Testing for localization with entropy-based measures

Roy Cerqueti

Sapienza University of Rome, Department of Social and Economic Sciences

Piazzale Aldo Moro, 5 - I-00185 - Rome, Italy.

Email: [roy.cerqueti@uniroma1.it](mailto:roy.cerqueti@uniroma1.it)

and

London South Bank University, School of Business

103 Borough Rd - SE1 0AA - London, UK

Eleonora Cutrini

University of Macerata, Department of Economics and Law

Via Crescimbeni, 14 I-62100, Macerata, Italy

Email: [eleonora.cutrini@unimc.it](mailto:eleonora.cutrini@unimc.it)

## Abstract

## This paper aims to give statistical significance to the measurement of spatial concentration in the context of entropy-based approaches. We simulate confidence intervals based on a null hypothesis able to capture systematic spatial concentration of firms from random patterns, and dissimilarities between the distributions of firms and employees. We implement this two-step methodology to the European manufacturing economy, and we find a substantive spatial clustering of establishments whereby the spatial divergence between employees and firms is significant both for small-scale industries typically considered as localized because of industry-specific Marshallian external economies and for those industries characterized by considerable internal scale economies. We suggest that a high heterogeneity in firm size may have positive implications for aggregate competitiveness at the sectoral level.

JEL codes: C43, C46, C12, L60, R12

Keywords: Statistical testing, Spatial Concentration, Entropy measures, Confidence interval

## Abstract

## This paper aims to give statistical significance to the measurement of spatial concentration in the context of entropy-based approaches. We simulate confidence intervals based on a null hypothesis able to capture systematic spatial concentration of firms from random patterns, and dissimilarities between the distributions of firms and employees. We implement this two-step methodology to the European manufacturing economy, and we find a substantive spatial clustering of establishments whereby the spatial divergence between employees and firms is significant both for small-scale industries typically considered as localized because of industry-specific Marshallian external economies and for those industries characterized by considerable internal scale economies. We suggest that a high regional heterogeneity in firm size may have positive implications for aggregate competitiveness at the sectoral level.

JEL codes: C43, C46, C12, L60, R12

Keywords: Statistical testing, Spatial Concentration, Entropy measures, Confidence interval

## Introduction Over the last few years, multidimensional concepts have been one of the most treated research topics. In particular, the research focused on the methodologies to assess complex phenomena such as wellbeing and competitiveness and the extent to which they can be quantified (See, among others, Mazziotta and Pareto, 2019; Krishnakumar and Nagar, 2008; Schlossarek *et al.,* 2019; Ciommi *et al.,* 2019; Greco *et al.,* 2019). Due to the recurrent economic crises from 2000s onwards, this discussion on the associated indicators and methodological advancements are fundamental to support the performance of the EU economic system, in a renewed interest in the analysis of the consequences of industrial policy (Aiginger and Rodrik, 2020), and particularly, after the “European Industrial Reinaissance” heralded by the European Commission in 2014 to bolster the competitiveness of EU industries through a combination of regional and industrial policy tools (European Commission, 2014). In this context, indicators and methodologies to evaluated competitiveness are fundamental to monitor the progresses towards the objectives of Europe 2020, and for a better definition and implementation of Smart Specialization Strategies at the regional level. Moreover, thanks to the growing availability and higher granularity at which data are available it is now feasible to conduct more robust evidence-based policies and evaluate competitiveness at the different spatial scales. In fact, one of the main issues for Europe in its path towards a “data economy” (EC, 2020) is to evaluate competitiveness, with a EU-wide perspective but also with a detailed analysis of the regional economic structures and their components as well as the spatial concentration of industries.

## In this respect, it is worth stressing that the two phenomena of regional competitiveness and agglomeration of industries are indeed strictly related. Interestingly, they found a statistical correspondence when global coefficients of localization are derived based on specific measures of concentration and specialization. In particular, specialization and concentration cannot diverge if relative measures are used (Aiginger and Davies, 2004; Mulligan and Schmidt, 2005; Cutrini, 2009). These first intuitions were further developed in the context of a Stochastic Independence Approach (Haedo and Mouchart, 2018).

## This paper deals with entropy-based methodologies for measuring the geographical concentration of industries. We first discuss their main advantages and limitations, then we introduce a novel method to test whether the spatial pattern of employment is significantly different from that of firms. We address a specific gap in the literature. In fact, although entropy-based measures have attracted considerable interest among academics and practitioners, to date, to the best of our knowledge, no study have still established the possible existence of statistical criteria that might enable us to test for the presence of significant relative concentration of both firms and employees. Without this kind of statistical tests, any concentration index has little value because, as suggested by Combes et al., (2008) *“we are unable to determine whether we are dealing with high or low concentration, or whether there is even any spatial concentration at all”* (p. 259).

To overcome these limitations, this article introduces a statistical testing procedure suitable for entropy measures applied to discrete space analysis. The testing procedure is developed in two steps. In the first step, we evaluate whether the observed spatial distribution of establishments is significantly different from randomness, while in the second step we value whether the spatial distribution of employees is significantly different from the observed spatial distribution of firms.

The methodological advancement proposed here allows improving the interpretation of the indices considering the underlying factors behind the concentration of each sector. In the first step, it is possible to infer whether location patterns of one sector is dominated by the strength of intra-industry externalities due to co-location of small firms or instead it is mostly driven by increasing returns at the plant level. In the second step of the testing procedure, the divergence between the regional distribution of employees and the regional distribution of establishments is considered.

The results of the performed tests suggest that the spatial concentration of manufacturing establishments was still a stylized fact in the European economic geography in 2007. The distribution of the employees however does not follow the spatial concentration of firms. We suggest that this divergence is mostly attributable to the heterogeneity in the average regional firms’ size.

The paper is organized as follows: Section 1 reviews the related literature. Section 2 introduces the analytical framework defining the entropy measures adopted. Section 3 presents the testing procedure. Section 4 includes an application based on regional European data. Section 5 offers some concluding remarks.

## 1. Testing for localization: some insights from the related literature

In the so-called New Trade Theory and New Economic Geography, trade and comparative advantage are driven not by relative cost (price) differences, but by increasing returns effects associated with economies of scale, monopolistic competition, and geographical agglomeration externalities.

On this background, the development of New Economic Geography theoretical models was inevitably accompanied by a rising interest in measuring localization of economic activities. Hence, the spatial distribution of economic activity has been the focus of growing research interest at the empirical level and several statistical tools were introduced to systematically evaluate the spatial agglomeration of industries.

In this context, we may distinguish two main lines of methodological advancements: one is related to the measures for empirical research over the discrete space - Ellison and Glaeser (1997)’s approach and the entropy-based methods; the other deals with the measurement of the geographical concentration over the continuous space (Marcon and Puech, 2003, 2010; Duranton and Overman, 2005; Lang *et al.*, 2020 among others).

Ellison and Glaeser (1997) started to think about systematic ways to evaluate agglomerations. They have the merit of having introduced a measure of spatial distribution of employees, which controls for the industrial structure. In the “dartboard approach” the measurement of relative concentration is rescaled by an index of absolute concentration of establishments (the Herfindal index), and the index is anchored to a theoretical model of location. Hence, it is possible to derive a test for the statistical significance of the index with a null hypothesis defined as perfect regularity, and indicate whether a sector’s distribution of activity across locations is significantly concentrated or dispersed beyond scale economies. Successive empirical studies (Maurel and Sédillot, 1999, Devereux *et al.*, 2004) have adopted the “dartboard approach” to investigate the spatial distribution of manufacturing industries in France and United Kingdom, respectively[[1]](#footnote-1). Recently, Cassey and Smith, 2014 show how the *ad hoc* thresholds suggested by Ellison and Glaser (1997) for a significant concentration can flaw the interpretation of results based on the Ellison and Glaeser’s index and they simulate confidence intervals that can be used for statistical testing.

Mori *et al.* (2005) have established an entropy index which is asymptotically normally-distributed and derived the confidence interval for the true level of localization. They used Monte Carlo simulations to check their test with small samples. Their null hypothesis is different from the one adopted in this paper. More specifically, their counterfactual distribution follows the productive area of regions and corresponds to the reference distribution included in the computation of the index. They formulated a probability model of complete spatial dispersion and then consider the deviations of each distribution from this benchmark model. They postulated that a completely dispersed distribution for an industry is one in which randomly sampled establishments are equally likely to be located anywhere within a given economic area. The authors also recognized the possibility to adopt alternative choices for the reference distribution, such as employees’ levels for all industries.

The identification of a null hypothesis and the procedure for testing significant localization is also a main feature of the distance-based approach, even since the original contributions on the evaluation of the spatial distribution of firms in France (Marcon and Puech, 2003) and United Kingdom (Duranton and Overman, 2005). In such contributions, the significant spatial concentration of an industry was denoted as being a departure from a theoretical random distribution of the same industry. Specifically, Marcon and Puech (2003) constructed the counterfactual with a no-localization scenario, being referred to a random distribution of establishments from the same industry reshuffled in all the sites occupied by the industry; Duranton and Overman (2005) projected the distribution of establishments in the industry in the wider area of all the possible locations of the manufacturing industry (the aggregate economic activity)[[2]](#footnote-2). Following this procedure, manufacturing industries are detected as dispersed (at some distance) if their degree of concentration at that distance is lower than the random generated confidence interval. In this manner, spatial point patterns do not control for the distribution of the economic activity as a whole. This is the main reason why Lang *et al.* (2020) have recently introduced a relative density function.

This feature – to control for the spatial distribution of overall manufacturing or another suitable benchmark - is instead typical in the construction of relative measures. Relative concentration indices - like the raw G-index of Ellison and Glaeser and relative entropies- usually measure the discrepancy between the spatial distribution of one industry and that of the aggregate activity selected as a benchmark (e.g. total manufacturing) (Brülhart and Traeger, 2005, Bickenbach *et al.*, 2008; Cutrini, 2009, 2010). However, there are further options for the reference distribution. For example, Mori *et al.* (2005) measured topographic concentration for Japan using a dissimilarity entropy measure -a discrete Kullback-Leibler divergence- to compare the distribution of establishments over the distribution of economic land area.

Relative entropies are becoming frequently used in the empirical assessment of specialization and concentration (e.g. Bagoulla and Péridy, 2011; Vechiu and Makhlouf, 2014; Stierle-von Schultz and Stierle, 2013, Gokan, 2010, Evans, 2010 among many others), mostly because they meet desirable separability properties across geographical scales and sectors (Mori et al., 2005; Brülhart and Traeger, 2005; Bickenbach and Bode, 2008, Cutrini, 2009, 2010). These decomposition properties are particularly appealing in the case of the Theil index, since the weight used in the within-entropy does not depend on the between-entropy (Bourguignon, 1979). Hence, they allow decomposing inequality across different spatial and sectoral scale in order to identify the contributions of individual regions (sectors) to the overall localization of economic activity. Moreover, testing procedure based on bootstrap methods have been proposed to evaluate significant change over time (Brülhart and Traeger, 2005). On the methodological terrain, relative entropy measures have been described on the basis of the axiomatic principles implicitly assumed by regional economists when using them to quantify the spatial concentration of a sector (continuity, symmetry, weak scale invariance, location division property, group division property, type I and type II independence) (Alonso-Villar and Del Río, 2013). Further studies investigated the discrepancy between absolute and relative entropy measures (Bickenbach et al., 2010) and their characterization in the context of a stochastic independence approach (Haedo and Mouchart, 2018).

To summarize, our contribution is in line with research on the simulation of confidence intervals for measures constructed over the discrete space such as the Ellison and Glaeser index to test the null hypothesis of no concentration (Cassey and Smith, 2014) and the recent advancements in the context of distance-based methods (Duranton and Overman, 2005; Marcon and Puech, 2003, 2010, Lang *et al.*, 2020) in that they also seek to detect significant localization patterns, although over the continuous space.

## 2. The framework: some essential premises and clarifications

In this section we define and characterized the absolute and relative entropy measures we use according to a taxonomy that is well-established in the reference literature (e.g Brülhart and Traeger, 2005; Bickenbach and Bode, 2008). We use the Theil index –the index belonging to the Generalised Entropy class GE(a) with parmeter a=1, GE(1) – in two versions.

In a first step, we focus on the spatial concentration of firms, and we use an absolute entropy measure.

The normalized Shannon index – or absolute entropy - is a deviation of the observed entropy from the maximum entropy associate to the uniform distribution (). This index is defined as:

(1)

where is the share in region of the distribution of establishments in sector , being the number of establishments in belonging to sector and the number of establishments in sector .

In this case the reference counterfactual distribution is the uniform distribution of establishments across regions. We therefore evaluate the distance between the geographical spreading of establishments in a specific industry and the theoretical uniform distribution across spatial units. This divergence signals the presence of localization economies within an industry due to the clustering of firms within specific locations.

In a second step, we use employees data and we measure the divergence between the regional distribution of employees in each industry *s* and the spreading of establishments in the same industry *s*.

Here, we refer to the idea that a high divergence between the regional distribution of employees and that of establishment signal the importance of high regional diversity in firm size within the industry, with positive implications for aggregate competitiveness at the sectoral level. In fact, among the possible measures of competitiveness one group is related to firms’ dynamics (WIFO, 2017). For example, Bartelsman et al. (2009) suggest that average firm size relative to entry by age is a useful indicator of the gap in size between entrants and incumbents. In particular, a smaller relative size of entrants tells about greater experimentation and thus higher competitiveness of a sector. Moreover, they mention the *share of gazelles*, which are firms that rapidly expand the number of their employees. And a higher share of gazelles within an industry is considered as an indication that the most innovative and productive companies are easily conquering market shares which is good for competitiveness.

The relative Theil index – or relative entropy- is defined as:

(2)

where is the share in region of the distribution of employees in sector , being the number of employees in belonging to sector and the number of employees in sector .

In this case the reference counterfactual distribution is the actual distribution of establishments.

When the relative Theil index is zero it means that the two distributions of cases and controls –to use the terminology of Diggle (1983) and Arbia *et al.* (2012) - are perfectly overlapping and hence the employees in the specific sector mimics the spatial distribution of establishments. In this case, there is no heterogeneity in firm size, each single firm operating within the industry is of the same size in terms of employees.

We assume that, under the null hypothesis, the location of employees follows the location of establishments.

In the next section we enter the details of the testing procedure.

## 3. A testing procedure for relative entropy measures

## The concept of competitiveness can be referred to different economic unit of analysis: firms (micro level), - sectors (meso level), - economy-wide (macro level) (WIFO, 2017) with several intersection with one another. We have highlighted that regional/country competitiveness found a counterpart in the geographical localization of industries when relative measures are adopted in the analysis. Such geographical localization can be the outcome of external economies and the associated increasing returns effects to the firms in the industry in question. Moreover, we suggested that firm heterogeneity across space within each industry is an additional useful information which may signal competitiveness at the firm and regional level.

## In the dartboard approach, the methodology to assess the magnitude of localization involves a comparison between the distribution of employees and the distribution of firms and it is anchored to a theoretical model of location. In similar approaches related to the discrete space that make use of lattice employees data, such as those based on relative entropy measures, the underlying location choice of firms can be viewed as derived from a probabilistic model of location choice, where each firm or plant sequentially chooses its location to maximize its profits, as in the Ellison and Glaeser’s approach.

We have also highlighted that the identification of a significant departure from a random location scenario for the distribution of firms across a homogeneous space - a “no localization” state, where even “first nature” location advantages are irrelevant - can be associated to a large entropic distance from the uniform distribution. Furthermore, in a frictionless world with perfect and instantaneous mobility of labor across regions, and equal-sized firms, the spatial distribution of employees should be tailored on the spatial distribution of the existing firms. In this respect, employees locate in space on the basis of the relative number of firms belonging to the available regions.

## However, the no-localization state seems to be not recurrent in the reality and the optimal firm size at the industry level is just a theoretical feature. Hence it is important to identify whether the actual spatial distribution of industries are the expressions of significant deviations from the counterfactual scenario.

## Consistently with the theoretical assumptions explained above, we test this hypothesis by developing confidence intervals for absolute and relative entropies defined in equations (1) and (2) and finding threshold levels able to distinguish systematic localization patterns purposely different from the ideal counterfactual scenario. The main difference between relative and absolute entropy lies in the form of the counterfactual distribution. The latter type of entropy is used when one has to compare a given distribution with the uniform one, while the former does not require the uniform distribution and can be used in a general case.

Testing for localization for absolute and relative entropy measures naturally involve testing the following null:

The rejection region is identified by the alternative hypothesis:

where represents one of the entropy statistics measuring either the spatial concentration of establishments (), defined in equation (1) or the dissimilarity between employees and establishments for each sector *s* (), defined in equation (2). The threshold represents the critical value that define the rejection regions. The dependence of and from is conveniently removed from the notation.

We notice that the shape of null and alternative hypothesis is the same in both of cases of firms and employees. We do not use two different notations for the sake of simplicity.

The counterfactual distribution associated to the null hypothesis of the test is constructed under a location scenario, which refers to the theoretical spatial distributions of firms and workers that would have arisen in the absence of various agglomeration forces. We elaborate on this assumption.

In a neutral world where regions are identical and decisions are taken without involving any specific preference, then location of firms is a purely random process. Instead, in the real world, location processes are related to factors such as demand distribution in space, transport costs and agglomerative forces such as increasing returns to scale (internal or external to the firm), as in the well-known synthesis of the New Economic Geography.

We postulate that firms create jobs and employees immediately follow. Hence, in the counterfactual distribution, the employees of each sector are distributed across regions as the respective distribution of establishments. Instead, under the alternative hypotheses the distribution of employees departs from the distribution of firms.

The null hypotheses are then constructed on the basis of two basic assumptions:

(a) uniform distribution of firms across regions, for the absolute entropy related to firms;

(b) employees assigned to regions proportionally to firms, for the relative entropy related to employees.

Accordingly, our testing procedure is based on two steps. The counterfactual reference distribution is the uniform distribution in the first step, and the distribution of firms, in the second step.

In the first step, we construct the counterfactual distribution of establishments, considering that, under the null hypothesis, firms’ location follows a uniform law with upper and lower bounds based on the actual distribution of firms across regions.

Assuming pure randomness means that the selection process of a region by a firm is not driven by natural advantages, internal scale economies or external Marshallian economies.

We run 10,000 Monte Carlo simulations for each industry establishments’ distribution, separately. The construction of counterfactual distributions for each industry is performed on a sample of 133 NUTS2 regions for 16 countries and for the reference year 2007 (see section 4.1 for the details on the data).

This procedure is adopted for assessing the values of the thresholds . Specifically, we calculate confidence intervals for each index of absolute entropy defined in equation (1) and we order the 10,000 simulated values of the 19 entropy measures. We select the threshold levels (1%, 5%, 10%) from the top of our generated distribution and then record the critical values. Such critical values are the required thresholds . If the real value of the index is higher than the thresholds defined as explained above, we can conclude that spatial distribution in sector *s* does not follow a uniform law. Hence, this implies that the sector is spatially clustered because its spatial distribution is significantly different from randomness.

In the second step, we construct the counterfactual distribution of employees, considering that, under the null hypothesis, employees follow a binomial law based on the actual spatial patterns of firms.

Hence, we assume that, for each industry with *Ls* employees and *Ns* establishments the expected share of workers in sector located in region is equal to the share of establishments in the same region. So, the counterfactual distribution of employees across regions for each sector is derived starting from the real distribution of establishments. Under the null hypothesis, we are implicitly assuming that all establishments of the same industry are of equal size and there are no peculiar aggregation forces on some specific establishments.

We run 10,000 simulations for each industry employees’ data, separately. Specifically, we calculate confidence intervals for each index of relative entropy and by ordering the 10,000 simulated values of the 19 distributions of sectoral employees across regions, we select a classical 5% risk-level rejection, the associated threshold level is identified excluding the 5% of the simulations from the top of our generated distribution. This gives the critical values of the tests at the related confidence levels to which the real values of the concentration indices are compared. The null hypothesis is thus rejected when the true value of the relative entropy index is higher than the threshold . In this case, since the spatial distribution of employees is significantly different from the location of establishments, we conclude that, in this industry, employees does not follow the counterfactual distribution of establishments.

## Empirical application

## 4.1 Data

We collected data on existing establishments and employees in Europe in 2007 from the Eurostat database. What follows is a brief description of the sources, dimensions, and choices made that led to the final database being used. We aimed to provide a detailed analysis of sectoral concentration of industries based on regional (NUTS2) data for a large sample of European countries.

The official source of data on the number of employees and establishments at the regional level is the Structural Business Statistics database contained in Eurostat[[3]](#footnote-3).

The analysis considers a sample of 133 NUTS 2 level in 16 European countries: Austria, Belgium, Bulgaria, Czech Republic, Finla­nd, France, Germany, Hungary, Italy, Norway, Poland, Portugal, Romania, Spain, Sweden, and the UK. Table 1 summarizes the geographical coverage of the dataset while the complete list of the regions is given in the Appendix. The regions included in the analysis were selected on the basis of the availability of information for both metrics that are of interest for our analysis, namely employees and establishments.

***Table 1 – Geographical coverage of the dataset***

|  |  |  |
| --- | --- | --- |
| Country | Administrative partition | Number of regions included |
| Austria | States (Bundesländer) - NUTS2 | 2 |
| Belgium | Provinces - NUTS2 | 1 |
| Bulgaria | Planning regions - NUTS 2 | 4 |
| Czech Republic | Cohesion region/Region soudržnosti - NUTS 2 | 7 |
| Finland | Suuralueet - NUTS2 | 3 |
| France | Régions - NUTS2 | 22 |
| Germany | Government regions (Regierungsbezirke) -NUTS2 | 9 |
| Hungary | Planning and statistical regions (Tervezési-statisztikai régiók) - NUTS 2 | 7 |
| Italy | Regioni - NUTS2 | 18 |
| Norway | Regions - NUTS 2 | 5 |
| Poland | Regions (regiony) - NUTS2 | 11 |
| Portugal | Regions - NUTS2 | 2 |
| Romania | Regions (Regiuni) - NUTS2 | 8 |
| Spain | Comunidades autónomas -NUTS2 | 6 |
| Sweden | National Areas (Riksområden) - NUTS2 | 2 |
| United Kingdom | Counties - NUTS2 | 26 |
|  |  | 133 |
| Total |

The analysis is restricted to the manufacturing services. We focus on 19 manufacturing industries according to the NACE rev.1.1 at three-digit classification (Statistical Classification of Economic Activities in the European Community). The final collection of sectors was chosen according to data availability and includes Textiles, Wearing apparel, Wood, Paper, Publishing, printing, Chemicals, Rubber and plastics, Other non-metallic mineral products, Basic metals, Fabricated metal products, Machinery and equipment, Office machinery and computers, Electrical machinery and apparatus n.e.c., Radio, television and communication equipment, Medical, precision and optical instruments, watches and clocks, Motor vehicles, Other transport equipment, Furniture; manufacturing n.e.c., Recycling. In a very limited number of cases, we complete data based on the preceding year (2006), when they were available. In any case, we had to exclude four manufacturing sectors from our analysis (*food, beverages and tobacco, leather and footwear, coke, refined petroleum*) because of the huge amount of confidential and missing data.

Table 2 reports for each manufacturing industry, some main descriptive statistics on the employment distribution across EU regions and on the distribution of establishments across the same spatial units.

***Table 2 – Descriptive statistics of variables***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Employment | |  | Establishments | |
|  | Obs | Mean | Std. Dev. |  | Mean | Std. Dev. |
| Textiles | 133 | 5150 | 10156.16 |  | 409 | 848.4534 |
| Wearing apparel | 133 | 7328 | 12536.82 |  | 790 | 1403.223 |
| Wood | 133 | 5668 | 4990.37 |  | 998 | 1228.153 |
| Paper | 133 | 3197 | 3006.707 |  | 105 | 129.2276 |
| Publishing, printing | 133 | 8011 | 10460.99 |  | 1129 | 1680.325 |
| Chemicals | 133 | 7859 | 10948.5 |  | 193 | 263.0213 |
| Rubber and plastics | 133 | 8214 | 7954.704 |  | 368 | 511.913 |
| Other non-metallic mineral products | 133 | 6928 | 6723.805 |  | 568 | 641.9258 |
| Basic metals | 133 | 5442 | 8095.078 |  | 86 | 141.6616 |
| Fabricated metal products | 133 | 19403 | 23201.59 |  | 2027 | 2976.906 |
| Machinery and equipment | 133 | 17984 | 23947.45 |  | 909 | 1380.023 |
| Office machinery and computers | 133 | 738 | 1002.606 |  | 49 | 63.2692 |
| Electrical machinery and apparatus n.e.c. | 133 | 7969 | 8445.964 |  | 417 | 732.1853 |
| Radio, television and communication equipment | 133 | 3660 | 5047.407 |  | 175 | 271.4882 |
| Medical, precision and optical instruments, watches and clocks | 133 | 4777 | 5911.84 |  | 476 | 655.6935 |
| Motor vehicles | 133 | 9365 | 16364.25 |  | 95 | 86.5742 |
| Other transport equipment | 133 | 4522 | 5089.113 |  | 135 | 191.5075 |
| Furniture; manufacturing n.e.c. | 133 | 8322 | 9151.03 |  | 1150 | 1439.624 |
| Recycling | 133 | 806 | 718.2209 |  | 111 | 116.0304 |

**4.2 Results**

Before presenting the results of the statistical testing procedure applied to entropy-based measures we present summary statistics for the average regional firm size for all the industries and total manufacturing (See Table 3).

Figures in Table 3 clearly highlight that the industries with smaller average firm sizes are *wood, recycling, wearing apparel, fabricated metal products, textiles, Medical, precision and optical instruments, watches and clocks, furniture, manufacturing nec, Other non-metallic mineral products, publishing printing*. In these low-scale industries the regional variation in the average firm size is also quite low.

***Table 3 Average regional firm size, aggregate regional data by sector***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Obs** | **Mean** | **Std. Dev.** | **Min** | **Max** |
|  |  |  |  |  |  |
| Textiles | 133 | 19 | 22.86 | 1 | 125 |
| Wearing apparel | 133 | 15 | 24.27 | 1 | 147 |
| Wood | 133 | 11 | 14.39 | 2 | 82 |
| Paper | 133 | 43 | 41.04 | 3 | 228 |
| Publishing, printing | 133 | 17 | 59.33 | 1 | 652 |
| Chemicals | 133 | 52 | 64.54 | 7 | 474 |
| Rubber and plastics | 133 | 29 | 22.18 | 7 | 117 |
| Other non-metallic mineral products | 133 | 17 | 12.44 | 3 | 69 |
| Basic metals | 133 | 74 | 74.74 | 3 | 410 |
| Fabricated metal products | 133 | 16 | 19.73 | 3 | 97 |
| Machinery and equipment | 133 | 28 | 31.38 | 5 | 169 |
| Office machinery and computers | 133 | 34 | 107.51 | 1 | 1142 |
| Electrical machinery and apparatus n.e.c. | 133 | 40 | 47.56 | 2 | 265 |
| Radio, television and communication equipment | 133 | 39 | 60.10 | 1 | 544 |
| Medical, precision and optical instruments, watches and clocks | 133 | 16 | 23.01 | 2 | 136 |
| Motor vehicles | 133 | 114 | 180.53 | 4 | 1492 |
| Other transport equipment | 133 | 61 | 84.88 | 1 | 720 |
| Furniture; manufacturing n.e.c. | 133 | 14 | 21.24 | 2 | 117 |
| Recycling | 133 | 11 | 11.24 | 3 | 74 |
| Total manufacturing | 133 | 23 | 31.50 | 4 | 192 |

Instead, higher average firm sizes are found in sectors characterized by increasing returns to scale at the plant level such as *Motor vehicles* (113 employees in each establishment, on average), *Basic metals* (73), *Chemicals* (52), *Other transport equipment* (61).

Figure 1 plots the data on average regional firm size with the respective data on the total number of establishments for each manufacturing industry. The graph shows the inverse relationship between internal scale economies and the number of firms. The spatial concentration of an industry can be the outcome of increasing returns (the intensive margin), the result of a wide number of small firms localized in space (the extensive margin), or a combination of the two driving forces.

***Figure 1 Average regional firm size (intensive margin) and N. of establishments (estensive margin), by sector***



Table 4 summarizes the results of the test performed in the first stage, when we compare the spatial distribution of firms with the counterfactual uniform distribution, as explained in Section 3. We report the values of the normalized Shannon index for each industry , as defined in equation (1).

We reject the null hypothesis (H0) of a random location scenario in all the manufacturing industries considered in our analysis in favor of the alternative hypothesis of significant spatial concentration of establishments. In a nutshell, agglomeration is significant both for small-scale industries typically considered as localized because of industry-specific Marshallian external economies, for those industries characterized by considerable internal scale economies, and for industries that usually gain cost advantages from proximity to natural resources or to the demand.

***Table 4 Indices of absolute spatial concentration of establishments and confidence intervals***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Real values | Sign | Critical values | | |
| *s* | p90 | p95 | Macintosh HD:private:var:folders:lj:xxc8ffq97xz83mgfptyhphwc0000gn:T:TemporaryItems:msoclip:0:clip_image001.png   |  | | --- | | p99 | |
| Textiles | 0.862 | + | 0.222 | 0.232 | 0.249 |
| Wearing apparel | 0.937 | + | 0.223 | 0.232 | 0.251 |
| Wood | 0.305 | + | 0.222 | 0.230 | 0.249 |
| Paper | 0.363 | + | 0.222 | 0.232 | 0.250 |
| Publishing, printing | 0.502 | + | 0.223 | 0.233 | 0.252 |
| Chemicals | 0.548 | + | 0.224 | 0.233 | 0.250 |
| Rubber and plastics | 0.343 | + | 0.218 | 0.227 | 0.246 |
| Other non-metallic mineral products | 0.343 | + | 0.215 | 0.224 | 0.241 |
| Basic metals | 0.668 | + | 0.220 | 0.229 | 0.247 |
| Fabricated metal products | 0.410 | + | 0.219 | 0.228 | 0.247 |
| Machinery and equipment | 0.521 | + | 0.216 | 0.224 | 0.243 |
| Office machinery and computers | 0.691 | + | 0.222 | 0.232 | 0.250 |
| Electrical machinery and apparatus n.e.c. | 0.429 | + | 0.223 | 0.233 | 0.253 |
| Radio, television and communication equipment | 0.626 | + | 0.224 | 0.234 | 0.252 |
| Medical, precision and optical instruments, watches and clocks | 0.494 | + | 0.215 | 0.225 | 0.241 |
| Motor vehicles | 0.733 | + | 0.222 | 0.231 | 0.250 |
| Other transport equipment | 0.492 | + | 0.223 | 0.232 | 0.251 |
| Furniture; manufacturing n.e.c. | 0.405 | + | 0.222 | 0.231 | 0.248 |
| Recycling | 0.336 | + | 0.217 | 0.225 | 0.243 |

*The table reports the real values of absolute entropies defined in equation (1). The sign “+” indicates statistically significant spatial concentration of establishments at the 95% confidence level. Columns 4, 5 and 6 report the upper bounds of the counterfactual 90%, 95% and 99% confidence intervals.*

Establishments are purposefully clustered (+) in all industries, they are highly heterogeneous in terms of technology and scale intensity. Some industries showing significant concentration are characterized by intermediate technology level and substantial internal scale economies (i.e. motor vehicles, basic metals, chemicals, other transport equipment), other industries are low-tech industries usually considered as localized because of Marshallian external economies (i.e. wearing apparel, textiles, publishing, printing) or small-scale knowledge intensive industries (e.g office machinery and computers, radio, television and communication equipment). We may posit that in this second group of industries the systematic difference between the actual spatial distribution relative to the counterfactual uniform is mainly the result of the strength of localization economies that is a mixture of advantages accruing to firms from being located close to other establishments in the same industry (industry-specific knowledge spillovers, labour market pooling, sharing of specialized suppliers).

Finally, other industries purposely clustered – i.e. rubber and plastic products, machinery, and non-metallic minerals, paper and furniture – are populated by specialized suppliers, producing intermediate products for other manufacturing sectors. Their location patterns should not be considered purely as a product of chance not only because of an asymmetrical distribution of natural resources, we suggest that their spatial distribution is related to the clustering of downstream industries that use their products as input.

Figure 1 provides additional evidence that significant spatial concentration can be the outcome of different source of intra-industry scale economies, depending on the characteristics of the sector. For example, in the case of wearing apparel a look at Figure 1 suggest that the significant spatial location of firms should be the outcome of Marshallian externalities, since the sector is dominated by a relatively high number of establishments with a small average size. On the contrary, in sectors like motor vehicles, other transport equipment and basic metals, the higher average size combined with the smaller number of establishments confirm that the significant localization that we find is more the result of internal scale economies than of external economies.

In the second step of the testing procedure, cross-regional firm size distribution is considered. Table 5 summarizes our results on relative concentration -that is the divergence between the spatial distribution of firms and employees - and its statistical significance for 2007: we report the relative concentration values for each industry , as defined in equation (2), and the thresholds of the testing procedure where the counterfactuals are based on the real distributions of establishments.

Interestingly, also in this second step, we reject the null hypothesis in 15 out of the 19 manufacturing industries considered in our analysis in favor of the alternative hypothesis of significant dissimilarity between the two distributions as explained in Section 3. We conclude that the spatial distribution of employees is significantly different from the theoretical distribution one would have expected by looking at the distribution of firms, and this also confirms the evidence that firms are not equally sized. In particular, the variety of establishments’ size across regions is mainly responsible of the dissimilarity between the spatial distribution of employees and the spatial distribution of firms.

We suggest that a significant divergence can be a sign of competitiveness, in line with the view that heterogeneity in firm size within a sector may signal greater experimentation of new entrants and successful dynamics of most innovative and productive companies (Bartelsman *et al.*, 2009).

***Table 5 Relative entropy measures***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Real values | Sign | Critical values | | |
|  | p90 | p95 | Macintosh HD:private:var:folders:lj:xxc8ffq97xz83mgfptyhphwc0000gn:T:TemporaryItems:msoclip:0:clip_image001.png   |  | | --- | | p99 | |
| Textiles | 0.223 | + | 0.209 | 0.209 | 0.210 |
| Wearing apparel | 0.335 | + | 0.251 | 0.251 | 0.251 |
| Wood | 0.195 |  | 0.202 | 0.202 | 0.203 |
| Paper | 0.295 | + | 0.070 | 0.070 | 0.070 |
| Publishing, printing | 0.226 |  | 0.228 | 0.229 | 0.229 |
| Chemicals | 0.260 | + | 0.071 | 0.071 | 0.071 |
| Rubber and plastics | 0.177 | + | 0.077 | 0.078 | 0.078 |
| Other non-metallic mineral products | 0.147 | + | 0.102 | 0.102 | 0.102 |
| Basic metals | 0.417 | + | 0.127 | 0.127 | 0.128 |
| Fabricated metal products | 0.225 | + | 0.039 | 0.039 | 0.040 |
| Machinery and equipment | 0.308 | + | 0.047 | 0.047 | 0.048 |
| Office machinery and computers | 0.623 | + | 0.228 | 0.228 | 0.230 |
| Electrical machinery and apparatus n.e.c. | 0.532 | + | 0.166 | 0.166 | 0.167 |
| Radio, television and communication equipment | 0.572 | + | 0.182 | 0.182 | 0.183 |
| Medical, precision and optical instruments, watches and clocks | 0.380 | + | 0.069 | 0.070 | 0.070 |
| Motor vehicles | 0.570 | + | 0.191 | 0.191 | 0.191 |
| Other transport equipment | 0.404 |  | 0.441 | 0.441 | 0.442 |
| Furniture; manufacturing n.e.c. | 0.198 | + | 0.047 | 0.048 | 0.048 |
| Recycling | 0.107 |  | 0.321 | 0.322 | 0.323 |

*The table reports the real values of relative entropies defined in equation (2). The sign “+” indicates statistically significant divergence between the distribution of employees and the distribution of establishments at the 95% confidence level. Columns 4, 5 and 6 report the upper bounds of the counterfactual 90%, 95% and 99% confidence intervals.*

The testing procedure led to not reject H0 in 5 sectors, namely *Wood, Publishing, printing, Other transport equipment, and Recycling*, even though establishments are spatially clustered in these industries (Cfr. Table 3). Particularly, in the sectors *Wood, Publishing, printing,* and *Recycling*, that are characterized by a small average firm size (11, 17 and 11, respectively; see Table 2), the similarity of the geography of employees with the spatial distribution of firms lead us to conclude that they may be driven by the same exogenous factors. Location choices in these sectors sound consistent with “first nature causes”, market size, and intra-industry complementarities and spillovers among firms rather than a diversified average firm size within the sector. Moreover, a comparison between Tables 3 and 4 suggests that *Wood* and *Recycling* –which are two sectors showing an evident similarity between the distribution of the employees and the one of firms- are also sectors with small standard deviation of the size. This outcome goes in the direction of further pointing out the relationship between the null hypothesis of the relative entropy test and the equally sized condition of the firms in the sector.

To summarize, our results suggest that the spatial distributions of firms do not follow a uniform law. They also point to a systematic deviation of the spatial concentration of employees from the distribution of firms. This evidence suggest that the process of industrial agglomeration is more complex as it is assumed to be, and sound consistent with the view that factors like path dependence (see, e.g., Arthur 1989; David 1985; Boschma and Lambooy, 1999; Martin and Sunley, 2006; Martin, 2012 among others), heterogeneity in individual preferences about spatial interaction (Papageorgiu and Smith, 1983) may lead to spatial configurations that are far from being a uniform distribution of identical individuals across space. Such theoretical arguments and possible explanations behind these patterns are discussed in a companion paper (Cerqueti and Cutrini, 2021) and goes far beyond the present analysis. These issues open directions for future research that are discussed in the final sections.

# Concluding remarks and further developments

In this paper we focus on the available method to assess localization of industries, that is a phenomenon inevitably associated to territorial competitiveness.

The approach to study localization based on entropy measures is limited so far since applied researchers lack the instruments for attaching statistical significance to the observed outcomes in comparative analysis. Moreover, the widespread practice of using aggregate employees data applied to discrete space analysis has left completely on the shadow the underlying location process by firms, which is instead central in spatial point patterns. We believe that the location process by firms should be taken in due consideration also for measures based on aggregate employees’ data to better evaluate the agglomeration forces behind the observed spatial patterns. We go in this direction with a two-step testing procedure that considers both employees and establishments.

We improve the economic interpretation of entropy-based indices by simulating confidence intervals that can be used to assess the spatial concentration of industries. To this purpose, we introduce a testing procedure developed in two steps. In the first step, we evaluate absolute concentration of establishments. In this case the counterfactual distributions for the null hypothesis are constructed on randomly sampled establishments that are equally likely to be located anywhere in space. In the second step, we compute relative entropies with counterfactual distributions following a binomial law based on the real distribution of establishments across regions.

We run 10,000 simulations for each industry and we determine 10,000 simulated values of the whole set of entropy measures of absolute and relative concentration. We identify critical values of type I error (at 5% level) by ordering the 10,000 simulated values of each entropy index and selecting those that correspond to the top of each generated distribution. As an illustrative exercise, we use our approach to calculate critical values of significant concentration of manufacturing industries in Europe in 2007.

Our findings suggest that spatial concentration of industries is widespread in Europe, as a significant departure from a uniform location scenario is common to all the manufacturing sectors investigated and does not depend on technology intensity, neither it is affected by internal scale economies.

As for relative entropies, we suggest that a significant divergence can be related to competitiveness, in line with the view that heterogeneity in firm size within a sector may signal greater experimentation of new entrants and successful dynamics of most innovative and productive companies but this first intuition requires to be corroborated. In particular, it is important to understand why some regions are more apt to favor heterogeneity and the growth of firms and establishments while other are not. These are promising lines of further research that may help policy makers to define appropriate development policies to foster competitiveness at the local level.

Although the study has successfully demonstrated some empirical regularities in the spatial distribution of employees and establishments, it is subject to at least three limitations, in terms of geographical coverage, sectoral coverage and theoretical explanations. We briefly recall them.

A first limitation of this study is that it focuses on the European context. Equally important would be to extend the analysis to the United States for which data is available on a larger scale and especially to other rapidly growing geographic areas, such as Asia.

Second, the present analysis used a convenience sample as detailed sectoral-regional data with a pan-European coverage are in short supply. Hence, the analysis concerns only the manufacturing sectors due to the absence of reliable data on services. It goes without saying that the analysis of localization of economic activities of contemporary economic systems cannot ignore the service sectors. In particular, services make up the largest sector in most European economies and there are services that are of great importance for the distribution of regional income and welfare. Any full assessment of spatial concentration in Europe should include them.

Third, despite its eminently descriptive nature, this study offers some insight into the possible theoretical explanations that lies behind the observed spatial configurations. Nevertheless, further investigation is needed in this area, to improve the economic interpretation and the robustness of our findings.

There are also possible improvements of the testing procedure. A promising direction for a methodological advancement could be to refine the testing procedure with a counterfactual scenario embedded in a consistent nested model of location choice able to take into account different spatial scales. This method is versatile, and with the appropriate definition of the variable of interest, counterfactual distributions and location model, it may be easily adapted to different research questions. Particularly, it could be used to evaluate the strength of other relevant agglomeration and dispersion forces like urbanization economies and the distribution of final demand (market potential).

Entropy measures and the methodology developed here to test for significant localization may be helpful in the context of public choices and public planning. In particular, location policies can benefit from an empirical analysis able to assess whether industries are significantly localized, so to evaluate the costs and benefits associated with the current state of the system and with its possible changes.

A further possible avenue for future research should be to verify whether the results on the significance of spatial concentration are consistent with evidence arising through the methods based on spatial point patterns.

# References

Aiginger, K. and Rodrik, D. (2020), Rebirth of Industrial Policy and an Agenda for the Twenty-First

Century, *Journal of Industry, Competition and Trade*, Vol. 20, 189–207.

Alonso-Villar O., Del Río C. (2013), Concentration of Economic Activity: An Axiomatic Approach, *Regional Studies* 47(5), 756-772.

Arbia G. (2001), Modelling the geography of economic activity on a continuous space. *Papers in Regional Science* 80, 411-424.

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.

Arthur W.B. (1989), Competing Technologies, Increasing Returns, and Lock-In by Historical Events, *Economic Journal*, 99, 116-131.

Bagoulla C., Péridy N. (2011), Market access and the other determinants of North–South manufacturing location choice: An application to the Euro-Mediterranean area, *Economic Systems* 35, 537-561.

Bartelsman, E., Haltiwanger, J., Scarpetta, S. (2009), Measuring and analyzing cross-country differences in firm dynamics, in: Dunne, T., Jensen, J.B., Roberts, M.J. (eds.), *Producer Dynamics: New Evidence from Micro Data*, Chicago: University of Chicago Press, 15-76.

Bickenbach F., Bode E. (2008), Disproportionality Measures of Concentration, Specialization and Localization, *International Regional Science Review,* 31(4), 359-388.

Bickenbach F., Bode E., Krieger-Boden C. (2010), Closing the gap between absolute and relative measures of localization, concentration or specialization, *Kiel Institute for the World Economy, Working Paper 1660*.

Billings S.B., Johnson E.B. (2012a), A non-parametric test for industrial specialization, *Journal of Urban Economics*, 71(3), 312–331.

Billings S.B., Johnson E.B. (2012b), The Location Quotient as an Estimator of Industrial Concentration, *Regional Science and Urban Economics*, 42(4), 642–647.

Boschma R.A., Lambooy J.G., (1999), Evolutionary economics and economic geography, *Journal of Evolutionary Economics*, 9(4), 411-429.

Bourguignon F. (1979), Decomposable income inequality measures, *Econometrica*, 47, 901-920.

Brülhart M., Träger R. (2005), An Account of Geographic Concentration Patterns in Europe, *Regional Science and Urban Economics*,35(6), 597-624.

Cassey A.J., Smith B.O. (2014), Simulating confidence for the Ellison–Glaeser index, *Journal of Urban Economics*, 81(C), 85-103.

Cerqueti R., Cutrini E., (2021), A framework for modelling economic regional location processes under uncertainty, *Journal of Quantitative Economics,* doi: 10.1007/s40953-021-00248-w.

Ciommi M., Gigliarano C., Emili A., Taralli S., Chelli, F.M. (2017), A new class of composite indicators for measuring well-being at the local level: An application to the Equitable and Sustainable Well-being (BES) of the Italian Provinces, *Ecological Indicators*, 76, 281–296.

Combes P.-P., Mayer T., Thisse J.-F. (2008), Economic Geography: The Integration of Regions and Nations, Princeton University Press.

Cutrini E. (2009), Using entropy measures to disentangle regional from national localization patterns, *Regional Science and Urban Economics,* 39(2), 243-250.

Cutrini E. (2010), Specialization and Concentration from a Twofold Geographical Perspective: Evidence from Europe, *Regional Studies,* 44(3), 315–336.

David, P. (2000), *Path dependence, its critics and the quest for ‘historical economics’*. Oxford and Stanford University.

Devereux M.P., Griffith R., Simpson H. (2004), The geographic distribution of production activity in the UK,*Regional Science and Urban Economics,* 34(5), 533-564.

Diggle P.J. (1983), Statistical analysis of spatial point patterns, Academic Press, London.

Duranton G., Overman H.G. (2005), Testing for Localisation Using Micro-Geographic Data, *Review of Economic Studies,* 72(4), 1077–1106.

Ellison G., Glaeser E.L. (1997), Geographic concentration in U.S. manufacturing industries: a dartboard approach, *Journal of Political Economy,* 105(5), 889–927.

European Commission (2011), Regional Policy for Smart Growth in Europe, 2020.

European Commission (2014), For a European Industrial Renaissance. Brussels: *Communication COM/2014/014 final*.

European Commission (2020), A European strategy for data, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 19 February 2020, Brussels.

Evans A.J. (2010), Complex spatial networks in application, *Complexity,* 16, 11–19.

Gokan T. (2010), On the usage of the measurements of geographical concentration and specialization with areal data, in: *Spatial Statistics and Industrial Location in CLMV: An Interim Report,* cap. 2, Kuroiwa ed., Chosakenkyu Hokokusho, IDE-JETRO.

Greco S., Ishizaka A., Tasiou M., Torrisi G. (2019), On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness, *Social Indicators Research*, 141(1), 61–94.

Guimarães P., Figueiredo O., Woodward D. (2009), Dartboard tests for the location quotient, *Regional Science and Urban* *Economics,* 39(3), 360–364.

Haedo C., Mouchart M. (2018), A stochastic independence approach for measuring regional specialization and concentration, *Papers in Regional Science,* 97(4), 1151-1168.

Krishnakumar J., Nagar A.L. (2008), On exact statistical properties of multidimensional indices based on principal components, factor analysis, MIMIC and structural equation models, *Social Indicators Research,86*(3), 481–496.

Lang G., Marcon E. Puech F. (2020), Distance-based measures of spatial concentration: introducing a relative density function, *The Annals of Regional Science*, v64, 243–265.

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. (2010), Measures of the Geographic Concentration of Industries: Improving Distance-based Methods, *Journal of Economic Geography,* 10(5), 745–762.

Marcon E., Puech F. (2017), A typology of distance-based measures of spatial concentration, *Regional Science and Urban Economics*, 62, 56-67.

Martin R. (2012), (Re) placing path dependence: a response to the debate, *International Journal of Urban and Regional Research*, 36(1), 179-192.

Martin R., Sunley P. (2006), Path dependence and regional economic evolution, *Journal of Economic Geography*, 6(4), 395–437.

Maurel F., Sédillot B. (1999), A measure of the geographic concentration in French manufacturing industries, *Regional Science and Urban Economics,* 29(5), 575-604.

Mazziotta M., Pareto A. (2019), Use and misuse of PCA for measuring well-being, *Social Indicators Research*, 142, 451–476.

Mori T., Nishikimi K., Smith T.E., 2005l A divergence statistic for industrial localization, *The Review of Economics and Statistics*, 87(4), 635–651.

﻿Mulligan G.F., Schmidt C. (2005), A note on localization and specialization, *Growth and Change*, 36 (4), 565–576.

OECD (2011), ISIC Rev. 3 Technology Intensity Definition - Classification of manufacturing industries into categories based on R&D intensities, OECD Publishing.

Papageorgiou Y.Y., Smith T.R. (1983), Agglomeration as local instability of spatially uniform steady-states, *Econometrica*, 51, 1109-1119.

Schlossarek M., Syrovatka M., Vencalek O. (2019), The importance of variables in composite indices: A contribution to the methodology and application to development indices, *Social Indicators Research*, 145, 1125–1160.

Stierle-von Schultz U., Stierle M.H. (2013), Regional Specialisation and Sectoral Concentration in an Enlarged EU, *INFER WP. 2/2013*.

Vechiu N., Makhlouf F. (2014), Economic integration and specialization in production in the EU27: does FDI influence countries’ specialization?, *Empirical Economics* 46, 543-572.

WIFO (2017), Measuring Competitiveness, Background documents for the European Semester .

# Appendix

**Table B1 List of NUTS 2 regions**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| NUTS2 | Region |  | NUTS2 | Region |  | NUTS2 | region |
| BE35 | Prov. Namur |  | ES30 | Comunidad de Madrid |  | ITD2 | Provincia Autonoma Trento (NUTS 2006) |
| BG32 | Severen tsentralen |  | FR10 | Île de France |  | ITD3 | Veneto (NUTS 2006) |
| BG34 | Yugoiztochen |  | FR21 | Champagne-Ardenne |  | ITD4 | Friuli-Venezia Giulia (NUTS 2006) |
| BG41 | Yugozapaden |  | FR22 | Picardie |  | ITD5 | Emilia-Romagna (NUTS 2006) |
| BG42 | Yuzhen tsentralen |  | FR23 | Haute-Normandie |  | ITE1 | Toscana (NUTS 2006) |
| CZ01 | Praha |  | FR24 | Centre (FR) |  | ITE2 | Umbria (NUTS 2006) |
| CZ02 | Strední Cechy |  | FR25 | Basse-Normandie |  | ITE3 | Marche (NUTS 2006) |
| CZ03 | Jihozápad |  | FR26 | Bourgogne |  | ITE4 | Lazio (NUTS 2006) |
| CZ05 | Severovýchod |  | FR30 | Nord - Pas-de-Calais |  | ITF1 | Abruzzo |
| CZ06 | Jihovýchod |  | FR41 | Lorraine |  | ITF3 | Campania |
| CZ07 | Strední Morava |  | FR42 | Alsace |  | ITF4 | Puglia |
| CZ08 | Moravskoslezsko |  | FR43 | Franche-Comté |  | ITF5 | Basilicata |
| DE11 | Stuttgart |  | FR51 | Pays de la Loire |  | ITF6 | Calabria |
| DE13 | Freiburg |  | FR52 | Bretagne |  | ITG1 | Sicilia |
| DE21 | Oberbayern |  | FR53 | Poitou-Charentes |  | ITG2 | Sardegna |
| DE30 | Berlin |  | FR61 | Aquitaine |  | HU10 | Közép-Magyarország |
| DEA1 | Düsseldorf |  | FR62 | Midi-Pyrénées |  | HU21 | Közép-Dunántúl |
| DEA2 | Köln |  | FR63 | Limousin |  | HU22 | Nyugat-Dunántúl |
| DEA5 | Arnsberg |  | FR71 | Rhône-Alpes |  | HU23 | Dél-Dunántúl |
| DED1 | Chemnitz (NUTS 2006) |  | FR72 | Auvergne |  | HU31 | Észak-Magyarország |
| DEG0 | Thüringen |  | FR82 | Provence-Alpes-Côte d'Azur |  | HU32 | Észak-Alföld |
| ES11 | Galicia |  | FR92 | Martinique (FR) |  | HU33 | Dél-Alföld |
| ES12 | Principado de Asturias |  | FR94 | Réunion (FR) |  | AT31 | Oberösterreich |
| ES13 | Cantabria |  | ITC1 | Piemonte |  | AT32 | Salzburg |
| ES21 | País Vasco |  | ITC4 | Lombardia |  | PL11 | Lódzkie |
| ES24 | Aragón |  | ITD1 | Provincia Autonoma Bolzano/Bozen (NUTS 2006) |  | PL12 | Mazowieckie |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| NUTS2 | Region |  | NUTS2 | region |  | NUTS2 | region |
| PL21 | Malopolskie |  | UKD5 | Merseyside (NUTS 2006) |  | NO04 | Agder og Rogaland |
| PL22 | Slaskie |  | UKE3 | South Yorkshire |  | NO05 | Vestlandet |
| PL32 | Podkarpackie |  | UKE4 | West Yorkshire |  | NO06 | Trøndelag |
| PL34 | Podlaskie |  | UKF2 | Leicestershire, Rutland and Northamptonshire |  |  |  |
| PL41 | Wielkopolskie |  | UKG1 | Herefordshire, Worcestershire and Warwickshire |  |  |  |
| PL42 | Zachodniopomorskie |  | UKG2 | Shropshire and Staffordshire |  |  |  |
| PL43 | Lubuskie |  | UKG3 | West Midlands |  |  |  |
| PL51 | Dolnoslaskie |  | UKH1 | East Anglia |  |  |  |
| PL52 | Opolskie |  | UKH2 | Bedfordshire and Hertfordshire |  |  |  |
| PT11 | Norte |  | UKH3 | Essex |  |  |  |
| PT16 | Centro (PT) |  | UKI1 | Inner London |  |  |  |
| RO11 | Nord-Vest |  | UKI2 | Outer London |  |  |  |
| RO12 | Centru |  | UKJ1 | Berkshire, Buckinghamshire and Oxfordshire |  |  |  |
| RO21 | Nord-Est |  | UKJ2 | Surrey, East and West Sussex |  |  |  |
| RO22 | Sud-Est |  | UKJ3 | Hampshire and Isle of Wight |  |  |  |
| RO31 | Sud – Muntenia |  | UKJ4 | Kent |  |  |  |
| RO32 | Bucuresti – Ilfov |  | UKK1 | Gloucestershire, Wiltshire and Bristol/Bath area |  |  |  |
| RO41 | Sud-Vest Oltenia |  | UKK2 | Dorset and Somerset |  |  |  |
| RO42 | Vest |  | UKK3 | Cornwall and Isles of Scilly |  |  |  |
| FI13 | Itä-Suomi (NUTS 2006) |  | UKL1 | West Wales and The Valleys |  |  |  |
| FI18 | Etelä-Suomi (NUTS 2006) |  | UKL2 | East Wales |  |  |  |
| FI19 | Länsi-Suomi |  | UKM2 | Eastern Scotland |  |  |  |
| SE22 | Sydsverige |  | UKM3 | South Western Scotland |  |  |  |
| SE31 | Norra Mellansverige |  | UKN0 | Northern Ireland (UK) |  |  |  |
| UKD2 | Cheshire (NUTS 2006) |  | NO01 | Oslo og Akershus |  |  |  |
| UKD3 | Greater Manchester |  | NO03 | Sør-Østlandet |  |  |  |

1. It is worth noting that the version of the index proposed by Maurel and Sédillot (1999) does not allow to control for the economic size of each region since they simply calculate G as a difference between the spatial concentration of the industry and the spatial concentration of the aggregate sector. Therefore in Maurel and Sédillot (1999), G measures the difference between two levels of absolute concentration (and do not measure relative concentration which is instead the G in Ellison and Glaeser). In this way they cannot grasp the dissimilarity in the spatial distribution of the two economic activities. Therefore, comparison across industries may lead to awkward interpretations from the standpoint of localisation, since in Maurel and Sedillot (1999) it may occur that two industries A and B (with the same industrial structure) have the same degree of geographical concentration even if the spatial distribution of employees across regions in industry A span exactly proportionally to total employees while industry B is mostly located in the less economically advanced regions.  [↑](#footnote-ref-1)
2. Concentration of a sector is detected comparing the Kernel density function of bilateral distances between pairs of establishments to a “theoretical” random distribution of an equivalent set of firms over a wider space, namely all sites (postcode level) where at least one manufacturing establishment is located. They simply reshuffle locations where at least one establishment of the aggregate economic activity exists. [↑](#footnote-ref-2)
3. See https://ec.europa.eu/eurostat/web/structural-business-statistics. [↑](#footnote-ref-3)