Concentration of Economic Activities in the Kingdom of Saudi Arabia

Salim BOURCHID ABDELKADER
King Khaled University, Kingdom of Saudi Arabia
smoualdi@kku.edu.sa

M'hamed HAMIDOUCHE
Djilali Bounâama University, Algeria
mah_hamdouche@yahoo.fr

Abstract
The comprehension of economic aggregates at the spatial level is important because it allows knowing the characteristics of each geographic region. So, the estimates of the economic activity's concentration in the Kingdom of Saudi Arabia that we developed are based on aggregated data from the national system of economic statistics which are based on an administratively determined geographical area scale. These concentrations are expressed as indexes, known as Hoover's which reveals a high concentration of the trade and construction sector in the Kingdom of Saudi Arabia, while the Kingdom's Gini index make known that was an equal distribution of wages. Whereas, the similar indexes which are Ellison-Glaeser and Maurel & Sédillot, reveal a strong concentration located in the cities of Riyadh, Makkah and Easte province. In addition, the Herfindahl index illustrates that Mining and quarrying is the sector that provides more jobs. And from another point of view, the Krugman index indicates that financial and insurance sector activity is a specialization in the Kingdom.
It's therefore appropriate to note that our concentration measurements are derived from aggregated data on economic activities in a spatial dimension defined by the administrative areas of the Kingdom of Saudi Arabia.

Keywords: spatial concentration; specialization; economic integration;

JEL Classification: P35; P47; R12;

DOI: http://doi.org/10.24818/ejis.2020.02

1. Introduction

The location of enterprises by sector of activity is expressed by geographical concentration indexes which very often indicate an uneven distribution of activities in a defined area. However, Duncan (1957) indicated that the results of the measurements of these indexes depend on the chosen configuration of the division of a given space. Thus, the measurement of the concentration indexes through geographical division allows responding to several questions: is the economic activity homogeneously distributed over the territory - national, regional or local level? Have there been any increased changes in activities for certain sectors and in favour of which area? Will recent trends make a significant contribution to the emergence of local specialization? Which areas are specialized and in which sectors are they oriented?

More precisely, the index of geographical concentration shall be determined by the manner in which employees are dispersed in the territory. In the case of a description of the structure of activities at the local level, i.e. the specialization of the area, we calculate the distribution of activities in the geographical area with the different sectors. This measure is often referred to as a sectoral concentration.

This study relates the various indicators of concentration indexes for the Kingdom of Saudi Arabia by sector of activity, such as: Gini, Ellison and Glaeser, Maurel and Sedillot, Herfindhal and Krugman Specialization index in order to analyse both the productive structure of an area by comparisons with a reference territory and study the geographical distribution of the activities of a sector in the different areas of a territory.

We also suggested a simplified and reduced methodology for the calculation of Gini index, which can be reproduced for all other indexes. In order to boost these calculations, we added the dimension of distance through the estimation of absolute and relative moments. We clarify that the data obtained in our analysis derive from General authority for statistics of Kingdom of Saudi Arabia. In fact, we would like to point that these data contain certain null values (extreme values), even though, they are extracted, the rest of the values don't fit the normal distribution (Fisher and Tippett, 1928; Gnedenko and Kolmogorov, 1954; Gumbel, 1958; deHaan, 1970; Pickands III *et al.*, 1975). In this regard, we consider that for any international comparison between countries, the values of geographical concentrations or specialization, whether obtained from aggregated or disaggregated data, require a logarithmic transformation (Gibra, 1931; Cowell, 1988; and Morrisson, 1996). Finally, our research will focus on the descriptive methodology that consists of analysing and explaining the obtained results and the all data.

To close, our paper is structured as follows: Introduction; Literature review on concentration concepts and measures; The limits of spatial distribution measurements; Zoning systems and concentration indexes in the Kingdom of Saudi Arabia; Concentration indexes based on distance (the relative moments, the absolute moments), and Conclusion.

2. Literature Review on Concentration Concepts and Measures

Von Thünen (1826) was considered to be a pioneer in localization studies considering the structure of spatial space around towns. Using a theoretical framework, he described the emergence of areas specializing in agricultural production, the specialization of which depends on the distance from the city-market. After, Marshall (1920) and Weber (1909; 1929) distinguished their work on the localization of industrial production. Then, the aim of Gini's indicator (1921) was to measure the inequalities in the distribution of income and the concentration of the allocation and the peculiarity which was deduced from the curve of Lorenz (1905). In his book, Gibrat (1931) made no distinction between measures of inequality within population and measures of industrial concentration, he reserved the term concentration for firms, which was defined as the unequal distribution of firms by importance. In addition, Simpson (1949) developed an index to measure the degree of concentration in which individuals were classified by type. The square root of this index was already demonstrated by Hirschman in 1945. The index was rediscovered by Herfindahl (1950). Later, according to Hart (1956; 1971), the Theil Index was used to calculate total industrial concentrations. In fact, Theil (1967) concentrated on income inequality through information theory.

From the point of view of the location of firms in the same sector, Arrow (1962) used the externalities of Marshall (1890) and emphasized on the knowledge spillovers that exist between firms. According to Arrow, local industrial specialization will be an effective way of growing knowledge externalities between firms and thereby boosting growth processes. After, Hall and Tideman (1967) compared the estimates of concentration from both measures and found that they produce similar results and stated that the concentration and inequality were different concepts. Jacobs (1969) clarified that the most significant transfer of knowledge comes from outside the sector. As a result, the variety and diversity of industries, which are contiguous within the territory, promotes innovation, growth, local competition and accelerates the adoption of technology. So, Rao (1969) used a matrix approach to show that the decomposition of the Gini index was written in a quadratic form which allows the identification of the weighted average concentration indicators within groups and sub-groups. In another context, Atkinson (1970) argued that at the basis of any synthetic statistics of inequality there was a concept of social well-being, and that the analysis should focus explicitly on this aspect. Afterwards, Pyatt (1976), with the matrix approach, as well as the game theory, gave a new interpretation of the Gini coefficient in the sense of the expected value (mathematical expectation). He introduced the fact that each individual can compare his income with that of another, which would be drawn at random. In his study on income inequality in OECD countries, Sawyer (1976) listed all indicators of inequality, including indices of classical concentration and entropy. Hannah and Kay (1977) proposed the variation of the Hirschman-Herfindahl index. But in the research of Shorrocks (1980) four properties should be respected by inequality indices: normalization, symmetry (or anonymity), invariance to replication (the Dalton population principle), the transfer principle, and continuity. Cowell and Mehta (1982) proposed a rule of thumb derived by interpolating the histogram of the industry's participants. They estimated the upper and lower bounds of the tail of the Lorenz Curve from the concentration ratio at different levels, and then, took a linear combination of these bounds to estimate the Gini coefficient of the industry.

Then, Dagum (1987) presented eight definitions to obtain a decomposition of the Gini index into three groups. From the point of view of technological externalities, Porter (1990) has developed the idea that they essentially increase within the same industry and that specialization is potentially favourable for growth, both for the industry and the agglomeration in which it is located. From another viewpoint, Krugman (1991a; 1991b) explained the physical advantages of decreased transport costs, a condition that enables consolidation within a single area while benefiting from economies of scale, thereby increasing both concentration and specialization. Ellison and Glaeser (1994) proposed an indicator based on a location choice model derived from the behaviour of firms, which deduces that a non-random distribution of economic activities means that decision-making by firms depends on variables such as technological externalities and/or natural advantages. Subsequently, Ellison and Glaeser (1997) measured the concentration levels of 459 industries four-digit classified and they found, at the state-level of spatial analysis, that over 97% of United States four-digit industries are agglomerated. In other context, Ciccone and Hall (1996) find that the knock-on effects of economies of scale are able to compensate the consequences of congestion. Later, Maurel and Sédillot (1999) have introduced improvements to the Ellison and Glaeser index and observed that there is more variation between the indices in relation to less localized industries. From another point of view, Combes (2000) concluded that the specialization and average size of firms had a negative effect in most sectors and that the diversity effect was less obvious. However, Duranton and Overman (2002) explained the difficulties associated with the measurement of concentration. In addition, they integrated distance to determine the density resulting from the measurement of the distribution of geographic distances between pairs of firms in the industry. Furthermore, Devereux *et al.* (2003) compared the Maurel and Sédillot indices of French and British industries with the Ellison and Glaeser indices of American firms and found that the results converged. The following Table 1 present a summary of the notions and properties of principal indexes which are cited.

Table 1. Notions and properties regarding the principal concentration indexes

Authors	Notions and proprieties	Observations
Von Thünen (1826)	The first modelling trying to explain the optimal allocation of land from freight transport to the centre of the city	-
Launhardt (1885)	Modelling the problems of horizontal and vertical differentiation of firms	-
Marshall (1890)	The geographical proximity of economies linked to spatial agglomeration (reduction of transport costs, availability of a specialized and stable labour force and, best dissemination of knowledge). Potential specialization of the territory in case of high concentration of a sector.	-
Weber (1909,1929)	The optimal location is the one that minimizes the total transportation cost between input and output suppliers. This model has the advantage that it is dependent only on distance and not on belonging to geographical areas.	-
Gini (1912)	The Gini index measures spatial phenomena and allows estimating the concentration by analogy of inequalities based on the Lorenz curve.	discrete index : the space is a part of a geographical entity and used in several fields
Hoover (1936)	Equidistributional index and allows comparing the distribution of a variable with a reference variable. These may be economies of internal scale specific to the company, sectoral economies or economies of urbanization.	
Herfindahl (1950)	The Hirschman-Herfindahl index is widely used in industrial economics to measure market concentration and to investigate the existence of an oligopoly or cartels in particular. This index has also been used as a measure of economic diversity and for macroeconomic specialization analyses.	
Theil (1967)	Decomposability index builds on information theory. It has been implemented for the analyses of specialization and concentration. The observations are weighted using their correspondent relative scores between regions and within regions.	discrete index: the space is a part of a geographical entity
Openshaw (1984)	Introduction of the concept MAUP (Modifiable Areal Unit Problem). It's a question of the influence of spatial partitioning on the results of statistical processing or modelling.	Boundaries aren't determinative due to the consideration of distances.
Krugman (1991a)	In order to understand trade, we must firstly understand the concentration of production. It calculates the share of aggregate which would have to be relocated to achieve an economics structure equivalent to the average structure of the reference group.	Index among the specialization measures.
Ellison & Glaeser (1997)	E&G and M&S indexes : • Are classified as discrete indexes: the space is a part of a geographical entity. • Takes into account the agglomeration of all activities in order to measure the concentration of a sector as well as the structure of the sector of activity. They allow verifying if the concentration of establishments is due to the geographical distribution of all activities over the whole territory.	E&G uses a probabilistic evaluation approach and is based on a theoretical localization model
Maurel & Sédillot (1999)		M&S indexes is a model that attempts to improve the index (Ellison & Glaeser, 1997) based on a location choice model.
Duranton & Overman (2002)	An index which allows comparison between different territories and which are insensitive to a modification of boundaries or to the choice of geographical unit and to the modification of sectoral classifications. Index that evaluates the density of establishments of a sector at each possible distance on the reference territory.	Continuous Index. Requires geolocalized data

Source: Authors' research

Lastly, within a spatial dimension, Rysman and Greenstein (2005) have developed a test of Multinomial Test for Agglomeration and Dispersion. Mori *et al.* (2005) proposed a statistical measurement index for the agglomeration based on a comparison of the divergence index of Kullback and Leibler (1951) for an industry that is a model of reference for spatial dispersion. Guillain and Le Gallo (2007) combined a discreet-space agglomeration model with a continuous space model. They distinguished between clustering, which can be defined by discrete space measurements, and agglomeration, for which clustering is a necessary but not a sufficient component. However, the coverage of the space for studying the location of firms differs according to the authors, some of which take into account that the area of study is a square area (Møller and Hakon, 2014), rectangular (Cole and Syms, 1999), circular (Szwagrzyk and Czerwczak, 1993), an administrative area (Arbia *et al.*, 2012) or a zoning (Lagache *et al.*, 2013).

3. The Limits of Spatial Distribution Measurements

One of the major issues in the study of spatial data is the statistical distortion generated by the use of surface units. Nevertheless, the use of administrative units for analysis can be justified by the difficulty of collecting data on a small scale or due to other forms of units. Therefore, the option of a division space has an effect on the results obtained. We note though, Gehlke and Biehl (1934) were the first to report that statistics could vary significantly from one zoning system to another. They observed that in the United States, the correlation between male youth delinquency and median monthly equivalent rent decreases monotonically with the scale of the area units. In addition, zoning figures can vary depending on the size of the units and the configuration of their limits (their shape). Thus, as reported by Kendall and Yule (1950) that: " *... we must insist on the need......., don't lose focus on the fact that our results depend on our units*". A new suggestion has been applied by Cramer (1964) regarding the statistical aggregation. And even though some authors such as Cliff and Ord (1973) discussed the question of spatial autocorrelation, or Martin *et al.* (1979) implemented the analysis of spatial-temporal processes to provide solutions to statistical problems, but the subject of spatial aggregation has not been considered. It's recognized that different meshing criteria give different results, such as optimizing the variance of the independent variable in compared to the other variable. The scope of research on this topic has grown over time, according to various viewpoints, approximately equal area, equal population (Blalock, 1964; Hannan, 1971), equal density of Sammons (1976), the maximum compactness of the zones and maximum spatial entropy of Batty and Sammons (1978), the maximum intra-zone homogeneity of Cliff *et al.* (1975) and minimization of the standard error of the regression coefficient of Williams (1976). Nonetheless, the reflection on the comparability of data according to electoral mapping systems stems from research in electoral geography to determine the extent to which electoral district divisions could affect election results (Taylor and Gudgin, 1976).

The impact of the change in scale is known in geography as MAUP (Modifiable Areal Unit Problem). It's defined as the manner of aggregating data in spatial units has a significant impact on the result, in particular on the search for explanatory factors using correlation Openshaw (1984). The observation scale describes the quality of the information of the spatial structures studied (Piron, 1992, 1993) and that the measuring instrument (or the observation tool) be adapted (Marceau, 1999; Mandelbrot, 1977; Nottalen, 1993). To solve these difficulties, other authors have provided detailed analyses using simple univariate statistics, such as mean and variance (Haggett, 1965, 1973; Sanders, 1989). In the same way, others have used multidimensional studies such as principal component analysis and correspondence factor analysis (Rozenblat, 1989). A multivariate analysis of the determinants of average family income for different zoning systems was carried out by (Fotheringham and Wong, 1991), they found the results to be highly unreliable in multivariate analysis of data from areal units. They also find a wide range of correlation and regression coefficients, which are positively (or negatively) significant for some data configurations, but not significant for others.

Amrhein and Reynolds (1996, 1997) have shown that distortions in size and form depend on whether the information is averaged or summed, as well as on the spatial organization of the raw data, such as the spatial autocorrelation coefficient. In fact, the scale covers two realities, one quantitative and the other qualitative. For the qualitative notion, Lacoste (1976) has introduced the multi-scalar concept (several scales), which consists in classifying the different categories of spatial entities, not according to the scales of representation, but according to their difference in size in reality. The quantitative concept refers to the relationship between the measurement of distance in the map and the measurement of distance in the earth (Baudelle and Regnauld, 2004; Volvey, 2005). Other authors as Marcon and Puech (2003, 2012) and Floch (2012) consider that space is not divided into fixed zones but into continuous areas and they introduce measures of spatial concentration of activities based on the distance between the units considered.

4. Zoning Systems and Concentration Indexes in the Kingdom of Saudi Arabia

The Kingdom of Saudi Arabia is divided into 13 administrative areas (AA), which are further divided into 136 governorates (G) divided in turn into 1530 centres (C). The data collected according to this configuration of administrative frontiers cannot be considered to be an accurate reflection of economic phenomena. In order to avoid this difficulty, some authors prefer to work with grids drawn arbitrarily in such a way that the smallest spatial unit rests on a clear economic basis. So, our empirical analysis is based on 2017 sectoral data at household level for the 136 governorates (entities forming 13 administrative zones). Aggregation in the above-mentioned large-scale zoning systems provides aggregates for employment, the number of production sites and economic units, wages and 26 industrial zones. As shown in the table below, depending on the zoning system, the economic characteristics differ greatly.

Table 2. Summary statistics (2017)

Zoning system	Administrative Area (AA)	Governorates (G)	Centres (C)
Number of units	13	136	1530
Land economic area (km^2)	2,111,908	129,183	471,808
Employment (workers)	1,211,982	93,229	8,912
Employment density (workers/km2)	0,57	0,273	0,026
Average National wage (local currency)	6,093		
Inter-area distance (km)	4,955	112,785	37
Operating surplus (Local money)	113,824,454	106,816,228	751,074

Source: https://www.stats.gov.sa

In the absence of a grid zoning system in the statistical system, we adopt a system that is more adaptable with units according to the administrative areas used on variables taken from the areas (the jobs) in a simple random or systematic random or stratified or multi-stage and clustered manner. For our case study, employment for the year 2018 in the Kingdom of Saudi Arabia by sector of activity and administrative area will constitute the element for calculating the spatial concentration.

4.1. Hoover index

The equipartition index of Hoover (1936, 1984) (or Robin Hood index) is defined as the proportion of employment that would have to be taken from the entities (administrative areas or activity sectors) with more than the average and redistributed to the rest of the entities with less than the average so that all these entities have the same number of employment. The index H is defined by: $\mathbf{H} = \frac{\frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|}{2\bar{x}}$ or $\mathbf{Hoover_k} = \frac{1}{2}\sum_{i=1}^{n}\left|s_{X_{ik}} - s_{\bar{X}_{ik}}\right|$

$s_{X_{ik}}$ and $s_{\bar{X}_{ik}}$ represent the proportion of the sector of activity (entity **i**) in national employment, respectively for the sector **k** (all sectors).

And for more visibility, in the Hoover dispersions due to the weight of the area (intra-sector) or the sector proper, the calculations in Table 3 have been formalized according to the following formulas: $\mathbf{E}_{\text{intersector}_i} = \frac{S_i - \bar{S}_S}{\sum_{i=1}^{n}|S_i - \bar{S}_S|}$ and $\mathbf{E}_{\text{intrasector}_i} = \frac{S_i - \bar{S}_{r_i}}{\sum_{i=1}^{n}\left|S_i - \bar{S}_{r_i}\right|}$ with:

$\mathbf{S_i}$: Number of employment in a given sector **i**.
$\mathbf{\bar{S}_S}$: The average number of employment in all sectors.
$\mathbf{\bar{S}_{r_i}}$ **:** The average number of employment in a sector compared to the total number of jobs in a given area.

The results of Hoover's index of the Kingdom of Saudi Arabia's activity sectors and dispersions for intra and inter-sectors are provided in the following table:

Table 3. Hoover index in %

Sectors	HOOVER Index	Dispersions Inter-sector	Dispersions Intra-sector	Spread = $\lvert$inter − intra dispersions$\rvert$
Other service activities	0.02	13.35	0.02	13.33
Agriculture, forestry and fishing	0.97	12.59	1.05	11.53
Electricity, gas and Water	1.04	12.54	1.12	11.41
Mining and quarrying	1.92	11.84	2.08	9.76
Transportation and communication	3.89	10.26	4.21	6.05
Other collective and social services	7.07	7.73	7.66	0.07
Manufacturing	9.29	5.96	10.07	4.11
Financial and insurance activities	9.50	5.80	10.29	4.50
Trade	23.57	5.40	25.54	20.13
Construction	35.04	14.53	37.96	23.42

Source: Authors' research

Hoover index expresses the proportion of employment in sector **k** that would have to be shifted between the entities **i** in order to achieve a geographical distribution equivalent to that of employment in all sectors combined. It varies between 0% (perfect equipartition) and 100% (theoretical maximum concentration). So, Hoover equal 23.57 for trade and

35.04 for construction that respectively indicates that 23,57 % and 35,04% of wages of the above sectors would have to be redistributed in order to achieve perfect equality (everyone's income is average income). For the rest of the sectors, the Hoover index is less than 10% which reflects that is a moderate equilibrium distribution of wages.

The ratios of the dispersions respectively equal to 5.40 for trade and 14.53 for construction of inter-sector and 25.54 and 37.96 for intra-sector of construction advocate that the disparity in the distribution of wages comes much more from the wages distributed for other sectors than from the sector proper.

4.2. Calculation of the Gini index with a random zoning system

On the zoning systems, employment at the level of the spatial units of the Kingdom of Saudi Arabia (13 administrative zones) is taken into account in return for the 10 economic activities.

The National Statistics Office of the Kingdom of Saudi Arabia has defined a set of 130 partitions of different administrative areas of the Kingdom which contain 1530 centres (spatial units). Thus, we proceed to estimate the Gini index according to the data provided by this office, which is shown in Appendix 1.

We indicate that the principle of calculating by random system, the Gini index from the data provided by the statistical system, is to simplify these calculations which are usually estimated by sector as established in Table 4 below. These estimates will be reduced to a single result for the whole country according to a matrix approach. We point that the Gini index can be written in a general as: $G = \sum_{i=1}^{k} q_{ij}$ and, can be rewritten as a matrix form according to: $G = p * q$.

Pyatt (1976) interpreted the Gini index as an expected value, in the sense of mathematical expectation, and he has been established that the Gini indicator is expressed as a matrix as: $G = (\mu'.p)^{-1}.q.E.p = \tau'.E^{*}.p$ (E^* is the normalized matrix of E), his approach was similar to Bhattacharya and Mahalanobis (1967).

In the continuity of Pyatt (1976) a new matrix approach of the Gini index has been formulated by Silber (1989) expressed by the following matrix: $\mathbf{G} = \mathbf{e}'.\mathbf{G}^{*}.\mathbf{s}$ (where e' is a line vector, s is a column vector and G* is a matrix (p x p).

After the data have been collected and arranged in matrix form (p x q), therefore, the steps to be employed as follows to estimate Gini index by random selection of data:

- We randomly choose an area, referred as the ($\mathbf{G}_{i,j}$), within the matrix of administrative areas ($\mathbf{G}_{13x10}$) which are in appendix 1. The data in the row and column assigned to this area is discarded in order to obtain a matrix. ($\mathbf{G}_{i-1,j-1}$), ie the matrix $G_{(12,9)}$. In practice, this means that all employment data for the administrative area are discarded which are randomly selected (column) and cross-referenced to the relevant economic activity (row). And we continue with another area designated ($\mathbf{G}_{i',j'}$) and so on. We execute the algorithm **n** more, with $\mathbf{n} = \frac{\frac{z^2 p(1-p)}{e^2}}{1+(\frac{z^2 p(1-p)}{e^2 N})}$ and:
  n: Sample size to be calculated
  N: Population size 130 in our case.
  e: Margin of error (we opt for 25%)

z: Confidence interval and a confidence level of 85 % is taken.
p: standard deviation (i.e. 50 % or p = 0,5)
For n ≈7 (corresponds to the cells colored yellow in table of Appendix 1).

- With a confidence level of 85% (**z-score** = 1,44) and for a proportion $\hat{\mathbf{p}}$, the confidence interval is given by the following formula:

$$\left[\hat{\mathbf{p}} - \mathbf{z}_{1-\frac{\alpha}{2}}\sqrt{(\mathbf{1}-\mathbf{f})\frac{\hat{\mathbf{p}}(\mathbf{1}-\hat{\mathbf{p}})}{\mathbf{n}-\mathbf{1}}}\ ,\ \hat{\mathbf{p}} + \mathbf{z}_{1-\frac{\alpha}{2}}\sqrt{(\mathbf{1}-\mathbf{f})\frac{\hat{\mathbf{p}}(\mathbf{1}-\hat{\mathbf{p}})}{\mathbf{n}-\mathbf{1}}}\right]$$

(Look appendix 2, among the 84 values of Gini indexes (7 x 12), 7 values indexes of Gini have been used for the matrix calculation, so, the proportion of the sample is $\hat{\mathbf{p}} = \frac{7}{84} = 0.08$). so, accordingly, with 85% of confidence level, the confidence interval level calculation is :

$$\left[\frac{7}{84} - 1{,}44\sqrt{\left(1-\frac{84}{130}\right)\frac{\frac{7}{84}\left(1-\frac{7}{84}\right)}{84}}\ ,\frac{7}{84} + 1{,}44\sqrt{\left(1-\frac{84}{130}\right)\frac{\frac{7}{84}\left(1-\frac{7}{84}\right)}{84}}\right] \approx [0.05\ ;\ 0.10]$$

- For purposes of further comparison, we will adopt the most widely used formula given in the World Bank's explanatory memorandum for the calculation of inequality in the incomes Brown (1994), which is as follows: $G = 1 - \sum_{i=1}^{n}(X_i - X_{i-1})(Y_i - Y_{i+1})$. Thus, the Gini indexes associated with each administrative area with random elimination of a given sector and the related administrative area are represented in Appendix 2.

- The Spearman correlations of the Gini spatial concentration index were estimate in order to choose the most correlated indices of Gini with two conditions, greater than 95% and the P-value (the significance) is less than 5%. The results of the calculations were presented in Appendix 3 & 4.

- So, we obtain the Gini index matrix $G_{(1,n)}$ and the matrix $G_{(n,1)}$ either for our case were $G_{(1,7)}$ and $G_{(7,1)}$. After, step is to obtain two matrices, the first one, a column matrix and the second one, a row matrix $G_{(1,n)}$ and $G_{(n,1)}$. The multiplication between these two matrices will give the value of the Gini index for the whole country, as a result of matrix multiplying: $[G_{(1,n)} \times G_{(n,1)}] = G_{(1,1)}$. Following the estimates listed in appendix 2, 3 and 4 and the two conditions proposed below (spearman correlation and p-value), the pair GINI_1 and GINI_4 was retained. The multiplication of these two matrixes for our case of Kingdom of Saudi Arabia, gives $\mathbf{G}_{(1,12)}$ and $\mathbf{G}_{(12,1)}$, thus, the value of the Gini index is calculated as follows:

$$\mathbf{G}[(1,12) * (12,1)] = \begin{bmatrix} 0.3760 \\ 0.4865 \\ 0.3046 \\ 0.3278 \\ 0.3713 \\ 0.4850 \\ 0.3716 \\ 0.3224 \\ 0.4737 \\ 0.4695 \\ 0.3361 \\ 0.2098 \end{bmatrix} \times [0.3810\ \ 0.4865\ \ 0.3072\ \ 0.3439\ \ 0.3815\ \ 0.4909\ \ 0.3712\ \ 0.4115\ \ 0.4731\ \ 0.4589\ \ 0.3758\ \ 0.2251] = \mathbf{0.1432}$$

Gini index of the Kingdom of Saudi Arabia = 14,32%

The Gini coefficient is between 0 and 1, where 0 represents perfect equality (everyone has the same resources) and 100 represents perfect inequality (resources are monopolized by a single person or category). So, the ratio 14,32% implies that no sector has a dominant position in the distribution of revenues. Consequently, for a better distribution of wages in the Kingdom, it would be advisable to take into consideration for each area, the density of employment (calculated on the basis of the ratio of the number of employment to the active population). Thus, the areas with the lowest rankings (in terms of employment density) should benefit both from fiscal and para-fiscal advantages and financial support within the budgetary policy context (spatial approach).

It should be noted that in the case when the Gini index was above or around 50%, these supports will be allocated to the:

- Disadvantaged sectors in order to achieve social justice (see the order of the Gini indexes in table 4, which 6 sectors were near 40%).
- The specialized sectors in order to increase the country's richness (take into consideration the higher ranks of Krugman index; see the table 4: financial and insurance activities, construction and mining & quarrying).

4.3. Formulas of Ellison & Glaeser and Maurel & Sédillot index

The index of Ellison & Glaeser (1997) is derived from the theoretical localization-choice model and has not been subject to criticisms such as those made to the Gini index. The Ellison-Glaeser index is defined as: $\boldsymbol{\gamma^{EG}} = \frac{\sum_{i=1}^{M}(s_i - x_i)^2 - (1-\sum_{i=1}^{M} x_i^2)\sum_{j=1}^{N} z_j^2}{(1-\sum_{i=1}^{M} x_i^2)(1-\sum_{j=1}^{N} z_j^2)}$

Hence:

$\mathbf{s_1}, \mathbf{s_2}, \ldots, \mathbf{s_M}$, the employment part of an industry in each of the regions.
N, the number of employment in an industry.
$\mathbf{z_1} \ldots \mathbf{z_j}$, the part of each employment in the total employment of the sector.
M, the number of administrative areas in the country.
$\mathbf{x_1} \ldots \mathbf{x_i}$: the part of each region in total employment.

Remind that the spatial Gini index: $\mathbf{G} \equiv \sum_{i=1}^{M}(\mathbf{s_i} - \mathbf{x_i})^2$ and Herfindahl: $\mathbf{H} = \sum_{j=1}^{N} \mathbf{z_j^2}$
Consequently, the index of Ellison and Glaeser is written as:

$$\boldsymbol{\gamma^{EG}} = \frac{G - \left(1 - \sum_{i=1}^{M} x_i^2\right) H}{\left(1 - \sum_{i=1}^{M} x_i^2\right)(1 - H)} = \frac{\frac{G}{\left(1 - \sum_{i=1}^{M} x_i^2\right)} - H}{(1 - H)} \equiv \frac{G - H}{(1 - \mathrm{H})}$$

Thus, the Ellison-Glaeser index is better adapted to compare economic aggregates for different spatial units.

Maurel and Sédillot replace the probability **p** with the frequency estimator, weighted by the size of the sectors employment (or firms) as follows:
The estimator of $\boldsymbol{\gamma}$ is then written: $\boldsymbol{\gamma_{EG}} = \frac{\mathbf{G_{EG} - H}}{\mathbf{1 - H}}$, they gets: $\boldsymbol{\delta_{MS}} = \frac{\mathbf{G_{MS} - H}}{\mathbf{1 - H}}$
With: $\mathbf{G_{MS}} = \frac{\sum_{i=1}^{M} s_i^2 - \sum_{i=1}^{M} x_i^2}{1 - \sum_i x_i^2}$

This implies that: $\boldsymbol{\gamma_{EG}} = \frac{\frac{\sum_{i=1}^{M} s_i^2 - \sum_{i=1}^{M} x_i^2}{1-\sum_i x_i^2} - \mathrm{H}}{1 - \mathrm{H}} = \frac{\frac{\sum_{i=1}^{M}\left[s_i^2 - x_i^2\right]}{1-\sum_i x_i^2} - \mathrm{H}}{1 - \mathrm{H}}$

4.4. Estimation of Saudi Arabia's concentration and specialization indexes

Following the formulas announced above, we have estimated the different indexes of employment concentration in the Kingdom of Saudi Arabia for the different sectors of activity as well as the Krugman specialization index for the year 2018, summarized in the table below, and whose basic data are given in Appendix 1.

It's recalled that the Herfindahl index of the sector measures the gross concentration of employment, when it is high, it informs that the workforce is not very dispersed, which is obvious for the mining and quarrying. Thus, for $\mathbf{G_{EG}}$ indicates the probability that two firms from the same sector will be localized in the same administrative area, it remains high for the sector of other service activities and mining & quarrying (respectively $\mathbf{G_{EG}}$ ≈81% & 70%) and on average ($\mathbf{G_{EG}}$ ≈26%) for the rest of the sectors (relatively minor).

Table 4. Concentration and specialization indexes in Saudi Arabia

Sectors / index	Index of concentration					Index specialization
	Gini	Herfindahl	G_{EG}	$\gamma^{EG} = \frac{G-H}{1-H}$	$\gamma^{MS} = \frac{G_{MS}-H}{1-H}$	Index of Krugman
Other service activities	0.3876	0.1411	**0.8114**	0.7804	0.6938	0.0001
Agriculture, forestry and fishing	0.4030	0.2033	0.1944	-0.0111	-0.3151	0.0004
Electricity, gas & water	0.4514	0.3221	0.3132	-0.0130	-0.3703	0.0023
Mining and quarrying	0.4473	0.7098	**0.7010**	-0.0303	-0.8650	0.0182
Transportation & communication	0.2763	0.2715	0.2627	-0.0121	-0.3446	0.0042
Other collective & social services	0.4080	0.2616	0.2528	-0.0119	-0.3399	0.0055
Manufacturing	0.2747	0.2388	0.2300	-0.0116	-0.3298	0.0134
Financial and insurance activities	0.3205	0.3965	0.3881	-0.0139	-0.4157	0.0510
Trade	0.4116	0.2515	0.2429	-0.0114	-0.3354	0.0053
Construction	0.3771	0.2340	0.2255	-0.0111	-0.3277	0.0256

Source: Authors' research

Moreover, Ellison-Glaeser index suggests that a sector is highly localized when its location correlation is greater than 5%, whereas if this correlation is between 2% and 5%, the sector will be said to be lightly agglomerated. A value of less than 2% would signify a sector in which the firms are not agglomerated (or just marginally consolidated in the territory). Applying this rule to $\boldsymbol{\gamma^{EG}}$, a single sector (other service activities) is highly localized while the rest of the sectors have low values and are highly dispersed (negative values indicate that firms tend to be more dispersed than the uniform distribution of activities). These results are near those of Maurel and Sédillot index, which are the exception in the sector of other service activities highly localized (zero values upset the estimates).

4.5. The relevance of concentration indexes

i- The principle of relevance

An index expresses a variable, based on measurements, and also represents as precisely as possible and necessary a defined phenomenon. An index is what is admitted to represent a criterion which is what is considered as a concern. Thus, the relevance of an index supposes that it possesses a quality. And according to UNESCO (2008), the qualities of index required:

- Pertinent: the data are relevant to decision-making and the issue to be measured;
- Timely: the data are made available quickly before they become out-of-date;
- Accurate: the data are correctly calculated and not subject to error;
- Frequency: the data collection can be repeated on a regular cycle to measure trends;
- Cost: data collection is not too expensive (few developing countries can afford dedicated surveys of more than top policy priorities);
- Valid: the data measure what they are intended to measure;
- Reliable: the data are stable, not changing too quickly to be captured;
- Consistency: indicators do not contradict each other, or individual responses contradict each other;
- Economy: it is preferable to pick the minimum number of indicators necessary in order to cover the maximum extent of the topic. This minimizes the burden of collection on countries;
- Independence: indicators should measure different aspects of a topic; they should not be inter-correlated though some indicators may be related;
- Transparency: the sources of data and how indicators have been calculated should be as clear as possible to the reader;
- Comparability: the use of data at the international level adds a further dimension of complexity, that data should be comparable across different cultures and economies.

And as per Maby (2002) five criteria are considered for the legitimacy or relevance of an index, and that are:

- Probative, i.e. it must allow to answer to a given problem;
- Discriminating, this purpose concerns, particularly geography in the sense that the different spatial units (countries, regions, localities) studied are distinguished from each other;
- Subsuming, the indexes should facilitate the elaboration of typologies of the aggregates studied;
- Analytically, the real is decomposed into elementary units, the process of creating indexes is associated with the search for inference chains of cause and effect links;
- Systemic, the indexes aim to restitute the complexity of a phenomenon by trying to uncover and render intelligible its different facets.

And for a relevance of quantitative or qualitative or composite or aggregated index, we consider that it must assure at the same the following three criteria:

- Measurement or representation criterion: index provides a representative measure of what is meant to be measured and involves the following three conditions:
  - Validity: must really measure the items that are supposed to be measured and extract an assertion or conclusion or be tested;
  - Reproducibility: a recalculation must give the same value;
  - Sensitivity: capacity to express significant variations.
- Operational criteria: easy to use, convenient and continuous monitoring and must assure at the same the following three criteria:
  - Simplicity: simple and usable with limited tools;
  - Availability of input data (cost and delay);
  - Ethics: compatible with social science values.
- Applicability criteria: useful in decision making and must assure at the same the following four criteria:

– Transparency: easy to understand and criticize by the user;
– Interpretable: intuitive and unambiguous interpretation;
– An objective related: the achievement of a measure in relation to an objective;
– Related to a decision: measures the factors that can be modified by a decision.

The all three criteria mentioned above ensure the performance of an index, but the advantages and disadvantages that accompany the index permit to deduce the global performance.

These advantages and disadvantages of an index are summarized as follows:

- Advantages (a-f, see Table 5):
    a. Can synthesis complex or multi-dimensional problems for decision making.
    b. Easier to interpret.
    c. Easy comparisons.
    d. Reduce the size of the data.
    e. Facilitates policy integration.
    f. Communicable to the public.
- Disadvantages (g-k, see Table 5):
    g. May cause the erroneous decisions
    h. May lead to simplistic conclusions
    i. Instrumentation, in case that the method of construction of the index is complex and can give results with low-quality data.
    j. The selection and allocation of indicators can be the subject of a political game.
    k. Hide weaknesses on certain dimensions.

So, the above performance criteria and their advantages and disadvantages are illustrated in table 5 relating to the global relevance of the concentration indexes.

Table 5. Global Relevance of the concentration index

Index/ criteria	Relevance										Advantage	Inconvenient
	Representation Y_1			Operationally Y_2			Applicability Y_3					
	Validity X_1	**Reproducibility X_2**	Sensitivity X_3	Simplicity X_4	Data availability X_5	Ethics X_6	Transparency X_7	Interpretable X_8	Linked to an objective X_9	Linked to a decision X_{10}		
Gini	X	X	X	X	X	X	X	X	X	X	b, c, e, f	g, h, i, j, k
Hoover	X	X	X	X	X	X	X	X	X	X	b, c, e, f	h, i, j, k
Herfindahl	X	X	X	X	X	X	X	X	X	X	b, c, e, f	h, i, j, k
Theil	X	X	X		X	X		X	X	X	b, c, e, f	g, h, i, j, k
Ellison & Glaeser	X	X	X	X	X	X	X	X	X	X	a, b, c, f	k
Maurel & Sédillot	X	X	X	X		X	X	X	X	X	a, b, c, f	k
Duranton & Overman	X	X	X	X		X	X	X	X	X	a, c, d, e, f	k
Krugman	X	X		X	X	X	X	X	X	X	b, c, e, f	g, h, i

Notation: The indexes highlighted are ensuring the three criteria at the same time (representation, operationally and applicability). Source: Authors' research

ii- Simple ranking of indicators according to relevance and global relevance

The indexes Gini, Hoover, Herfindahl and Ellison & Glaeser are relevant because they adopt ten criteria: validity, reproducibility, sensitivity, simplicity, data availability, ethics, transparency, interpretable, linked to an objective and linked to take a decision. In our case, the rank of the relevance of an index is in favour of the one that brings more advantages than disadvantages, i.e. the ranking is in the order of the number resulting from the gap between the number of advantages and number of disadvantages. From this approach, the ranking of the performance of the concentration indexes is summarized in Table 6 below:

Table 6. Ranking of the concentration indexes according to global performance

Rank	Indexes / criteria	Advantage	Inconvenient	Gap (Advantage - Inconvenient)
1	Ellison & Glaeser	a, b, c, f	k	4 – 1 = 3
2	Hoover	b, c, e, f	h, i, j, k	4 - 4 = 0
3	Herfindahl	b, c, e, f	h, i, j, k	4 - 3 = 1
4	Gini	b, c, e, f	g, h, i, j, k	4 - 5 = -1

Source: Authors' research

5. Concentration Indexes based on Distance

5.1. The relative moments

The measurement of the distance as the crow flies between two points on the territory is obtained by: $\mathbf{d(A,B)} = \sqrt{([(\mathbf{X_B} - \mathbf{X_A})^2 + (\mathbf{Y_B} - \mathbf{Y_A})^2])}$

Where $\mathbf{X_A}$ and $\mathbf{X_B}$ **:** point latitude of A and B & $\mathbf{Y_A}$ and $\mathbf{Y_B}$: point longitude of A and B With these coordinates expressed in radian and for the reason of the curvature of the earth whose radius is 6371 Km, the distance from A to B $\mathbf{d(A,B)}$ is (Dall'erba, 2004):

$$\mathbf{d(A,B)} = \mathbf{ArcCos(Sin(X_A))*Sin(Y_A)+Cos(X_A)*Cos(Y_A)*Cos(X_B - Y_B) * 6371} \qquad (1)$$

Employments based on distances give rise to the calculation of relative moments for each zone according to the following formula: $\boldsymbol{\mu_i} = \sum_{i=1}^{n} \mathbf{Z_i}.\,\mathbf{d_i(A,B)} = \mathbf{d_i(A,B)} * \sum_{i=1}^{n} \mathbf{Z_i}$

$\boldsymbol{\mu_i}$: Relative moment for each administrative area;
$\mathbf{d_i(A,B)}$: The average distance for each administrative area in relation to all the distances between all the towns in the same area (expressed in Km);
$\mathbf{Z_i}$: Number of employment, wages or number of firms in a given sector.

And the total moment of a country is: $\boldsymbol{\mu_{T_R}} = \sum_{i=1}^{n} \boldsymbol{\mu_i}$
$\boldsymbol{\mu_i} = \sum_{i=1}^{n} \mathbf{Z_i}.\,\mathbf{d_i(A,B)} \neq \mathbf{d_i(A,B)} * \sum_{i=1}^{n} \mathbf{Z_i}$ as all these distances need to be determined.

For our case, we have adopted the data of inter-city distances from General Authority for statistics for Kingdom Saudi Arabia (from Table 7) and the employment for each sector and administrative area given in Appendix 1.

In fact, the moment expresses the weight of economic agents or aggregates and the potential for displacement that requires energy. Thus, the moment relates the degree of mobility formulated on employments.Km. If we had more means we would divide this measure for each region by the average costs of employment (over a day as a unit of time) necessary for each employee to be at work expressed in currency and which represents the attraction of employment (the attraction to employment must be greater than the reaction of for not working, it would be unhelpful or irrational for a employment to provide more expenditure than income by going to that employment (or activity). The best formulation of the moment would be in unity employment.Km/$ US or employment.Km/SDR (Special Drawing Rights or SDR: money of International Monetary Fund).

The moments for each administrative area are estimated as follows.

Table 7. Relative Moment of employment by area

Area Administrative	Intra-area distance (km) Intra Governorate	Relative moment (Total x inter-distance) $\mu_i = \sum_{i=1}^{n} Z_i . d_i(A,B) = d_i(A,B) * \sum_{i=1}^{n} Z_i$
North bord	98.607	3887148
AL - Baha	8.887	369823
AL - Jouf	60.719	3499401
Tabuk	49.476	4486309
Hail	22.413	2445469
Najran	29.058	3391245
Jazan	10.362	1245620
Asir	14.608	4192008
Madinah	27.344	8354723
Qassim	39.203	12529522
Makkah	15.144	29609551
Riyadh	41.482	137866940
Easte. Prov	63.337	116013961
Total		327891719

Source: Authors' research

Relative Total Moment (RTM) = sum of the relative moments of all the areas

In our case the Kingdom of Saudi Arabia (KSA):

Relative Total Moment (RTM_{ASK}) Kingdom of Saudi Arabia = 327.891.719 emplois.Km

RTM_{ASK} = 328 Giga E.Km

The relative total moment 328 Giga.E.KM reflects the distances covered by income (wage) obtained.

The values obtained from the calculation of the relative total moment can be classified into three groups: smaller, moderate and significant value as shown in the following table:

Table 8. Value classification of the Relative Moment of employment

	Area with low employment	Areas with high employment
Condensed area	moderate value **(a)**	significant value **(b)**
Sparse area	smaller value **(c)**	moderate value **(d)**

Source: Authors' research

Then from Table 8, the smallest value (**c**) represents in the area that the wages are disadvantaged by the cost of transport (the areas concerned by this classification according to the values obtained in table 7: North bord, AL - Baha, AL-Jouf, Tabuk, Hail, Najran, Jazan, Asir and Madinah).
For the moderate value (**a**) or (**d**) which describes the proximity of wages (the areas concerned by this classification according to the values obtained in table 7: Qassim and Makkah). The significant value (**b**) reveals that wages are intended to be well considered in addition to the reduced cost of transportation (the areas concerned by this classification according to the values obtained in table 7: Riyadh and Easte Province).

5.2. The absolute moments

For a point distribution, the measurement of the central tendency of the distribution allows to calculate its centre of gravity. Its coordinates are calculated using **X** and **Y** of each of the points composing the distribution. It is therefore the point of equilibrium of all the points, the average localization. The coordinates of the centre of gravity are calculated as follows:

$$(\overline{\boldsymbol{X}}_{\boldsymbol{GC}}, \overline{\boldsymbol{Y}}_{\boldsymbol{GC}},) = \left(\frac{\sum_{i=1}^{n} X_i}{n}\right), \left(\frac{\sum_{i=1}^{n} X_i}{n}\right)$$

With: $\overline{\mathbf{X}}_{\mathbf{GC}}, \overline{\mathbf{Y}}_{\mathbf{GC}}$ centre of gravity coordinates, $\mathbf{X_i}$ and $\mathbf{Y_i}$ points coordinates,
n: number of coordinates (expressed in points).

And we use the same method given above for the relative moment, i.e.:

Absolute Total Moment (ATM) of country = sum of the absolute moments of the regions

We remind that for the absolute moments of each area with itself, is equal to the absolute moment already calculated for each area (or region). These first elements of calculation constitute the diagonal of the matrix of calculations of the country's total absolute moment. And for the calculations of the absolute moment of each zone, the estimate is as follows:

$$\boldsymbol{\mu_i} = \sum_{i=1}^{n} \mathbf{Z_i} . \mathbf{d_i}(\mathbf{A}, \mathbf{B}) = \mathbf{d_i}(\mathbf{A}, \mathbf{B}) * \sum_{i=1}^{n} \mathbf{Z_i}$$

$\boldsymbol{\mu_i}$: Absolute moment calculated for each administrative area.
$\mathbf{d_i}(\mathbf{A}, \mathbf{B})$: The distance for each administrative area from all other areas of the country (expressed in Km).
$\mathbf{Z_i}$: Number of employment, wages or number of firms in a given sector.

And the Absolute Total Moment is: $\mathbf{ATM} = \boldsymbol{\mu}_{\mathbf{T_A}} = \sum_{i=1}^{n} \boldsymbol{\mu_i}$

As a result of the above discussion, the distances between administrative areas have been calculated using Google maps (see Appendix 5) and the absolute moments for each administrative region calculated for each administrative region given in Appendix 6.

Total moment of the Kingdom of Saudi Arabia is equal to the sum of the moments of each region, i.e:
Absolute Total Moment (ATM_{ASK}) = 224089287650 emplois.Km= 224 Giga E.Km

The absolute moment is a measurement that represents the mobility of workers and that characterizes the geometry of wages in our case. So, the absolute moment is the distribution of the chosen parameter work, or all other parameters, such as the number of firms or salaries paid in the chosen area. It is evident from this explanation of absolute moment that more wages are dispersed over a sparse distance, the greater there is this moment and vice versa. Otherwise, absolute moment can be interpreted as the cost of employment (moving employees to the workplace). This cost can be converted to the transport costs of a given sector by multiplying the absolute moment of the given sector or of whole sectors by the transportation price per kilometre multiplied again by two (going and return from the house to the work).

It is to be noted that the absolute moment has a maximum limit as soon as the displacement for the job affects a large part of the received wage. In the case that the absolute moments of the two countries or two sectors have the same value, this equality is justified by the equality of the following quotients $d_1/d_2 = Z_2/Z_1$ (d: distance and Z: number of jobs).

6. Conclusion

The calculation of spatial concentration has been the subject of much interest by economists for more than a century, giving rise to a variety of indicators that are the source of critique from researchers, as well as suggestions to develop these indicators for a better analysis of the economic context of agglomerations. Therefore, we are not content to use only these indicators; it was a question of proposing a method for calculating the Gini index based on a simple random selection of data from statistical system for the area and the economic activity sustained by the Spearman coefficients and the significance of the values (p-value). It's a simplified calculation method that could be used by other researchers to confirm or refute its veracity for the calculation of other indexes in the case of disaggregated data or for other indexes and not necessarily for concentration or specialization.

In addition to the results, the Gini indicator has been summed for all sectors and regions and, in the event that of this calculation method, it's possible to make an international comparison regardless the number of economic sectors and agglomerations. Nevertheless, this comparison can be effective only if the estimations are accompanied by their averages. We must recognize the difficulty for any comparison of the available data, outside the reference years, the Gini index that we estimated at 14.32% for the Kingdom of Saudi Arabia based on employment, isn't comparable to that of 35.80% for Tunisia based on consumption and to 32%, the average of OECD countries based on income inequalities.

Moreover, we have calculated other indicators adopted by the scientific community, such as the concentration coefficient for Kingdom of Saudi Arabia of Gini, Herfindahl, Ellison and Glaeser, Maurel and Sédillot and the specialization coefficient of Krugman. Their interpretations converge towards more analysis of the global distribution of economic activity between intra- and inter-sector according to the agglomeration.

Therefore, we have proposed simple moments (relative or absolute) that take into account the aggregate under consideration and the distances between them expressed in units reduced in aggregate per kilometre. This is a revealing index of spatial disparity that could be used for any economic aggregate.

Finally, an index must be easily usable considering the available data and must be comparable according to the sector, allow measuring significantly the differences between areas, periods or sectors, taking into account the geographical agglomeration of the activities and being insensitive to the sector classification or the geographical division.

Acknowledgments:

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through research groups' program under grant number GRP 299-40.

References:

Amrhein, C.G., & Reynolds, H. (1996): Using Spatial Statistics to Assess Aggregation Effects, *Geographical Systems*. 3: 83–101.

Amrhein, C.G., & Reynolds, H. (1997): Using the Get is Statistic to Explore Aggregation Effects in Metropolitan Toronto Census Data, *The Canadian Geographer*. 41(2): 137–149.

Arbia, G., Espa, G., Giuliani, D., Mazzitelli, A. (2012): Clusters of firms in an inhomogeneous space: The high-tech industries in Milan. *Economic Modelling*. 29(1): 3–11.

Arrow, K.J. (1962): The Economic Implications of Learning by Doing. *The Review of Economic Studies*. 29 (3): 155-173.

Atkinson, Anthony.B. (1970). On the measurement of inequality. *Journal of Economic Theory*. 2(3): 244-263.

Baudelle, G., Regnauld, H. (2004): *Echelles et temporalités en géographie*. Paris. Sedes.

Batty, M., Sammons, R. (1978). On searching for the most informative spatial pattern. *Environment and Planning*. A, 10: 747-780.

Bhattacharya, N., & Mahalanobis, B. (1967): Regional Disparities in Household Consumption in India. *Journal of the American Statistical Association*. 62: 143-161.

Blalock, H.M. (1964): *Causal Inferences in Nonexperimental Research*. North Carolina. Chapel Hill.

Brown, M.C. (1994): Using Gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on Alberta data. *Social Science and Medicine*. 38(9): 1243-1256.

Cliff, A.D., Haggett, P., Ord, J.K., Bassett, K. & Davies, R. (1975): *Elements of Spatial Structure*. Cambridge, Cambridge Univ. Press.

Cliff, A.D., & Ord, J.K. (1973): *Spatial Autocorrelation*. London. Pion.

Cole, R.G., & Syms, C. (1999): Using spatial pattern analysis to distinguish causes of mortality: an example from kelp in north-eastern New Zealand. *Journal of Ecology*. 87(6): 963–972.

Cowell, F. and Mehta, F. (1982). The estimation and interpolation of inequality measures. *The Review of Economic Studies*. 49: 273-290.

Ciccone, A, & Hall, R.E. (1996): Productivity and the Density of Economic Activity, *American Economic Review*. 86(1): 54–70.

Cramer, J.S. (1964): Efficient grouping regression and correlation in Engel curve analysis. *Journal of the American Statistical Association*. 59: 233-250.

Combes, P.P. (2000): Economic Structure and Local growth: France, 1984-1993. *Journal of Urban Economics*. 47: 329-355.

Cowell, F.A. (1988): Inequality decomposition: three bad measures. *Bulletin of Economic Research*. 40(4): 309-312.

Dagum, C. (1987): Measuring the economic affluence between populations of income receivers. Journal of Business and Economics Statistics. Vol. 5(1): 5–12.

Dall'erba, S. (2004): Les politiques de développement régional en Europe à la lumière des outils récents de la science régionale. Thèse de doctorat sciences économiques, Pau. France.

deHaan, L.F.M. (1970): *On regular variation and its application to the weak convergence of sample extremes*. Amsterdam. Mathematics Centrum.

Devereux, M.P., Griffith, R., & Simpson, H. (2003): The geographic distribution of production activity in the UK. *Regional Science and Urban Economics*. 34: 533–564.

Duncan, O.D. (1957): The measurement of population distribution. *Population Studies*. XI (1). pp. 27-45.

Duranton, G. and Overman, H.G. (2002): Testing for localization using micro-geographic data. Discussion Paper 3379. CEPR.

Ellison, G. and Glaeser, E. (1994): Geographic concentration in U.S. Manufacturing industries: a dartboard approach. NBER Working Paper 4840.

Ellison, G. and Glaeser, E.L. (1997): Geographic concentration in U.S. Manufacturing Industries: A Dartboard Approach. *Journal of Political Economy*. 105(5): 889-927.

Floch, J.M. (2012): Détection des disparités socio-économiques - L'apport de la statistique spatiale. Document de travail de la DDAR. INSEE H2012/04.

Fotheringham, A.S., & Wong, D.W.S. (1991): The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning*. 23: 1025–1044.

Fisher, R.A., & Tippett, L.H.C. (1928): Limiting forms of the frequency distribution of the largest or smallest member of a sample. In: Mathematical Proceedings of the Cambridge Philosophical Society. Vol. 24. Cambridge University Press. pp. 180–190.

Gehlke, C., & Biehl, K. (1934): Certain Effects on Grouping upon the Size of the Correlation Coefficient in Census Tract Material, *Journal of the American Statistical Association*. 29(185): 169–170.

Gibrat, R. (1931): *Les inégalités économiques*. Librairie du Recueil Sirey de France. Paris.

Gini, C. (1912): Variabilità e mutabilità. Memori di Metodologia Statistica, Vol. 1, Variabilità e Concentrazione. Libreria Eredi Virgilio Veschi. Rome. pp 211-382.

Gini, C. (1921): Measurement of inequality of incomes. *The Economic Journal*. pp. 124-126.

Gnedenko, B., & Kolmogorov, A. (1954): *Independent random variables*. Cambridge, Massachusetts: Addison-Wesley.

Guillain, R., Le Gallo, J. (2007): Agglomeration and dispersion of economic activities in Paris and its surroundings: An exploratory spatial data analysis. Regional Economics Applications Laboratory Discussion. Paper 06-T-10.

Gumbel, E.J. (1958): *Statistics of extremes*, Columbia University Press. New York.

Haggett, P. (1965): *Localization analysis in human geography*. London. Edward Arnold Ltd.

Haggett, P. (1973): *L'analyse spatiale en géographie humaine*. Paris. Armand Colin.

Hall, M. and N. Tideman (1967). Measures of Concentration. *Journal of the American Statistical Association*. 62: 162-168.

Hannan, M.T. (1971). *Aggregation and Disaggregation*. Lexington. Mass. D.C. Heath.

Hannah, L. and Kay, J.A. (1977). *Concentration in Modern industry*. Macmillan. London.

Hart, P.E., & Prais, S. J. (1956): The analysis of business concentration. J. r. STATIST. Soc. Series A, part 2. pp. 150-191.

Hart, P.E. (1971): Entropy and Other Measures of Concentration. *Journal of the Royal Statistical Society*. 134 (1): 73-85.

Herfindahl, O.C. (1950): Concentration in the U.S. Steel Industry. Unpublished doctoral dissertation, Columbia University.

Hirschman, A.O. (1945): *National power and the structure of foreign trade*. Berkeley.

Hoover, E.M. Jr. (1936): The Measurement of Industrial Localization. *Review of Economics and Statistics*. 18: 162-171.

Hoover, E.M. Jr. (1984): *An Introduction to Regional Economics*. ISBN 0075544407.

Jacobs, J. (1969): *The economy of Cities*. Vintage. New York.

Kendall, M.G., & Yule, G.U. (1950): *An Introduction to the Theory of Statistics*. London. Griffin.

Krugman, P. (1991a): *Geography and Trade*. Leuven: Leuven University Press and Cambridge. MA: The MIT Press.

Krugman, P. (1991b): Increasing returns and economic geography. *Journal of Political Economy*. 99: 483-499.

Kullback, S, & Liebler, R.A. (1951): *Information Theory and Statistics*, New York: Wiley.

Lacoste, Y. (1976): *La géographie ça sert d'abord à faire la guerre*. Paris. La Découverte.

Lagache. T, Lang. G, Sauvonnet. N., & Jean-Christophe O-M. (2013): Analysis of the Spatial Organization of Molecules with Robust Statistics. *Plos One* 8.12, e80914.

Launhardt, W. (1885): *Mathematische Begründung der Volkswirthschaftslehre*. W. Engelmann.

Lorenz, M.O. (1905). Methods for measuring concentration of wealth, *Journal of the American Statistical Association*, 9(70): 209-219.

Maby, J. (2002): Approche conceptuelle et pratique des indicateurs en géographie. Objets et indicateurs géographiques. UMR Espace. Actes Avignon. France. pp. 16-41.

Marceau, D.J. (1999): The scale issue in social and natural sciences. *Canadian Journal of Remote Sens*. 25(4): 347-356.

Martin, R.L., Thrift, N.J., Bennett, R. (1979): *Towards the Dynamic Analysis of Spatial Systems*. London, Pion.

Mandelbrot, B. (1977): *The fractal geometry of nature*. New York. W. H. Freeman and Compagny.

Marcon, E., & Puech, F. (2003): Evaluating the Geographic Concentration of Industries Using Distance-Based Methods. *Journal of Economic Geography*. 3(4): 409-428.

Marcon, E., & Puech, F. (2009): Generalizing Ripley's K function to inhomogeneous populations. HAL halshs-00372631.

Marshall, A. (1890): *Principles of Economics*. Macmillan. London.

Maurel, F., & Sédillot, B. (1999): A measure of the geographic concentration in French manufacturing industries. *Regional Science and Urban Economics*. 29: 575-604.

Møller. J., & Hakon, T. (2014): Geometric Anisotropic Spatial Point Pattern Analysis and Cox Processes. *Scandinavian Journal of Statistics. Monographs on Statistics and Applies Probabilities*. 41(2): 414–435.

Mori, T., Nishikimi, K., & Smith, T.E. (2005): A Divergence Statistic for Industrial Localization. *The Review of Economics and Statistics*. 87(4): 635–651.

Morrisson, C. (1996): La répartition des revenus, Presses universitaires de France. Collection Thémis. Économie, 1ère édition, Paris.

Nottalen, L. (1993): *Fractal space-time and microphysics. Toward a theory of scale relativity*. Singapour. World scientific. XIV.

Openshaw, S. (1984): *The modifiable areal unit problem*. Norwich: Geo Books, CATMOG 38.

Pickands III, J. *et al.* (1975): Statistical inference using extreme order statistics. *The Annals of Statistics*. 3(1): 119–131.

Piron, M. (1992): Analyse statistique d'un système d'échelles, Bondy-ORSTOM. Document de travail. Réseau ADOC, 4.

Piron, M. (1993): Changer d'échelle: une méthode pour l'analyse des systèmes d'échelles. *L'espace géographique.* 2: 147-165.

Porter, M.E. (1990): *The Competitive Advantage of Nations*. Free Press. New York.

Pyatt, G. (1976): On the interpretation and disaggregation of Gini coefficients. *Economic Journal*. 86: 243-255.

Rao, V.M. (1969). Two decompositions of concentration ratio. *Journal of the Royal Statistical Society*. Series A 132: 418-425.

Rozenblat, C. (1989): Différenciations spatiales et sociales dans un espace urbain : les quartiers de Paris. *L'Espace géographique*. 3: 264-272.

Rysman, M, & Greenstein, S. (2005): Testing for agglomeration and dispersion. *Economics Letters*. 86(3): 405–411.

Sanders, L. (1989): *L'analyse des données appliquées à la géographie*. Montpellier. GIP-Reclus.

Sammons, R. (1976): Zoning systems for spatial models. *Reading Geographical Papers*, 52. Reading (England), Univ. of Reading, Dep. of Geogr.

Sawyer, M. (1976). La répartition des revenus dans les pays de l'O.C.D.E. Perspectives Économiques de l'O.C.D.E. Études spéciales. pp. 3-33.

Shorrocks. A.F. (1980). The class of additively decomposable inequality measures. *Econometrica*. 48 (3): 613-626.

Silber J. (1989): Factor components, population subgroups and the computation of the Gini index of inequality. *Review of Economics and Statistics*. 71: 107-115.

Simpson, E. H. (1949): Measurement of diversity. *Nature*. 163(4148): 688.

Szwagrzyk, J., & Czerwczak, M. (1993): Spatial patterns of trees in natural forests of East-Central Europe. *Journal of Vegetation Science* 4(4): 469–476.

Theil, H. (1967). *Economics and Information Theory*, Amsterdam.

UNESCO Institute for Statistics (2008): *Towards information literacy indicators*. UNESCO. Paris.

Volvey, A. (2005): *Echelles et temporalités*. Paris, Atlande.

Von Thünen, J.H. (1826): *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Hamburg, Perthes. Translation: (1966). The Isolated State. Oxford. Pergammon Press.

Weber, A. (1909): *Ueber den standort der industrien*, volume 2. J.C.B. Mohr, Tübingen.

Weber, A. (1929): *Theory of the location of industries*. University Press. Chicago.

Williams, I.N. (1976): Optimistic theory validation from spatially grouped regression: theoretical aspects. Transactions of the Martin Centre. 1:113-145.

Appendix 1

Number of Employees by administrative area and main groups of economic activities (2018) in in the Kingdom of Saudi Arabia

Administrative Area	Riyadh	Makkah	Madinah	Qassim	Easte. Prov	Asir	Tabuk	Hail	North Bord	Jazan	Najran	AL - Baha	AL - Jouf
Other service activities	7	2	0	0	1500	0	0	0	0	0	0	0	0
Agriculture, forestry & fishing	33558	14833	2543	11664	10725	1571	3643	5759	126	2802	801	140	2366
Electricity, gas & Water	45158	23664	802	253	18919	6446	166	395	123	290	135	89	64
Mining and quarrying	8512	12994	2490	555	149555	1667	118	394	892	589	467	204	66
Transportation and communication	154222	81426	9975	11113	68928	13125	3639	3233	2453	3485	6389	913	3447
Other collective & social services	273192	158002	28036	18152	109090	21466	11986	9712	3184	11029	6532	2734	5600
Manufacturing	289873	216198	36472	28314	211623	27910	10341	11090	3952	13983	6352	3637	5884
Financial & insurance activities	511885	176281	8114	7351	128982	25165	2618	1924	2360	7214	8089	2394	2678
Trade	830977	608136	101193	70806	357445	65947	25309	22084	8542	46526	22975	20884	14543
Construction	1 176111	663649	115889	171359	774863	123655	32807	54496	17690	34282	64937	10610	22924
Total	**3323495**	**1 955185**	**305514**	**319567**	**1831630**	**286952**	**90627**	**109087**	**39322**	**120200**	**116677**	**41605**	**57572**

Appendix 2

Gini indexes by administrative area of Kingdom Saudi Arabia

area Administrative	$GINI_1$	$GINI_2$	$GINI_3$	$GINI_4$	$GINI_5$	$GINI_6$	$GINI_7$
North. Bord	0.3760	0.5119	0.4095	0.3810	P.M	0.3463	0.5065
AL - Baha	0.4865	0.5808	0.4855	0.4865	0.6317	0.4675	0.5726
AL - Jouf	0.3046	0.3970	0.2565	0.3072	0.1857	P.M	0.3981
Tabuk	0.3278	0.3896	0.2390	0.3439	0.2161	0.1942	0.4056
Hail	0.3713	P.M	0.3098	0.3815	0.0502	0.2451	0.4204
Najran	0.4850	0.6157	0.5117	0.4909	0.3795	0.4933	0.6125
Jazan	0.3716	0.4673	P.M	0.3712	0.4362	0.3080	0.4600
Asir	0.3224	0.5023	0.3822	0.4115	0.3247	0.3423	P.M
Madinah	0.4737	0.5388	0.4218	0.4731	0.4488	0.3905	0.5400
Qassim	0.4695	0.5229	0.4275	0.4589	0.2131	0.3859	0.5147
Makkah	0.3361	0.4968	0.3716	0.3758	0.4232	0.3368	0.4808
Riyadh	P.M	0.4821	0.3481	P.M	0.3893	0.3197	0.4680
Easte. Prov.	0.2098	0.3888	0.4184	0.2251	0.1312	0.2003	0.3617

- **NB:** $\mathbf{GINI_1}$: estimation of GINI indexes by zone with the exclusion of the Riyadh administrative zone and other activity services related to this zone for n=1 and so on. ….. Up until $\mathbf{GINI_7}$.
- **PM:** For memory.

Appendix 3

Spearman's correlations of the Gini indices

	GINI 1	GINI 2	GINI 3	GINI 4	GINI 5	GINI 6	GINI 7
GINI 1		0.5343	0.4872	**0.9518**	0.2500	0.5704	**0.7709**
GINI 2			0.3620	0.6098	0.4469	**0.9852**	0.4934
GINI 3				0.5387	0.3141	**0.3493**	**0.7694**
GINI 4					0.3349	0.6462	0.8772
GINI 5						0.3409	0.4567
GINI 6							0.5055
GINI 7							

Appendix 4

P-value

Indices	GINI 1	GINI 2	GINI 3	GINI 4	GINI 5	GINI 6	GINI 7
GINI 1		0.07356	0.10820	**0.00189**	0.43320	0.05279	**0.00334**
GINI 2			0.24750	**0.03525**	0.14520	**0.00001**	0.10310
GINI 3				0.07073	0.32000	0.26570	**0.00343**
GINI 4					0.28720	**0.02319**	**0.00018**
GINI 5						0.27830	0.13560
GINI 6							0.09365
GINI 7							

Appendix 5

Distances between the administrative areas of the Kingdom of Saudi Arabia

	North. Bord	AL - Baha	AL - Jouf	Tabuk	Hail	Najran	Jazan	Asir	Madinah	Qassim	Makkah	Riyadh	Easte. Prov
North.Bord.		1682	664	1039	569	2149	2067	1724	1009	637	1438	905	1232
AL - Baha	1682		1610	1341	1182	623	502	316	707	993	310	954	1336
AL - Jouf	664	1610		421	424	286	81	1745	818	633	1305	985	1383
Tabuk	1039	1341	421		642	1903	1678	1675	682	946	1030	1299	1696
Hail	569	1182	424	642		1560	1544	1317	13	308	893	660	1058
Najran	2149	623	286	1903	1560		331	258	1243	1518	964	1262	1556
Jazan	2067	502	81	1678	1544	331		201	1120	1436	738	1180	1473
Asir	1724	316	1475	1675	1317	258	201		692	1093	613	837	1219
Madinah	1009	707	818	682	13	1243	1120	692		501	437	833	1230
Qassim	637	993	633	946	308	1518	1436	1093	501		930	354	751
Makkah	14381	310	1305	1030	893	964	738	613	437	930		961	1343
Riyadh	905	954	985	1299	660	1262	1180	837	833	354	961		400
Easte. Prov.	1232	1336	1383	1696	1058	1556	1473	1219	1230	751	1343	400	

Appendix 6

Absolute moment of each administrative area

Administrative Area	North bord	AL - Baha	AL - Jouf	Tabuk	Hail	Najran	Jazan	Asir	Madinah	Qassim	Makkah	Riyadh	Easte. Prov
North bord	**3887147**	69979610	38227808	94161453	62070503	250738873	248453400	494705248	308263626	203564179	28117515485	3007762975	2256568160
AL - Baha	69979610	369822	92690920	121530807	128940834	72689771	60340400	90676832	215998398	317330031	606107350	3170614230	2447057680
AL - Jouf	38227808	92690920	**3499401**	38153967	46252888	33369622	9736200	423254200	249910452	202285911	2551516425	3273642575	2533144290
Tabuk	94161453	121530807	38153967	**4486309**	70033854	222036331	201695600	480644600	208360548	302310382	2013840550	4317220005	3106444480
Hail	62070503	128940834	46252888	70033854	**2445469**	182016120	185588800	377915784	3971682	98426636	1745980205	2193506700	1937864540
Najran	250738873	72689771	33369622	222036331	182016120	**3391244**	39786200	74033616	379753902	485102706	1884798340	4194250690	2850016280
Jazan	248453400	60340400	9736200	201695600	185588800	39786200	**1245619**	57677352	342175680	458898212	1442926530	3921724100	2697990990
Asir	494705248	90676832	423254200	480644600	377915784	74033616	57677352	**4192008**	211415688	349286731	1198528405	2781765315	2232756970
Madinah	308263626	215998398	249910452	208360548	3971682	379753902	342175680	211415688	**8354722**	160103067	854415845	2768471335	2252904900
Qassim	203564179	317330031	202285911	302310382	98426636	485102706	458898212	349286731	160103067	**12529521**	1818322050	1176517230	1375554130
Makkah	28117515485	606107350	2551516425	2013840550	1745980205	1884798340	1442926530	1198528405	854415845	1818322050	**29609550**	3193878695	2459879090
Riyadh	3007762975	3170614230	3273642575	4317220005	2193506700	4194250690	3921724100	2781765315	2768471335	1176517230	3193878695	**137866940**	732652000
Easte. Prov	2256568160	2447057680	2533144290	3106444480	1937864540	2850016280	2697990990	2232756970	2252904900	1375554130	2459879090	732652000	**116013960**
Total	**35155898467**	**7394326685**	**9495684659**	**11180918886**	**7035014015**	**10671983695**	**9668239083**	**8776852749**	**7964099845**	**6960230786**	**47917318520**	**34869872790**	**26998847470**