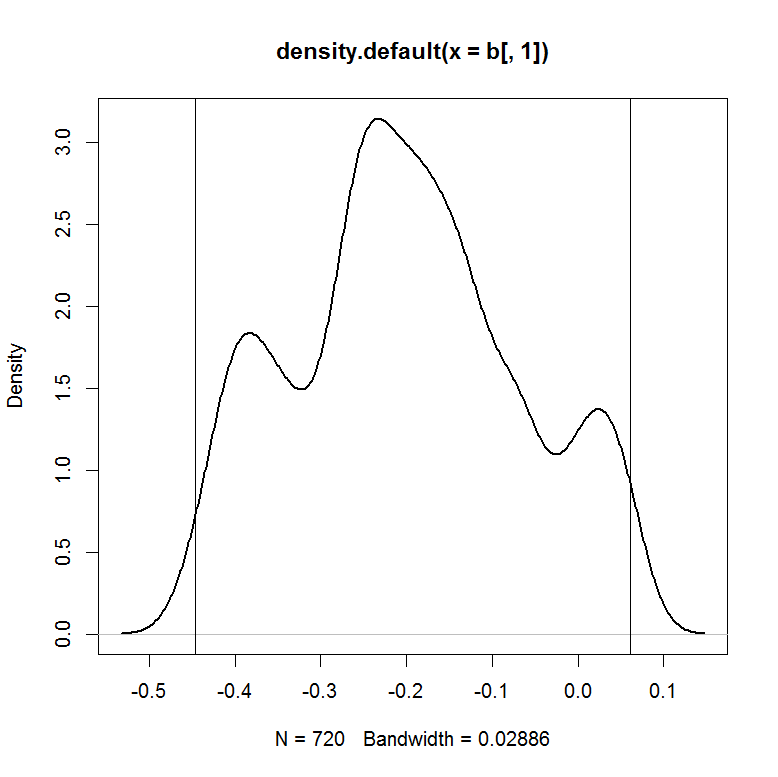
* + 1. **Spatial Concentration Measure**

Arbia & Piras (2009) construct a measure of spatial concentration, which by assumption will be sensitive to spatial permutations. This feature is indeed important, as the majority of cluster-based measures is a-spatial and spatial reconfiguration of individual regions changes the phenomenon, without changing the measure. Arbia & Piras (2009) propose the sectoral formula as follows:

where *Xj*­ is the phenomenon observed in a given location, *Xj\** is the value of the phenomenon which maximises the autocorrelation obtained from permutation of X values, μ is the average value of the phenomenon, m is the number of regions, n is the sector. Results of X permutation are tested for autocorrelation with Moran’s I or Getis-Ord G. Permutation of X which maximizes these statistics is treated as X\*. For easier interpretation of coefficients, one can set B=λ2, which is limited from 0 to 1. With λ=0 (as well with B=0), the existing spatial pattern is the same as the one maximizing the spatial concentration, so it is extreme concentration. On the contrary, λ=-1 or 1 (as well with B=1) is for total diversification of activity over space, without any concentration. This measure refers to both spatial concentration in a counter, as well as to inter-regional a-spatial concentration in the nominator.

To operationalize this method, one should prepare all possible permutations of the vector of activity analysed. For those values one computes Moran’s I and look for the permutation maximizing the spatial autocorrelation. For the example data with 6 regions, there were 720 permutations, which gave the distribution of Moran’s I as on Figure 2.2 below. The permutations of values with the highest Moran’s I=0,06\*\* are as follows (for regions 1 to 6): X\*1=(1, 70, 6, 11, 21, 10), X\*2=(1, 70, 10, 21, 11, 6), X\*3=(11, 10, 6, 1, 21, 70), X\*4=(21, 6, 70, 11, 1, 10), comparing to original allocation X=(1, 11, 21, 70, 10, 6). The average value of X is always 19.8.

**Figure 2.2: Density of Moran’s I for all permutations (with horizontal lines for max and min)**



*Source*: Own calculations

**Table 2.37: Components of**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total for regions** |  |
| **nominator components** | (1-19.8)2=  354.7 | (11-19.8)2=  78.0 | (21-19.8)2=  1.4 | (70-19.8)2=  2516.7 | (10-19.8)2=  96.7 | (6-19.8)2=  191.4 | 3238.8 |  |
|  |  |  |  |  |  |  |  |  |
| **counter components** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** | **Total for regions** | **lambda** |
| **permutation 1** | (1-19.8)⋅  (1-19.8)  =354.7 | (70-19.8)⋅  (11-19.8)  =-443.1 | (6-19.8)⋅  (21-19.8)  =-16.1 | (11-19.8)⋅  (70-19.8)  =-443.1 | (21-19.8)⋅  (10-19.8)  =-11.5 | (10-19.8)⋅  (6-19.8)  =136.0 | -423.2 | -423.2/  3238.8=  -0.13 |
| **permutation 2** | (1-19.8)⋅  (1-19.8)  =354.7 | (70-19.8)⋅  (11-19.8)  =-443.1 | (10-19.8)⋅  (21-19.8)  =-11.5 | (21-19.8)⋅  (70-19.8)  =58.5 | (11-19.8)⋅  (10-19.8)  =86.9 | (6-19.8)⋅  (6-19.8)  =191.4 | 236.8 | 236.8/  3238.8=  0.07 |
| **permutation 3** | (11-19.8)⋅  (1-19.8)  =166.4 | (10-19.8)⋅  (11-19.8)  =86.9 | (6-19.8)⋅  (21-19.8)  =-16.1 | (1-19.8)⋅  (70-19.8)  =-944.8 | (21-19.8)⋅  (10-19.8)  =-11.5 | (70-19.8)⋅  (6-19.8)  =-694.0 | -1413.2 | -1413.2/  3238.8=  -0.44 |
| **permutation 4** | (21-19.8)⋅  (1-19.8)  =-22.0 | (6-19.8)⋅  (11-19.8)  =122.2 | (70-19.8)⋅  (21-19.8)  =58.5 | (11-19.8)⋅  (70-19.8)  =-443.1 | (1-19.8)⋅  (10-19.8)  =185.2 | (10-19.8)⋅  (6-19.8)  =136.0 | 36.8 | 36.8  /3238.8=  0.01 |

*Source*: Own calculations

Thus the would be for those four combinations as follows: , , , . This proves that this method is in fact sensitive to spatial distributions, but with a small number of regions it cannot be conclusive, as for permutation 3 the empirical distribution is far from concentrated, oppositely to permutation 4, which proves that the empirical distribution is almost perfectly concentrated.

* + 1. **Relative Industrial Relevance**

Carlei and Nuccio (2014) develop an Relative Industrial Relevance index, which is based on Self-Organising Maps. This method can identify different spatial patterns of industrial agglomeration and co-agglomeration. It is based on neural network architectures, which follows the Kohonen (2001). This index is calculated for every single industry. Computational complexity makes it impossible to present the calculations below.

* 1. **Comparison of cluster-based measures**

This chapter presents the overview of the most of existing indicators of sectoral and geographical concentration that are applied in regional studies on regions and their sectoral composition. A review of these indicators brings some insight into their construction and the results they yield.

Firstly, one could notice that indicators described in this chapter start usually with the same input information – a two dimensional table of employment by regions and sectors, plus possibly other supplementing information on the size of companies, dynamics and next period values, the relative locations of regions and distances to neighbours, as well as size of the region’s territory. One can expect that with the same information delivered, the same result is being obtained. This will be tested below.

Secondly, one could observe that there is in fact much confusing in terms of terminology. Indicators described above are to measure concentration, specialization and agglomeration. Names differ even if the construction of the indicator is similar. Following the study from Chapter 1, indicators should be named consequently to the results they offer. There are in fact four kinds of indicators:

* For regions – when they yield a single value for each region (in the case of the sample data in this chapter they give six values)
* For industries - when they yield single value for each sector (in the case of the sample data in this chapter they give four values)
* For the whole economy – when they yield a single value for all regions and sectors (in the case of the sample data in this chapter they give one value)
* For local markets – when they yield a single value for every cell in the analysed regional-industrial matrix (in the case of the sample data in this chapter they give 24 values)

The dimension of the result the indicators yield, defines its type:

- Indicators for regions, which analyse the internal structure of the regional economy, with reference or no to benchmarks and other regions, are the indicators of sectoral concentration. Usually they are called “specialisation” indicators, but as proved in Chapter 1 it seems to be exaggerated, as it shows simply the over- or under-representation of employment in a region and consequently should be referred to as sectoral concentration measures. This group can also include dynamic indicators, which assess the changes in internal structure of regional economies.

- Indicators for industries, which analyse the distribution of firms of a given sector among the regions, with reference or no to benchmarks and other sectors or the whole economy, are the indicators of geographical concentration. This group can also include dynamic indicators to measure the changes in inter-regional business allocation for a given industry.

- Indicators for the whole economy, which analyse simultaneously the distribution of firms between regions and sectors.

- Indicators for a single “cell” sector in a region.

Thirdly, indicators are defined with the reference to different underlying benchmarks. There are mainly two possibilities: random distribution and empirical distribution. Radom or uniform distribution, which in fact means equal shares of firms among regions or sectors, is being assumed in case of entropy measures (Relative H, Theil’s H, Shannon’s H) of both sectoral and geographical concentration as well as in Ogive index and refined diversification index for sectoral concentration and in Kullback-Leibler Divergence (KLD) in case of geographical concentration; and in Theil total which is both for sectors and regions. Empirical underlying distribution appears in majority of indicators: for sectoral concentration in National Averages Index (NAI), Krugman dissimilarity index, Relative Diversity Index (RDI) (inverse Krugman), Hachman Index, Hallet Index, Kullback-Leibler Divergence (KLD) and Lilien index (dynamic index); for geographical concentration in Krugman concentration index, Bruhlart & Traeger index, Agglomeration V and Clustering index (Bergstrand index), and for Location Quotient (LQ). There are also measures with strongly transformed empirical underlying benchmark distribution, as Gini, locational Gini, Relative Specialisation Index, Ellison-Glaeser index (EG) and Maurell-Sedillot index (MS). There are also some measures with no simple distribution playing the role of benchmark as in case of Herfindahl (HH) index, Absolute Diversity Index (ADI) (inverse HH) or Guillain & LeGallo (Moran I for LQ).

Classification of the indicators into those four groups is as follows (see Table 2.38):

**Table 2.38: Classification of indicators**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Benchmark**  **distribution** | **sectoral concentration measures for regions** | **geographical concentration measures for industries** | **overall concentration measures**  **for whole economy** | **measures for single “cell”**  **/for sector in region/** |
| Uniform distribution | Relative H  Theil’s H  Shannon’s H  Ogive Index  Refined Diversification Index | Relative H  Theil’s H  Shannon’s H  Kullback-Leibler Divergence (KLD) | Theil total |  |
| Empirical  distribution | National Averages Index (NAI)  Relative Diversity Index (RDI) (inverse Krugman Index)  Hachman Index  Hallet Index  Kullback-Leibler Divergence (KLD)  Krugman Dissimilarity Index  Lilien Index (dynamic index) | Krugman concentration index  Bruhlart & Traeger index  Agglomeration V  Clustering index (Bergstrand index) | Geographic concentration index | Location Quotient (LQ) |
| Transformed empirical distribution | Gini  Relative Specialisation Index (RSI=max LQ) | Gini  Locational Gini  Ellison-Glaeser Index (EG)  Maurell-Sedillot Index (MS) |  |  |
| No distribution | Herfindahl (HH)  Absolute Diversity Index (ADI) (inverse HH) | Guillain & LeGallo (Moran I for LQ) |  |  |

*Source*: Own classification

What is more, almost none of them is an agglomeration coefficient as they operate on clustered data and involvement of spatial aspects, which can indicate the density of business of a given area is very poor, as it is limited to spatial links between regions. Clustered data, by nature, does not include information about the spatial distribution inside the regions, which is the clue for the agglomeration measurement.

Indicators can be tested about their behavior and information capacity within the functional groups delimited above. The core issue is to answer the question if they give the same or different information on regional structure of employment = sectoral concentration or allocation of business units among regions = geographical concentration.

Below we present the comparative analytics of measures introduced in this chapter. They were calculated for the same dataset as in the examples. Codes to R CRAN software for these indicators can be found in Appendix 3. This summary includes only the overview of indicators used, with measures listed and their results tabulated. There is no analysis of inter-relations between the measures, as the artificial dataset and a small sample are not in favour of this kind of analysis. The study on correlation and joint relations between measures was presented in Chapter 4, which includes wider analysis on a real and bigger dataset, which makes it more reliable and robust.

***Sectoral concentration measures for regions***

This chapter summarised several indicators for regions. Indicators differ in variation between extreme values; some of them are much more sensitive than others, which can be measured with the coefficient of variation (*cv*). The highest sensitivity is in the case of RDI (*cv*=154%) and NAI (*cv*=88%), the lowest in the case of Herfindahl and ADI (*cv*=21%) and entropy measures, Shannon and relative H (*cv*=13%).

**Table 2.39: Summary of sectoral concentration indicators**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |  | ***mean*** | ***std.dev*** | ***Coeff. of variation*** |
| **NAI** | 1,33 | 0,27 | 0,00 | 0,17 | 0,89 | 0,88 |  | *0,59* | *0,52* | *0,88* |
| **KLD** | 0,56 | 0,13 | 0,00 | 0,09 | 0,38 | 0,40 |  | *0,26* | *0,22* | *0,84* |
| **Lilien** | 0,14 | 0,08 | 0,05 | 0,03 | 0,23 | 0,10 |  | *0,10* | *0,07* | *0,72* |
| **Ogive** | 0,69 | 0,37 | 0,03 | 0,32 | 0,94 | 0,46 |  | *0,47* | *0,31* | *0,67* |
| **Theil’s H** | 0,35 | 0,16 | 0,02 | 0,17 | 0,41 | 0,27 |  | *0,23* | *0,14* | *0,62* |
| **refined diversification index** | 0,52 | 0,27 | -0,02 | 0,32 | 0,58 | 0,42 |  | *0,35* | *0,21* | *0,62* |
| **Krugman index** | 0,91 | 0,47 | 0,05 | 0,35 | 0,81 | 0,69 |  | *0,55* | *0,32* | *0,59* |
| **Hallet** | 0,46 | 0,24 | 0,03 | 0,18 | 0,41 | 0,34 |  | *0,27* | *0,16* | *0,59* |
| **Gini** | 0,95 | 0,50 | 0,07 | 0,42 | 0,87 | 0,85 |  | *0,61* | *0,34* | *0,56* |
| **RSI (max LQ)** | 3,42 | 1,75 | 1,06 | 1,61 | 2,60 | 2,95 |  | *2,23* | *0,90* | *0,40* |
| **Herfindahl** | 0,42 | 0,34 | 0,26 | 0,33 | 0,48 | 0,37 |  | *0,37* | *0,08* | *0,21* |
| **Relative Diversity Index (RDI)**  **(inverse Krugman)** | 1,10 | 2,13 | 19,50 | 2,85 | 1,23 | 1,46 |  | *4,71* | *7,28* | *1,54* |
| **Hachman** | 0,43 | 0,79 | 1,00 | 0,85 | 0,53 | 0,53 |  | *0,69* | *0,22* | *0,32* |
| **Absolute Diversity Index (ADI)**  **(inverse HH)** | 2,37 | 2,92 | 3,87 | 3,03 | 2,07 | 2,74 |  | *2,83* | *0,62* | *0,21* |
| **Shannon’s H** | 1,03 | 1,23 | 1,39 | 1,22 | 0,97 | 1,12 |  | *1,16* | *0,15* | *0,13* |
| **Relative H** | 0,75 | 0,88 | 0,99 | 0,88 | 0,70 | 0,81 |  | *0,84* | *0,10* | *0,12* |

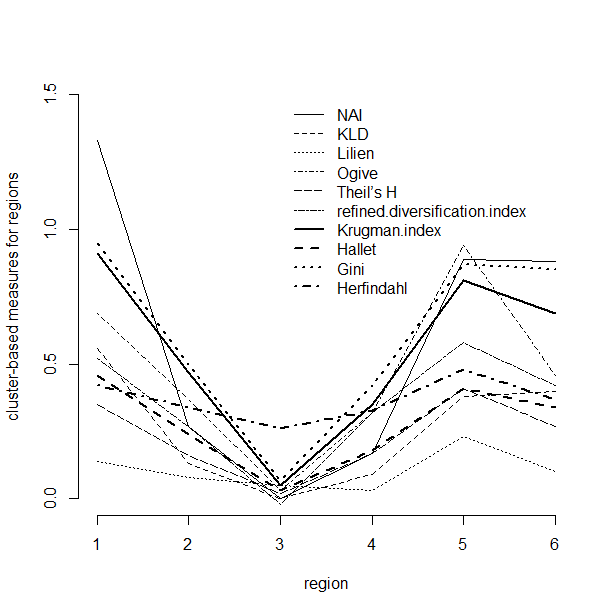
*Source*: Own calculations

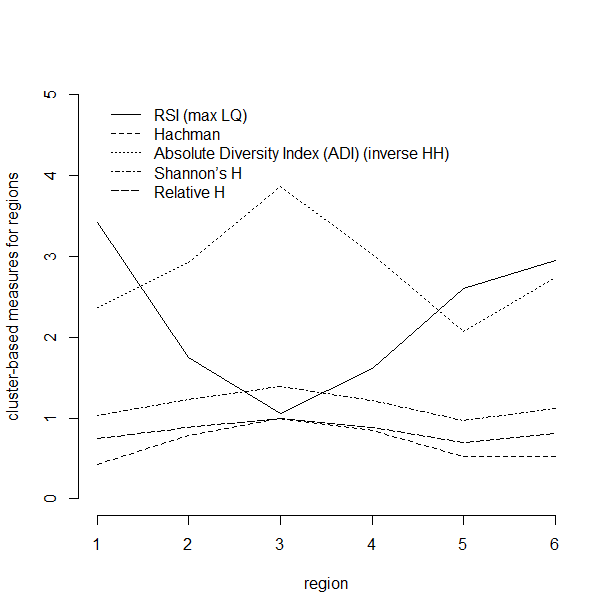
**Table 2.40: Interpretation of indicators of sectoral concentration**

|  |  |
| --- | --- |
| **Indicator** | **Rules of interpretation** |
| **NAI** | From 0 to max  0 for low disparity between regional and national structure  Max for significant disparity between regional and national structure |
| **KLD (Kullback-Leibler divergence)** | 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** | For dynamic data  From 0 to max  0 for structural stability over time  max for significant shits between industries over time |
| **Ogive** | From 0 to max  0 for equal shares of industry  The more unequal the shares, the higher the Ogive measure |
| **Theil’s H** | From 0 to max  0 for equal (uniform) distribution of sectors within the region  Max for fully unequal distribution of sectors within the region |
| **refined diversification index** | From 0 to 1  0 for the full diversification of region (equal shares)  1 for complete non-diversification |
| **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** | 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). |
| **Gini** | 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 |
| **Herfindahl H** | From 0 to 1  0 for even distribution of employment among the firms  1 for extreme concentration of employment in few (or single) firm (monopolisation) |
| **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 |
| **Hachman** | 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 |
| **Absolute Diversity Index (ADI)**  **(inverse HH)** | Inverse Herfindahl index  From 0 to max  0 for even distribution of employment among the firms  max for extreme concentration of employment in few (or single) firm (monopolisation) |
| **Shannon entropy H** | From 0 to ln(n)  0 for full concentration of industry  Max for equal share (full diversification) |
| **relative entropy H** | From 0 to 1  0 for full concentration of industry  1 for equal shares of industries within region |

*Source*: Own summary

**Figure 2.3: Sectoral concentration indicators for example data**





*Source*: Own calculations

***Geographical concentration measures for industries***

This chapter also summarized few indicators for industries. Variation of indicators differs: the lowest sensitivity is in case of agglomeration V index (*cv*=14.3 %) and the highest in case of MS index (*cv*=396 %). Significantly different is the clustering index (Bergstrand), its otherness may result from the data on distances between regions, which was incorporated in the indicator.

**Table 2.41: Summary of geographical concentration indicators**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Industry**  **1** | **Industry**  **2** | **Industry**  **3** | **Industry**  **4** |  | ***mean*** | ***std.dev*** | ***Coeff. of variation*** |
| **Gini** | 0,490 | 0,540 | 0,570 | 0,740 |  | 0,585 | 0,108 | 0,185 |
| **Krugman** | 0,445 | 0,305 | 0,440 | 0,576 |  | 0,442 | 0,111 | 0,251 |
| **Bruhlart &Traeger** | 0,053 | 0,045 | 0,083 | 0,123 |  | 0,076 | 0,035 | 0,465 |
| **locational Gini** | 0,089 | 0,189 | 0,324 | 0,385 |  | 0,247 | 0,133 | 0,540 |
| **Guillain & LeGallo**  **(Moran for LQ)\*** | -0,240 | -0,269 | -0,485 | -0,032 |  | -0,260 | 0,185 | -0,718 |
| **Ellison-Gaeser** | 0,044 | 0,006 | 0,059 | 0,079 |  | 0,047 | 0,031 | 0,657 |
| **agglomeration V** | 0,937 | 0,928 | 1,147 | 1,231 |  | 1,061 | 0,152 | 0,143 |
| **clustering index (Bergstrand)** | 1,61 | 2,30 | 2,88 | 2,67 |  | 2,366 | 0,56 | 0,237 |
| **KLD** | 0,550 | 0,350 | 0,300 | 0,090 |  | 0,323 | 0,189 | 0,586 |
| **MS** | 0,180 | 0,014 | 0,006 | -0,088 |  | 0,028 | 0,111 | 3,958 |
| **Shannon’s H** | 1,24 | 1,44 | 1,49 | 1,70 |  | 1,47 | 0,19 | 0,13 |
| **Relative H** | 0,69 | 0,80 | 0,83 | 0,95 |  | 0,82 | 0,11 | 0,13 |
| **Theil’s H** | 0,55 | 0,35 | 0,30 | 0,09 |  | 0,32 | 0,19 | 0,58 |

*\* here all Moran’s I insignificant*

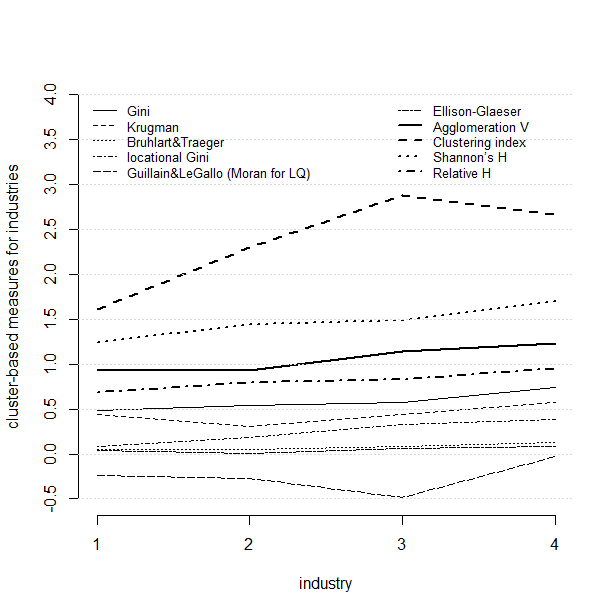
*Source*: Own calculations

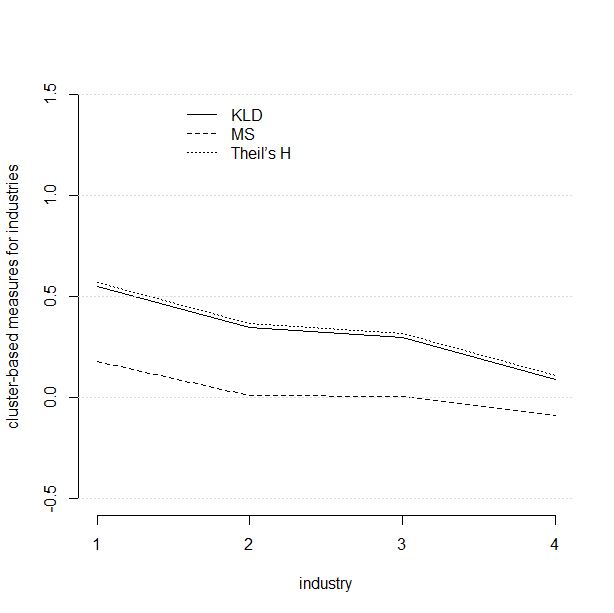
**Table 2.42: Interpretation of indicators of geographical concentration**

|  |  |
| --- | --- |
| **Indicator** | **Rules of interpretation** |
| **Gini** | From 0 to 1  0 for uniform distribution of activity among regions within the sector  1 for full geographical concentration in the sector |
| **Krugman** | From 0 to max  0 for firms from industry allocated proportionally to the region’s size  Max for firms from industry allocated to singe region only |
| **Bruhlart & Traeger GE** | From 0 to max  0 for dispersion of industry (equal shares)  max for geographical concentration of industry |
| **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 |
| **Guillain & LeGallo**  **(Moran for LQ and Gini)** | Joint interpretation of Moran’s I & Gini or LQ index  - high Gini and low Moran’s I – apparent agglomeration does not sprawl over the territory and is just located 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 |
| **Ellison-Glaeser (EG)** | From –max to +max  EG<0 for spatial dispersion  EG=0 for random distribution of firms among regions  EG>0 geographical concentration of business:  EG<0.02 low concentration  EG>0.05 high concentration |
| **Agglomeration index 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 bigger 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  1 for similar distribution of activity in the sector and in the whole economy, weighted with the distance.  C > 1 for neighbouring regions that have similar shares of given activity; the higher the C value, the stronger clustering |
| **KLD (Kullback-Leibler divergence)** | From 0 to max  0 for complete spatial dispersion of business  max for extreme geographical concentration |
| **Maurel & Sedillot (MS)** | From –max to +max  MS<0 for spatial dispersion  MS=0 for random distribution of firms among regions  MS>0 geographical concentration of business:  MS<0.02 low concentration  MS>0.05 high concentration |
| **Shannon’s entropy H** | From 0 to ln(n)  0 for full geographical concentration of industry (all firms from given sector in one region)  Max for equal share (fully equal allocation of sectoral business to regions) |
| **relative entropy 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 employment between regions within the sector  Max for fully unequal distribution of firms among the regions |

*Source*: Own concept

**Figure 2.4: Geographical concentration indicators for example data**





*Source*: Own calculations

***Overall and detailed concentration measures for whole economy***

There are also indicators of concentration for the whole economy – yielding a single value for all sectors and regions as well the opposite model, yielding separate values for all sectors and regions. Their value are incomparable because of different scales.

**Table 2.43: Summary of single-value indicators**

|  |  |
| --- | --- |
| **Indicator** | **Rules of interpretation** |
| **Geographic concentration index** | 0 for no geographic concentration (full geographic diversification)  1 for full geographic concentration (no diversification) |
| **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 bigger the gap  Proportion of regional and inter-regional components shows the source of concentration |

*Source*: Own concept

**Table 2.44: Summary of indicator for every cell - LQ**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **LQ** | **region 1** | **region 2** | **region 3** | **region 4** | **region 5** | **region 6** |
| **Industry 1** | 0.22 | 0.49 | 0.91 | 1.61 | 0.65 | 0.59 |
| **Industry 2** | 0.21 | 1.75 | 1.02 | 0.90 | 0.32 | 1.07 |
| **Industry 3** | 1.23 | 0.66 | 1.03 | 0.78 | 2.60 | 0.11 |
| **Industry 4** | 3.42 | 1.09 | 1.06 | 0.44 | 0.34 | 2.95 |

*Source*: Own calculations

**Table 2.45: Interpretation of LQ indicator**

|  |  |
| --- | --- |
| **Indicator** | **Rules of interpretation** |
| **LQ** | LQ>1 concentration of activity in the region  LQ>1.25 potential exporter  LQ<1 underrepresentation of activity in the region, potential importers |

*Source*: Own concept

\* \* \*

Practitioners after reading this chapter will stay with the question, which indicator should I use? The answer is not trivial.

First of all the above overview was conducted on an artificial dataset, which assumed both small and big regions as well as an equal and unequal allocation of activity. The collection of all indicators shows a trend to their behaviour. Anyway, there might be a dataset or different aggregation level which will reveal other patterns of indicators (exceptions and tricky cases are always possible). Also the variance (and coefficient of variation) differ, which proves that sensitivity of measures is different. This can generate small shifts in rankings of regions or industries when different measures are applied. Wider analysis to conclude better on the behaviour of measures was presented in Chapter 4.

Secondly, one should not neglect the works of the last 80 years, as most of the improvements in the indicators were an answer for some problems and issues in measurement. Anyway, when the data input is similar, one cannot expect very different results – new indices cannot make a revolution in results. The most important issue is to see the similarities and differences in the measures, their references and components included, which impact the final value of the measure.

Thirdly, it is to know which indicator measures what. Many papers and studies interpret them not very rigorously, often making the conclusions exaggerated. As classified above, there are measures for regions – and they analyse sectoral composition and concentration patterns; for industries – and they analyse geographical concentration and business allocation patterns, and overall and very detailed measures which look globally or locally at sectoral and geographical concentration simultaneously. In fact, neither of the indicators can answer the question on agglomeration patterns, which is highly reserved for distance-based indicators, operating on individual points, not aggregated data.

Finally, this chapter was to collect measures, which are well dispersed in the literature. It works as a guide for users and practitioners willing to apply measures to their cases. Appendix 3 includes codes to R to operationalise these quantitative concepts presented here.

1. We omit in further descriptions some of indicators as an agglomeration index (by Uchida & Nelson, 2008), which is based on grid for population and population density with regards to borders of city center and travel time. [↑](#footnote-ref-1)
2. “*The NAI is accepted as a more reasonable standard with which to gauge a region’s industry structure than other alternatives” (Sherwood-Call, 1990)*. [↑](#footnote-ref-2)
3. In inter-sector analysis for firms one can count also the equivalent number of firms of the same size that would give the same entropy as in studied distribution of firms (see i.e. Nawrocki & Carteer, 2010). [↑](#footnote-ref-3)
4. Entropy measure was found to be asymptotically normally distributed and allows for inference from random sample (Wasylenko & Erickson, 1978). [↑](#footnote-ref-4)
5. Bickenbach, Bode, and Krieger-Boden (2012) propose relative and absolute Theil index decomposition. [↑](#footnote-ref-5)
6. Consequences also appear for entropy when companies merge. [↑](#footnote-ref-6)
7. Concentration measures can be applied to a single firm in an industry as well to industrial distribution within the region. Ávila et al. (2010) give a good review of concentration measures applied to individual firms, as well as methods to obtain concentration measures (as HHI) from aggregate data. A review of concentration measures for firms from the banking industry is in Bikker & Haaf (2002). [↑](#footnote-ref-7)
8. GE(1) is the notation from Generalized Entropy GE(α), where α is the parameter, here α=1. See for more Bruelhart & Traeger (2005). [↑](#footnote-ref-8)
9. Most of the texts referring to negative EG cites the phrase: “*The Ellison- Glaeser index can be negative if, by design or agreement, establishments are located far from each other to prevent competition (which could explain* *the negative index for sports teams and clubs) or to provide more uniform geographic coverage than the population (which could explain monetary authorities and blood and organ banks).*” [↑](#footnote-ref-9)
10. The European NTS system organizes the statistical division of the territory. NTS5 are the lowest territorial units (communes), which grouped give NTS4 (provinces). It is grouped to NTS3 (subregions) and then to NTS2 (regions). Level NTS1 is treated as macro-regions and NTS0 the whole country. The general rule of decomposition is that lower level units belong only to one upper-level unit, so no cross-belonging is possible. [↑](#footnote-ref-10)