# A TERRITORIAL INDEX OF POTENTIAL AGGLOMERATION ECONOMIES FROM URBANIZATION

**André LEMELIN**
**Fernando RUBIERA-MOROLLÓN**
**Ana GÓMEZ-LOSCOS**

**Centre - Urbanisation Culture Société**

INRS
Université d'avant-garde

**www.ucs.inrs.ca**

# A TERRITORIAL INDEX OF POTENTIAL AGGLOMERATION ECONOMIES FROM URBANIZATION

**André LEMELIN**
**Fernando RUBIERA-MOROLLÓN**
**Ana GÓMEZ-LOSCOS**

**André Lemelin**
Centre – Urbanisation Culture Société, INRS
**andre.lemelin@ucs.inrs.ca**


**Fernando Rubiera-Morollón**
REGIOlab - University of Oviedo, Oviedo (Spain)


**Ana Gómez-Loscos**
Banco de España (Spain)

*This document can be downloaded without cost at:*
www.ucs.inrs.ca/sites/default/files/centre_ucs/pdf/Inedit03-12.pdf

**Abstract**

Cities form and grow to exploit economies of agglomeration. Whence the need in empirical spatial analysis for some type of variable that informs about agglomeration economies. What is typically measured is agglomeration itself (city population, employment...), a measure of the potential for agglomeration economies, not of agglomeration economies themselves. We develop an index to measure a territory's potential for agglomeration economies. Our index has several desirable properties. It has a graphical representation closely analogous to the development of the Gini coefficient from the Lorenz curve and can be formulated as an extension of the Hirschman-Herfindahl concentration index. It varies between 0 and 1, conforms to the Pigou-Dalton transfer principle, and lends itself to territorial aggregation and disaggregation. It can be related to the elasticity parameter of the Pareto distribution, a generalization of Zipf's rank-size rule. We apply our index to the Spanish NUTS III regions using local labor markets as the basic units for its construction, instead of administrative territorial divisions. The index's performance is evaluated by examining its correlation with the location quotients of activities known to be highly sensitive to agglomeration economies.

**Key Words:**
Urban and Regional Economics, Urbanization, Agglomeration Economies, Indexes and Spain

---

**Résumé**

Les villes se constituent et croissent pour exploiter les économies d'agglomération. D'où le besoin, dans les études empiriques, d'un indicateur des économies d'agglomération. C'est généralement le degré d'agglomération (population, emploi total...) qui est mesuré, c'est-à-dire le *potentiel* d'économies d'agglomération, non les économies elles-mêmes. Nous élaborons un indice pour mesurer le potentiel d'économies d'agglomération d'un territoire. Notre indice a plusieurs propriétés désirables. Sa représentation graphique est similaire à celle du coefficient de Gini à partir de la courbe de Lorenz et il peut être formulé comme une extension de l'indice de concentration de Hirschman-Herfindahl. Il varie entre 0 et 1, respecte le principe de transfert de Pigou-Dalton et se prête à l'agrégation ou la désagrégation. Il peut être mis en relation avec le paramètre d'élasticité de la distribution de Pareto, une généralisation de la loi rang-taille de Zipf. Nous appliquons cet indice aux régions NUTS III d'Espagne, avec, pour unités de base, les

marchés locaux du travail, plutôt que des divisions administratives. La performance de l'indice est évaluée selon son degré de corrélation avec les quotients de localisation d'activités connues pour leur sensibilité extrême aux économies d'agglomération.

**Mots clés :**
Économie régionale et urbaine, Urbanisation, Économies d'agglomération, Nombres indices et Espagne

**Resumen**

El papel de las economías externas de aglomeración en el crecimiento económico de los territorios es fundamental. De ahí que en estudios espaciales y regionales necesitemos frecuentemente disponer de una medida precisa de la dimensión de las aglomeraciones urbanas por zonas o regiones. En este trabajo se propone un índice para medir el potencial de un territorio para generar economías de aglomeración. Nuestro índice tiene varias propiedades deseables: (i) tiene una representación gráfica estrechamente análoga al modo de representar el coeficiente de Gini de la curva de Lorenz; (ii) se puede formular como una extensión del índice de concentración de Hirschman-Herfindahl; (iii) varía entre 0 y 1; (iv) cumple con el principio de transferencia de Pigou-Dalton; (v) se presta a la agregación y desagregación territorial y (vi) se puede relacionar con el parámetro de elasticidad de la distribución de Pareto o con una generalización de la regla de Zipf rango-tamaño. A modo de ilustración aplicamos esta propuesta de índice de aglomeración urbana a las provincias españolas (regiones NUTS III). Se usan datos los mercados de trabajo locales, en lugar de las divisiones territoriales administrativas, como las unidades básicas para su construcción. Su capacidad es evaluada mediante el examen de su correlación con los cocientes de localización de actividades altamente sensibles a las economías de aglomeración obteniendo resultados muy interesantes.

**Palabras claves :**
Economía regional y urbana, Urbanización, Economías de aglomeración, Números índices y España

## INTRODUCTION

Cities are fundamental to understand industrialization, service development, the behaviour of developed economies and globalization. Jane Jacobs (1969, 1984) is a pioneer of this viewpoint, and there has been much research in the last five decades confirming how important cities are for economic performance. Numerous studies have confirmed the positive relationship between *per capita* income and urbanization (see Fay and Opal, 2000; Jones and Koné, 1996 or Lemelin and Polèse, 1995 among others). Other studies have repeatedly proved the disproportionate contribution of urban areas to national income and production (for example, Prud'Homme, 1997; Petersen, 1991 or World Bank, 1991). Still others have found a positive link between productivity and the agglomeration of economic activity in cities (see Ciccone and Hall, 1976; Glaeser, 1994; Henderson, 1988, 2003 or Krugman, 1991 among others).

Indeed, it is commonplace to say that cities form and grow to exploit economies of agglomeration. This is undoubtedly a crucial phenomenon in the processes of economic and social development and industry specialization of a territory. Consequently, there is a need in empirical spatial analysis for some type of variable that informs about agglomeration economies, and the importance of this variable is clearly increasing. But measuring agglomeration economies poses a great challenge, most notably because they cannot be observed directly.

As a matter of fact, what is typically measured is agglomeration itself (for instance, when the unit of analysis is the individual urban area, city population, or employment). A measure of agglomeration is not a measure of agglomeration *economies*, but rather a measure of the *potential* for agglomeration economies. There are several aspects of agglomeration which may be relevant: total population, population with a university degree, total employment, sectoral employment, the size of the workforce in certain occupational categories, etc.

Additional difficulties arise when the unit of analysis is a territory comprising several urban areas, instead of the individual urban area: we then have to deal with the problem of aggregation. In particular, when using normative regions[1], which is the common practice, the analysis of territorially aggregated data may be subject to the ecological inference fallacy (first introduced by Robinson, 1950, and studied by many other authors since; see for example Richardson, 1992), or the modifiable areal unit problem (Openshaw and Taylor, 1981 and Arbia, 1989). In other words, the area or region created is not necessarily homogeneous, a problem which is also referred to in the literature as the aggregation bias (Fotheringham and Wong, 1991; Paelinck, 2000).

---

[1]  The expression « normative regions » refers to institutional territorial divisions, as opposed to analytical regions. It is used in particular to characterize the basic principles applied to define the EU nomenclature of territorial units for statistics (NUTS).

This is the issue which we address in this paper: how to measure potential agglomeration economies for a territory which does not constitute a single urban area. More specifically, we propose an index of the potential for agglomeration economies of the urbanization type[2].

A good index of potential agglomeration economies should be grounded in a clear understanding of their nature. Firstly, agglomeration economies arise in urban areas. Cities provide the fundamental ability to interact "productively". Urban areas are places where people can congregate in safety to trade, communicate and work. Cities allow goods, ideas and people to come together for purposes of exchange and production, in turn allowing society to reap the gains from trade and specialisation. It is difficult to imagine a modern market economy without market places. Indeed, it can be argued that this is the (economic) essence of a city. Most cities and towns first arose as market centres and centres of distribution and finance, well before the advent of the modern industrial era[3]. When urban concentration increases, firms within the same industry benefit through lower recruitment and training costs (shared labor force), knowledge spillovers, lower industry-specific information costs and increased competition (see Rosenthal and Strange, 2001; Beardsell and Henderson, 1999 and Porter, 1990). Certain infrastructures – international airports, post-graduate universities, research hospitals – are viable only in large metropolises. Recent literature stresses the positive link between productivity and the presence of a diversified, highly-qualified and versatile labor pool (see Duranton and Puga, 2002 or Glaeser, 1998). As highlighted by Hall (2000), large metropolises stimulate the exchange of knowledge. Activities that are characterized by a need for high creativity and innovation will, in general, choose to locate in major metropolitan areas or close to them.

The first thing we note in this respect is that the degree of urbanization (percentage of population living in an urban area) cannot be a satisfactory indicator of the potential for agglomeration economies. Beyond the difficulty of drawing a line between urban and non-urban (to which we return below), the urban/non-urban dichotomy is not sufficient to characterize agglomeration, because agglomeration economies are also related to density, concentration, mass. Consider two territories with the same degree of urbanization, one with several small-to-medium sized cities, the other with a single metropolis. Even casual observation reveals that *coeteris paribus* the second is likely to display greater economic dynamism.

---

[2]  "Urbanization economies", as opposed to "localization economies", which arise from many firms in the same industry locating close to each other

[3]  See Henderson (2003), Glaeser (1998) or Quigley (1998), among others, to extend this idea.

The index presented in this paper is constructed from local area data, which it aggregates to characterize the agglomeration economies potential of a broader territory[4]. Thus, it is implicitly assumed that the component local areas are defined in an economically meaningful way; otherwise, the validity of the underlying data is questionable, and the index can only be as good as its underlying data. Specifically, we illustrate the usefulness of our proposal by applying it to the Spanish provinces (NUTS III regions). As basic units for the construction of the index, we use local labor markets (henceforth LLMs: Sforzi and Lorenzini, 2002; ISTAT, 2006; Boix and Galleto, 2006).

Spain is a very good example because data availability forces much of the empirical research to be conducted at the NUTS II or NUTS III regional level, even though a different geography may be preferable. The Spanish National Statistical Institute (INE) provides most of the main economic information (e.g. GDP, stock of capital, wages or employment data) for the whole country or at this level of disaggregation. There is little available data concerning finer aggregation levels, such as municipalities. Fortunately, population data, that is the one we just need to compute or index is available. To find lower aggregation levels of economic information we have to look up some very specific databases or use the data provided by taxes or unemployment registers. This scarceness of finer information is also found in many other countries, and our proposed index could be useful in all such cases, as well as being valuable for carrying out cross-country comparative analyses. Moreover, we show that our index can also be applied with less data requirements.

After a brief survey of the literature (Section 2.1), the proposed index is first presented graphically, in a manner that is closely analogous to the development of the Gini coefficient from the Lorenz curve (Section 2.2). As it turns out, our index can be formulated as an extension of the Hirschman-Herfindahl concentration index (hereinafter HH), and it can be computed in a straightforward way. In Section 2.3, the index is extended to deal with the boundary problem that arises when economically significant urban areas (LLMs) extend into more than one territory (Spanish province). We then explore the properties of the index in some detail (Section 2.4): it varies between 0 and 1, conforms to a version of the Pigou-Dalton transfer principle, and lends itself to territorial aggregation and disaggregation. Finally, we investigate the index's relationship with the elasticity parameter of the Pareto distribution, a generalization of Zipf's rank-size rule (Section 2.5). The proposed index is applied to the Spanish provinces (Section

---

[4] It could be argued that, while the degree of urbanization cannot be a satisfactory indicator of the potential for agglomeration economies, our proposed index could be used as an alternative measure of urbanization.

3.1), and its performance is evaluated by examining its correlation with the location quotients of activities that are known to be very sensitive to agglomeration economies, such as knowledge intensive business services, and by comparing it with that of other measures (Section 3.2). The main conclusions and future extensions are presented in the final section. In Appendix I, we develop a version of the index which can accommodate data in a less detailed format. In Appendix II, we present an alternative approach to the boundary problem arising from LLMs extending beyond provincial borders, and compare results. Appendix III verifies how the Pigou-Dalton principle applies to the version of the index that deals with the boundary problem. Finally, Appendix IV develops the derivative of the index relative to the Pareto parameter.

## PROPOSED INDEX

### Measures of urbanization and potential agglomeration economies

The United Nations (UN) define the *degree of urbanization* as the share of total national population living in an urban environment. The most common way to measure the degree of urbanization is just to compute the proportion of population living in urban areas above a certain size, for instance 50 000 inhabitants. However, definitions of what constitutes an urban area vary across countries: for example, two neighboring population concentrations each below the urban threshold may or may not be amalgamated into a single urban area, depending on the rules that define how urban areas are delineated. Different definitions of urban areas can give different rankings of urbanization in cross-country comparisons. Furthermore, a country with a large population in large cities could show a similar degree of urbanization to a country with a small population and small cities. In other words, the degree of urbanization suffers from the lack of a standard definition. Moreover, as we mentioned earlier, the degree of urbanization is not a suitable option for measuring the potential for urbanization economies. Indeed, our view of agglomeration economies implies that a territory which has two cities with populations of, say, 50 000 inhabitants each has less potential than a region with a single city of 100 000 inhabitants.

As for the share of urban population that lives in the largest city, called *primacy* or *urban primacy* (see Ades and Glaeser, 1995), it is a very coarse measure of urban concentration, and scarsely a candidate for measuring the potential for agglomeration econcomies. More sophisticated indexes of spatial concentration have been used. One is the Hirschman-Herfindahl index (Wheaton and Shishido, 1981); the proposed index is closely related to the Hirschmann-Herfindahl index (see below, after equation [14])[5]. Brülhart and Sbergami, (2009) apply Theil entropy indexes scaled to regional areas to measure regional concentration in 16 European

---

[5]  Although frequently applied as a index of urban concentration within a country, it was originally proposed as a measure of market concentration, or market power.

countries. Rosen and Resnick (1980) examine the Pareto parameter for a sample of 44 countries, finding that it is quite sensitive to the definition of the city and the choice of city sample size. Long and Nucci (1997) recalculate the Hoover index and consider the population distribution across territorial units to show the scale at which concentration or dispersion occurs across US states. But, as measures of the potential for agglomeration economies, concentration indexes share a conceptual deficiency with the degree of urbanization: they are insensitive to absolute population magnitudes.

Uchida and Nelson (2010) have recently proposed a broader agglomeration index, based on three factors: population density, population in a large urban center and travel time to that urban center. However, applying their index is hampered by the difficulty of obtaining the required information, and by the fact that measures of accessibility are not standard. On the other hand, Arriaga (1970) had proposed an urbanization index which is extremely close to our proposed index (it is further discussed below, after equation [12]).

Our proposal presents quite a few advantages over other approaches. Firstly, it takes into consideration both concentration and size. Secondly, it does not require defining an urban threshold. Thirdly, its data requirements are modest[6], and its computation can be adapted to situations where data is less detailed (see Appendix I). Finally, it has several desirable properties, as will be shown in Section 2.3 below.

**Interpretation and computation of the index**

To develop the idea underlying our proposal, we shall first illustrate our approach using fictitious data as an example[7]. Table 1 gives the population of each of 5 urban areas in one of the territories under investigation.

**Table 1 - Fictitious data**

|  | (1) Population | (2) Population share (%) | (3) Cumulated population | (4) Cumulative distribution (%) |
|---|---|---|---|---|
| **Urban area 1** | 35 000 | 7.6 | 35 000 | 7.6 |
| **Urban area 2** | 55 000 | 12 | 90 000 | 19.6 |
| **Urban area 3** | 70 000 | 15.2 | 160 000 | 34.8 |
| **Urban area 4** | 100 000 | 21.7 | 260 000 | 56.5 |
| **Urban area 5** | 200 000 | 43.5 | 460 000 | 100.0 |
| **Total** | **460 000** |  |  |  |

---

[6]  Although it may be challenging to define economically meaningful areas to compute the underlying population data.

[7]  The use of fictitious data is strictly for expositional convenience. In the Spanish data we use later on, provinces display a tight clustering of small urban areas which makes graphs less clearly illustrative of the concepts we want to present.

As mentioned before, a common measure of the degree of urbanization is the proportion of population living in urban areas above a certain size. For instance, given the data in Table 1, the degree of urbanization could be measured as the proportion of population living in urban areas with a population of at least 50 000. In our example this would be 92.4% (425 000 / 460 000). Referring to the cumulative distribution (Table 1, column 4), this way of measuring urbanization amounts to lumping together all categories but the first. The conventional measure of urbanization is therefore based on a highly simplified representation (Figure 1) of the more detailed distribution of population, represented in Figure 2.

**Figure 1 – Cumulative distribution underlying the conventional measure of urbanization**
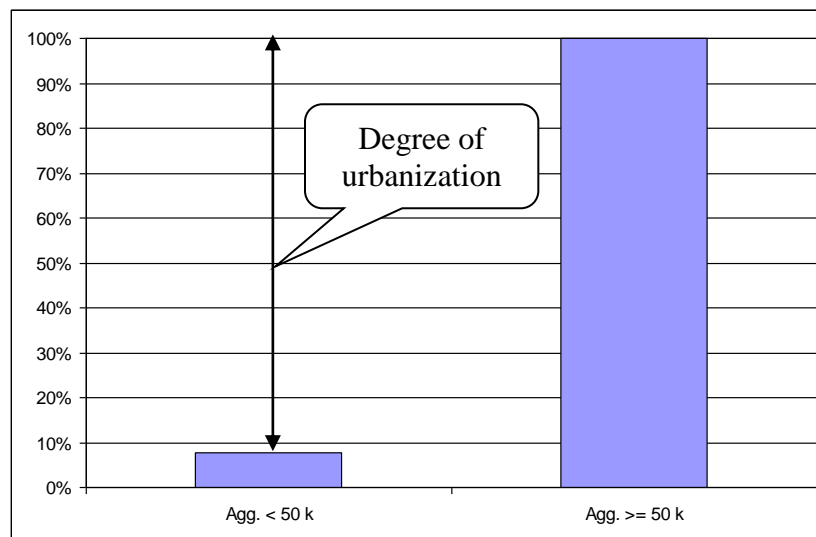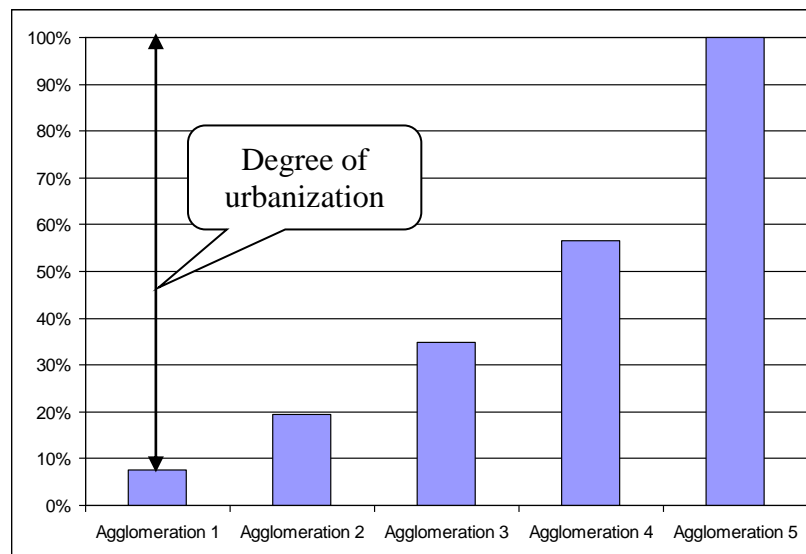


**Figure 2 – Cumulative distribution of population among urban areas**

In our example, the conventional degree of urbanization is the sum of population shares of all urban areas but the first (Table 1, column 2). We shall try to make use of the more detailed information given by the cumulative distribution in Figure 2. One way of doing so would be to take the sum of distances above each bar in Figure 2. Algebraically, let $f_i$ be the fraction of population residing in the $i$th urban area (LLM), and let $F_i = \sum_{j=1}^{i} f_j$ be the cumulative distribution[8]. The sum of distances above the bars is given by:

$$\sum_{i=1}^{5}(1-F_i) = \sum_{i=1}^{5}\left(1-\sum_{j=1}^{i} f_j\right) = \sum_{i=1}^{4}\sum_{j=i+1}^{5} f_j = \sum_{i=2}^{5} f_i + \sum_{i=3}^{5} f_i + \sum_{i=4}^{5} f_i + \sum_{i=5}^{5} f_i = \sum_{i=2}^{5}(i-1)f_i \qquad [1]$$

It is a weighted sum of all urban area sizes except for the one that has a population below the urban threshold of 50 000; the weight of each urban area is its rank $i$ minus 1.

But note that the bar chart in Figure 2 does not reflect the relative sizes of urban areas. This is correctly taken into account in Figure 3.

**Figure 3 – Cumulative distribution with city sizes on the X-axis**
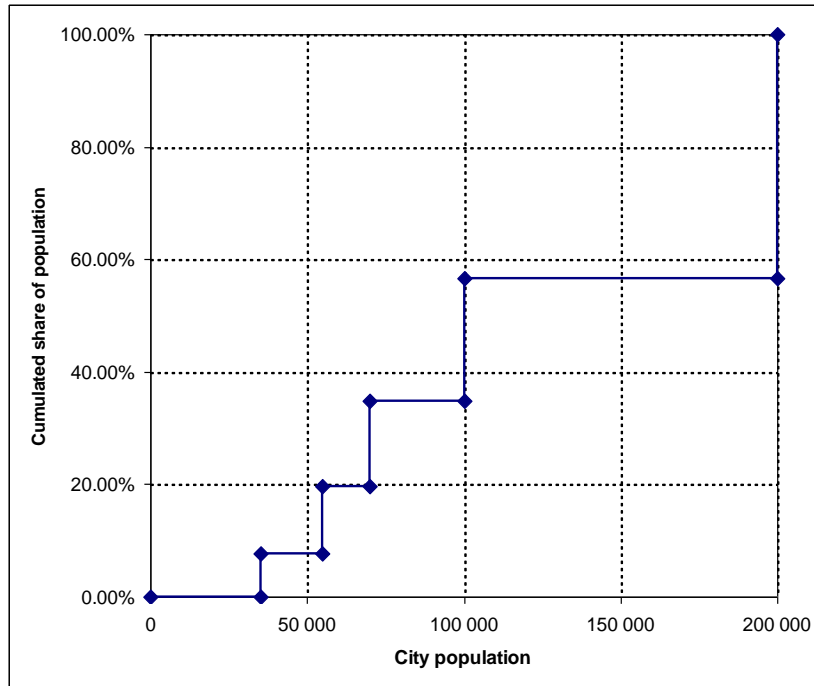


---

The measure of potential agglomeration economies which we formally develop below corresponds to the area above the cumulative distribution in Figure 3. This is remindful of the Gini coefficient which is equal to the area above the Lorenz curve9. But before proceeding, one last development is in order. A convenient measure of the degree of urbanization should vary from zero to 1, or 100%. In order to achieve this, we need some kind of normalization rule. One way of doing this is to rescale the horizontal axis of the cumulative distribution in terms of the size of the largest city in the urban system under study (in our application it will be Madrid).

So let $x_i$ be the population size of urban area $i$, relative to the largest city in the urban system considered:

$$x_i = \frac{\text{Population of urban area } i}{\text{Population of the largest city in the urban system}} \qquad [2]$$

Define $x_0 = 0$ and let $N$ be the number of urban areas. As before,

$$f_i = \frac{x_i}{\sum_{j=1}^{N} x_j} = \frac{x_i}{\sum_{j=0}^{N} x_j} \qquad [3]$$

is the fraction of population residing in the $i^{\text{th}}$ urban area (with $f_0 = 0$), and $F_i = \sum_{j=1}^{i} f_j = \sum_{j=0}^{i} f_j$ is the cumulative distribution (with $F_0 = 0$). Urban areas are assumed to be ordered from smallest to largest. Our proposed index is:

$$I = \sum_{i=1}^{N} (1 - F_{i-1})(x_i - x_{i-1}) \qquad [4]$$

Note that the first term of formula [4] is the area above the curve to the left of the first urban area in Figure 3. This implies that there is no threshold below which a populated area is considered non-urban. Deleting that first term would be an error, since it would result in an index that is insensitive to a horizontal shift in the cumulative distribution. In Figure 4, the blue line reproduces the distribution in Figure 3, while the red line represents a rightward shift in the distribution. If we do not count rectangles ABCD and AECF, the area above both distributions is the same, and so is the value of the index, whereas all urban areas in the shifted (red) distribution are larger. Of course, it would be possible to take into account a minimum urban size by

---

9   A major difference, however, is that, in the Lorenz curve, the abscissa is also a cumulative distribution.

measuring the surface above the cumulative distribution only to the right of some threshold value (say, 50 000 inhabitants); but that would only make calculations messy. Moreover, it could be argued that there is some potential for agglomeration economies even in small villages.

**Figure 4 – Cumulative distribution with city sizes on the X-axis**
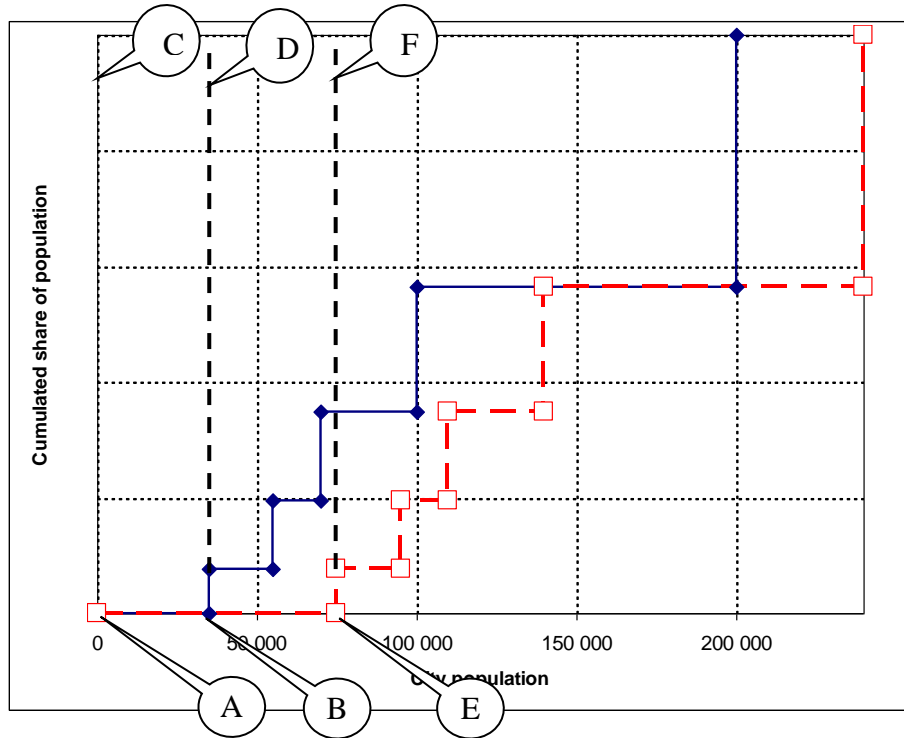


It is easily verified that our index weighs the population shares of urban areas according to their relative sizes. Equation [4] can be written as:

1) $I = \sum_{i=1}^{N} \left( 1 - \sum_{j=0}^{i-1} f_j \right) (x_i - x_{i-1})$ [5]

2) $I = \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} \left( \sum_{j=0}^{i-1} f_j \right) x_i - \sum_{i=1}^{N} x_{i-1} + \sum_{i=1}^{N} \left( \sum_{j=0}^{i-1} f_j \right) x_{i-1}$ [6]

3) $I = \sum_{i=1}^{N} \left( \sum_{j=i}^{N} f_j \right) x_i - \sum_{i=1}^{N} \left( \sum_{j=i}^{N} f_j \right) x_{i-1}$ [7]

Remebering that $x_0 = 0$,

1) $\quad I = \sum_{i=1}^{N}\left(\sum_{j=i}^{N} f_j\right) x_i - \sum_{i=2}^{N}\left(\sum_{j=i}^{N} f_j\right) x_{i-1}$ [8]

2) $\quad I = \sum_{i=1}^{N}\left(\sum_{j=i}^{N} f_j\right) x_i - \sum_{i=1}^{N-1}\left(\sum_{j=i+1}^{N} f_j\right) x_i$ [9]

3) $\quad I = f_N x_N + \sum_{i=1}^{N-1}\left(\sum_{j=i}^{N} f_j\right) x_i - \sum_{i=1}^{N-1}\left(\sum_{j=i+1}^{N} f_j\right) x_i$ [10]

4) $\quad I = f_N x_N + \sum_{i=1}^{N-1}\left(\sum_{j=i}^{N} f_j - \sum_{j=i+1}^{N} f_j\right) x_i$ [11]

5) $\quad I = f_N x_N + \sum_{i=1}^{N-1} f_i x_i = \sum_{i=1}^{N} f_i x_i$ [12]

The formula in equation [12] has the merit of being much simpler and easier to compute than the one in equation [4]. In addition, formula [12] lends itself to the interpretation given in Arriaga (1970, p. 209)[10]: since $f_i$ is the probability that a randomly chosen individual reside in urban area $i$, then the average individual in the province lives in an urban area whose size is $I$ times the size of the largest city in the urban system under study. Let us pursue the development, substituting from [3],

6) $\quad f_i = \dfrac{x_i}{\sum_{j=1}^{N} x_j} = \dfrac{x_i}{\sum_{j=0}^{N} x_j}$ [3]

we find:

7) $\quad I = \left(\sum_{i=1}^{N} f_i^2\right)\left(\sum_{j=0}^{N} x_j\right)$ [13]

---

[10] Indeed, the proposed index is identical to the theoretical index proposed by Arriaga, except for its division by the population of the largest city in the urban system under study. Arriaga investigates the implications of using a truncated index which ignores the bottom end of the size distribution of agglomerations, and concludes that a truncated index is a good approximation, under mildly restrictive hypotheses. But in our case, there is no truncation, because the local labor markets (LLMs) cover the whole territory: see 3.1 below. Appendix I presents a version of the index that is based on information aggregated by size classes and therefore deals with truncation.

where:

8) $H = \sum_{i=1}^{N} f_i^2$  [14]

is the HH index of concentration. So:

9) $I = \left( \sum_{j=0}^{N} x_j \right) H$  [15]

Our index is the HH index of concentration, multiplied by a factor that reflects the size of the province relative to the size of the largest city of the urban system considered (in our empirical application, Section 3, it will be the city of Madrid). In other words, our index accounts for both concentration and (relative) scale, which we consider as two aspects of the potential for economies of agglomeration. But, contrary to the HH index, its inverse does not have a straightforward interpretation[11].

It is shown in the Appendix I that it would also be possible to compute the index, and obtain similar results, when the underlying population data is available only in city-size categories, rather than for each urban area. This is a strength of the proposal as it would allow comparing the potential for agglomeration economies of countries with different ways of providing the statistical information (in city-size categories or for each urban area) without affecting the index's performance.

**Boundary problem**

So far, the proposed index has been developed under the assumption that every urban area (LLM) $i$ is entirely contained in a single territory (province). But that is seldom true. In the Spanish case, several LLMs include some municipalities that are located in a different province. As before, we define $f_i$ as the fraction of population residing in the $i$th urban area. Given definition [2],

1) $x_i = \dfrac{\text{Population of urban area } i}{\text{Population of the largest city in the urban system}}$  [2]

---

[11] Adelman's (1969) has shown the "numbers-equivalent" property of the Hirchman-Herfindahl index ($H$): its inverse (1/$H$) can be interpreted as the number of equal-sized urban areas which would exhibit a concentration level equal to $H$.

if only part of urban area *i* belongs to the territory under consideration, then equation [3] is no longer true, and

2) $f_i \neq \dfrac{x_i}{\sum\limits_{j=1}^{N} x_j}$ [16]

However, bearing in mind Arriaga's (1970) interpretation, the average (relative) size of the urban area where a randomly chosen individual lives is still given by formula [12]:

3) $I = f_N x_N + \sum\limits_{i=1}^{N-1} f_i x_i = \sum\limits_{i=1}^{N} f_i x_i$ [12]

But the tight relationship of our index with the HH concentration index breaks down.

In Section 3 below, our index combines the relative sizes $x_i$ of LLMs, and population shares $f_i$ computed at the scale of component municipalities, based on the province in which each municipality is actually located. In the Appendix II, this is compared with an index for which each LLM has been attributed in its entirety to the province where its centroid is located, which is tantamount to redrawing provincial boundaries. Interestingly, at least in the Spanish case, the two versions of the index are tightly correlated across the 52 provinces.

**Properties of the index**

The proposed index in non-negative, and its lower bound is zero. This extreme case would be approximated if all of the population lived in rural areas, in very small villages (of, say, 5 000 inhabitants); the curve would then be close to an upside-down «L», with the horizontal line at the 100% level, and the vertical line at the 5 000 population level, as illustrated in Figure 5.

**Figure 5 – Cumulative distribution, no urbanization**
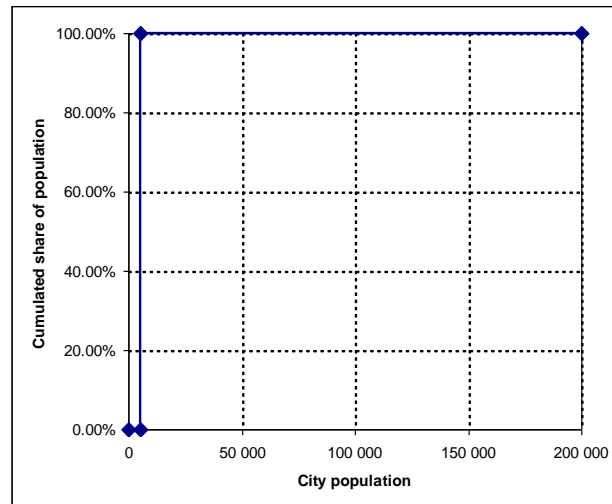


The upper bound of the index is 1. This would correspond to a province whose population would be concentrated in a single urban area equal in size to the reference city, Madrid. The curve would then be a mirror-image of an «L», with the vertical bar to the right. Actually, this latter case is not too different from the Madrid province, where 98.2% of the population lives in the Madrid metropolitan area, as shown in Figure 6.

**Figure 6 – Cumulative distribution, Madrid, 2001**
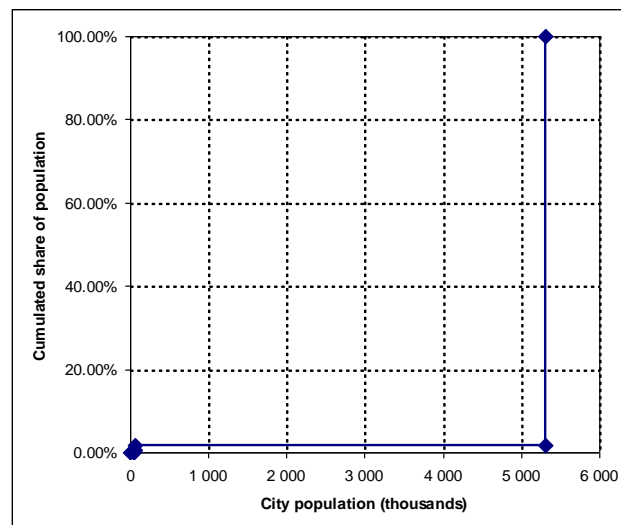
A key property of the index is that it correctly reflects the change in the potential for agglomeration economies of any reallocation of population. This property is close to the Pigou-Dalton transfer principle for measures of inequality, which states that any change in the distribution that unambiguously reduces inequality must be reflected in a decrease in its measure.

Let $\Delta x_i$ represent the change in the relative population size of urban area $i$. A reallocation of population is restricted by the condition that $\sum_{i=1}^{N} \Delta x_i = 0$. Any reallocation can be represented as a series of reallocations between two urban areas, and any reallocation between two urban areas can be represented as a series of reallocations between an urban area and the following or preceding one when urban areas are ordered according to size. Therefore, we need only to consider a reallocation of population from urban area $s-1$ to urban area $s$ (from an urban area to the next higher ranking one in terms of size):

$$\Delta x_s = -\Delta x_{s-1} > 0, \text{ and } \Delta x_i = 0 \text{ for } i \neq s, s-1 \tag{17}$$

According to our theoretical *a priori*, such a reallocation raises the potential for agglomeration economies. What effect does it have on the index?

Following equation [3], define[12]:

$$\Delta f_i = \frac{\Delta x_i}{\sum_{j=1}^{N} x_j} = \frac{\Delta x_i}{\sum_{j=0}^{N} x_j} \tag{18}$$

where, in view of [17],

$$\sum_{j=1}^{N} x_j - \Delta x_{s-1} + \Delta x_s = \sum_{j=1}^{N} x_j \tag{19}$$

and

$$\Delta f_s = -\Delta f_{s-1} > 0 \tag{20}$$

The value of the index after the reallocation is:

---

[12] The argument that follows can be generalized, albeit laboriously, to the version of the index that deals with the boundary problem. See Appendix III.

$$I' = \sum_{i \neq s-1, s} f_i x_i + \left(f_{s-1} - \Delta f_s\right)\left(x_{s-1} - \Delta x_s\right) + \left(f_s + \Delta f_s\right)\left(x_s + \Delta x_s\right) \quad [21]$$

$$I' = \sum_{i \neq s-1, s} f_i x_i + f_{s-1} x_{s-1} - f_{s-1}\Delta x_s - \Delta f_s x_{s-1} + \Delta f_s \Delta x_s$$
$$+ f_s x_s + f_s \Delta x_s + \Delta f_s x_s + \Delta f_s \Delta x_s \quad [22]$$

$$I' = \sum_{i \neq s-1, s} f_i x_i + f_{s-1} x_{s-1} + f_s x_s + \left(f_s - f_{s-1}\right)\Delta x_s + \Delta f_s\left(x_s - x_{s-1}\right) + 2\Delta f_s \Delta x_s \quad [23]$$

$$I' = \sum_{i=1}^{N} f_i x_i + \left(f_s - f_{s-1}\right)\Delta x_s + \Delta f_s\left(x_s - x_{s-1}\right) + 2\Delta f_s \Delta x_s \quad [24]$$

$$I' = I + \left(f_s - f_{s-1}\right)\Delta x_s + \Delta f_s\left(x_s - x_{s-1}\right) + 2\Delta f_s \Delta x_s \quad [25]$$

Given that the urban areas are ordered from the smallest to the largest, $x_S > x_{S-1}$, and $f_S > f_{S-1}$, so that $I' > I$.

Another interesting property of the proposed index is that it lends itself to territorial aggregation or disaggregation. Consider partitioning a territory in such a way that agglomeration $i$ belongs to subdivision $A$ if $i \in A$, and to subdivision $\overline{A}$ if $i \notin A$. Starting with formula [13], we have

$$I = \sum_{i \in A} f_i^2 \left(\sum_{j=0}^{N} x_j\right) + \sum_{i \notin A} f_i^2 \left(\sum_{j=0}^{N} x_j\right) \quad [26]$$

$$I = \sum_{i \in A} f_i^2 \left(\sum_{j \in A} x_j\right)\left(\frac{\sum_{j=0}^{N} x_j}{\sum_{j \in A} x_j}\right) + \sum_{i \notin A} f_i^2 \left(\sum_{j \notin A} x_j\right)\left(\frac{\sum_{j=0}^{N} x_j}{\sum_{j \notin A} x_j}\right) \quad [27]$$

Now let

$$F_A = \left(\frac{\sum_{j \in A} x_j}{\sum_{j=0}^{N} x_j}\right) \quad [28]$$

$$f_{i|A} = \frac{x_i}{\sum\limits_{j\in A} x_j} = f_i \frac{\sum\limits_{j=0}^{N} x_j}{\sum\limits_{j\in A} x_j} = \frac{f_i}{F_A} \qquad [29]$$

And we can rewrite [27] as

$$I = \sum_{i\in A}\left(f_{i|A}F_A\right)^2\left(\sum_{j\in A} x_j\right)\left(\frac{1}{F_A}\right) + \sum_{i\notin A}\left(f_{i|\overline{A}}(1-F_A)\right)^2\left(\sum_{j\notin A} x_j\right)\left(\frac{1}{1-F_A}\right) \qquad [30]$$

$$I = F_A I_A + (1-F_A)I_{\overline{A}} \qquad [31]$$

The index of a territory is a weighted average of the indexes of its subterritories, where the weights are given by population shares.

**Relationship with the Pareto distribution**

The empirically estimated exponent of the Pareto city-size distribution (a generalization of Zipf's rank-size rule) has been used as a measure of the concentration of an urban system (Rosen and Resnick, 1980). Following the notation established above, the (discrete) Pareto distribution can be written as:

$$N + 1 - i = Ax_i^{-a} \qquad [32]$$

where $N$ is the number of cities (ranked from the smallest to the largest), $x_i$ is the size of city $i$ [13], and $A$ and $a$ are parameters. Parameter $A$ can be calibrated from the size of the largest city:

$$N + 1 - N = 1 = Ax_N^{-a} \qquad [33]$$

$$A = x_N^{a} \qquad [34]$$

Inverting [22], we obtain:

$$x_i^{a} = \frac{A}{N+1-i} \qquad [35]$$

$$x_i = \left(\frac{A}{N+1-i}\right)^{1/a} \qquad [36]$$

Total urban population is:

---

[13] It makes little difference whether city sizes are absolute or relative to some benchmark, such as Madrid above.

$$X = \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} \left( \frac{A}{N+1-i} \right)^{1/a} \qquad [37]$$

And so it is quite straightforward to construct a cumulative distribution similar to the one in Figure 3 reflecting a theoretical Pareto distribution. It is then possible to apply our proposed index to a theoretical Pareto distribution using formula [12]:

$$I = \sum_{i=1}^{N} f_i x_i \qquad [12]$$

There results

$$I = \frac{\sum_{i=1}^{N} \left( \frac{A}{N+1-i} \right)^{2/a}}{\sum_{i=1}^{N} \left( \frac{A}{N+1-i} \right)^{1/a}} \qquad [38]$$

where we exploit the identity[14]:

$$I = \sum_{i=1}^{N} f_i x_i = \sum_{i=1}^{N} \left( \frac{x_i}{\sum_{j=1}^{N} x_j} \right) x_i = \frac{\sum_{i=1}^{N} x_i^2}{\sum_{j=1}^{N} x_j} \qquad [39]$$

If we assume that the number of cities $N$ and the size of the largest city $x_N$ are fixed, then, using [34], [38] can be written as[15]:

$$I = \frac{\sum_{i=1}^{N} \left( \frac{x_N^a}{N+1-i} \right)^{2/a}}{\sum_{i=1}^{N} \left( \frac{x_N^a}{N+1-i} \right)^{1/a}} \qquad [40]$$

---

[14] Here again, we ignore the boundary problem, which the Pareto distribution approach does not handle anyway.

[15] Note that the denominator of [40] is a CES aggregator function.

$$I = \frac{\sum_{i=1}^{N} x_N^{2} \left( \dfrac{1}{N+1-i} \right)^{2/a}}{\sum_{i=1}^{N} x_N \left( \dfrac{1}{N+1-i} \right)^{1/a}} \qquad [41]$$

$$I = x_N \frac{\sum_{i=1}^{N} \left( \dfrac{1}{N+1-i} \right)^{2/a}}{\sum_{i=1}^{N} \left( \dfrac{1}{N+1-i} \right)^{1/a}} \qquad [42]$$

It is shown in the Appendix IV that the derivative of the index relative to the Pareto parameter is

$$\frac{\partial I}{\partial a} = x_N \left( \frac{1}{a} \right)^{2} \frac{2 \sum_{j=1}^{N} \sum_{i=1}^{N} \left[ j^{-1/a} i^{-2/a} \ln i \right] - \sum_{j=1}^{N} \sum_{i=1}^{N} \left[ j^{-2/a} i^{-1/a} \ln i \right]}{\left( \sum_{i=1}^{N} i^{-1/a} \right)^{2}} \qquad [43]$$

The sign of that derivative is the sign of its numerator, but we could not determine that sign analytically. Using numerical simulations[16], we obtain that the derivative is negative for low values of $a$, and positive for high values. The sign reversal of the derivative is explained by the fact that, for a given number of cities, the size of the smallest city under the rank-size rule, $x_1 = x_N N^{-1/a}$, increases with $a$, leaving a larger gap to the left of the first point on the cumulative distribution (see Figure 3). Referring to index computation formula [5],

$$I = \sum_{i=1}^{N} \left( 1 - \sum_{j=0}^{i-1} f_j \right) \left( x_i - x_{i-1} \right) \qquad [5]$$

it is easily verified that its first term is equal to $x_1$. Indeed, our numerical simulations confirm that, if that first term is omitted, our index is a monotonically decreasing function of parameter $a$. This is illustrated in Figure 7.

---

[16] Spreadsheet calculations were performed for values of $N$ from 1 to 100, and $a$ from 0.01 to 2 in increments of .01.

**Figure 7 – Relationship of the proposed index to the Pareto elasticity parameter**

*N* = 5

*N* = 100
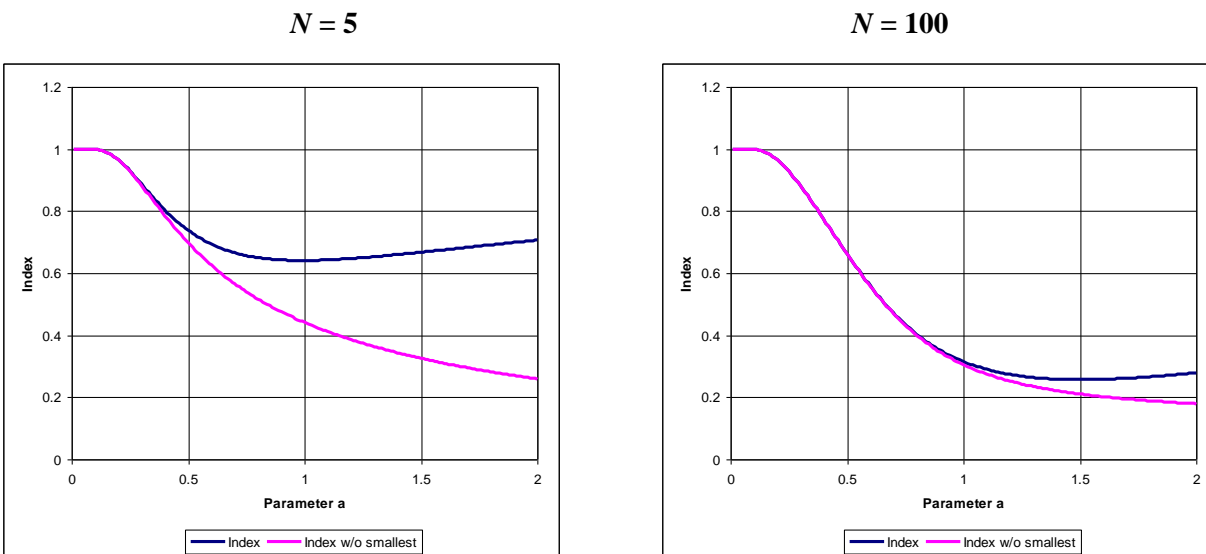


## EMPIRICAL APPLICATION

### The Spanish provinces

Administratively, Spain is divided into 8 105 municipalities that are aggregated into 52 provinces (NUTS III level) and seventeen Autonomous Communities or NUTS II regions (Figure 8 - Maps 8A and 8B). The number of municipalities within each province ranges from 34 (Las Palmas) to 371 (Burgos). Furthermore, there are Autonomous Communities with several provinces, for example, Andalusia with eight, and others with only one, like Asturias. For comparison with other European Union member-states, the seventeen Autonomous Communities can be aggregated into seven administrative regions or NUTS I regions (Figure 8, map 8C), which have no real internal political or administrative meaning.

**Figure 8 – Spanish administrative division of the territory into Provinces (NUTS III), Autonomous Communities (NUTS II) and NUTS I.**

Map 8A. Provinces or NUTS III regions*       Map 8B. Autonomous Communities or NUTS II regions       Map 8C. NUTS I regions



* Bold lines are provincial boundaries; fine lines are municipal boundaries.

Source: Own.

It is important to point out that we do not use municipalities as our territorial units of analysis (Figure 9 – map 9A). The municipality is an administrative division of the territory with no economic significance, because, in many cases, there is a high level of commuting between neighboring municipalities. An urban area could transcend municipal boundaries, to make up metropolitan areas, which might include several population nuclei surrounding a core one. To consider this fact, we aggregate the information offered by municipalities into LLMs. The regionalization method developed by Sforzi and Lorenzini (2002) and ISTAT (2006), applied to Spain by Boix and Galleto (2006), identifies the LLMs through a multi-stage process. Applying an algorithm that consists of four main stages and a fifth stage of fine-tuning, these authors aggregate the 8 106 Spanish municipalities into 806 LLMs. The algorithm starts with the municipal administrative unit and it generates the LLMs by using data of resident employed population, total employed population and displacements from the place of residence to the workplace, from the 2001 Spanish Population and Housing Census (INE)[17]. Map 9B shows the 806 LLMs defined by these authors[18].

---

[17] This is the most recent data as the 2011 Spanish Population and Housing Census (INE) is not yet available.

[18] More details and applications of LLMs in Fernandez and Rubiera (2012). Rubiera and Viñuela (2012) perform an economic evaluation of this spatial unit.

**Figure 9 – Spanish local areas: administrative basic unit (municipalities) and economic local unit (LLMs) (2001)**

Map 9A. Spanish division of the territory into 8,105 municipalities*

Map 9B. Spanish division of the territory into 806 LLMs



* Bold lines are provincial boundaries; fine lines are municipal boundaries.

Source: Own.

We have aggregated municipal employed population data from the 2001 Spanish Population and Housing Census (INE) to LLM data, following Boix and Galleto (2006). The LLM data are used to compute our index of potential agglomeration economies for each province. In the cases in which the LLM exceeds the limits of a province, we recalculate the part that belongs to each province, as proposed in Section 2.3[19]. Then, for each province, we calculate the cumulative distribution of population according to the relative LLM size, in terms of the Madrid LLM, the biggest in the country.

Figure 10 represents this cumulative distribution for the province of Asturias (being similar to Figure 3, but using real data and considering the LLM size as a percentage of that of Madrid). Next, the area above the curve of the cumulative distribution is calculated. In our example Asturias obtains the 82,6% level of the index.

---

[19] Nevertheless, in Appendix 3, we also present the results assuming that the population of these special cases belongs to the province in which is located the centroid and, we find minimum changes with respect to our approach.

**Figure 10 – Cumulative distribution of Asturias (2001).**
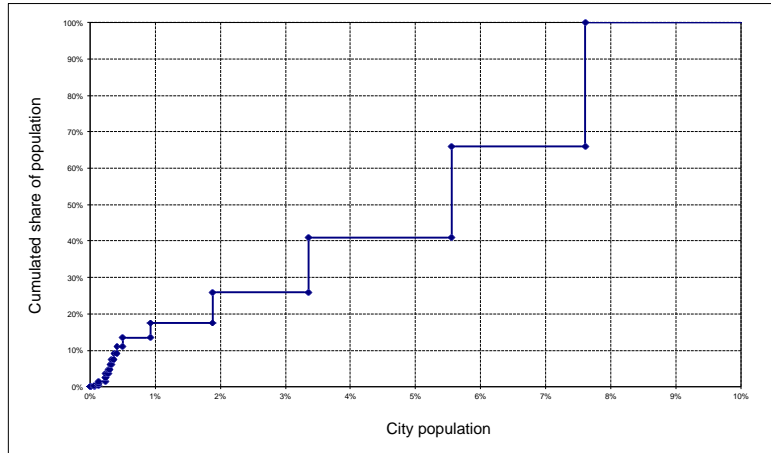


Source: Own based on INE (2001) data.

All the Spanish provinces are plotted in Figure 11, ranked by the value of the index. Madrid is followed by Barcelona and, at a much greater distance by Vizcaya, Valencia and Seville, which are among the biggest cities of the country. The rest of the distribution is intuitively reasonable, on the basis of our practical knowledge of the Spanish geographical economy.

**Figure 11 – Proposed index, ranked in decreasing value, Spain (2001).**
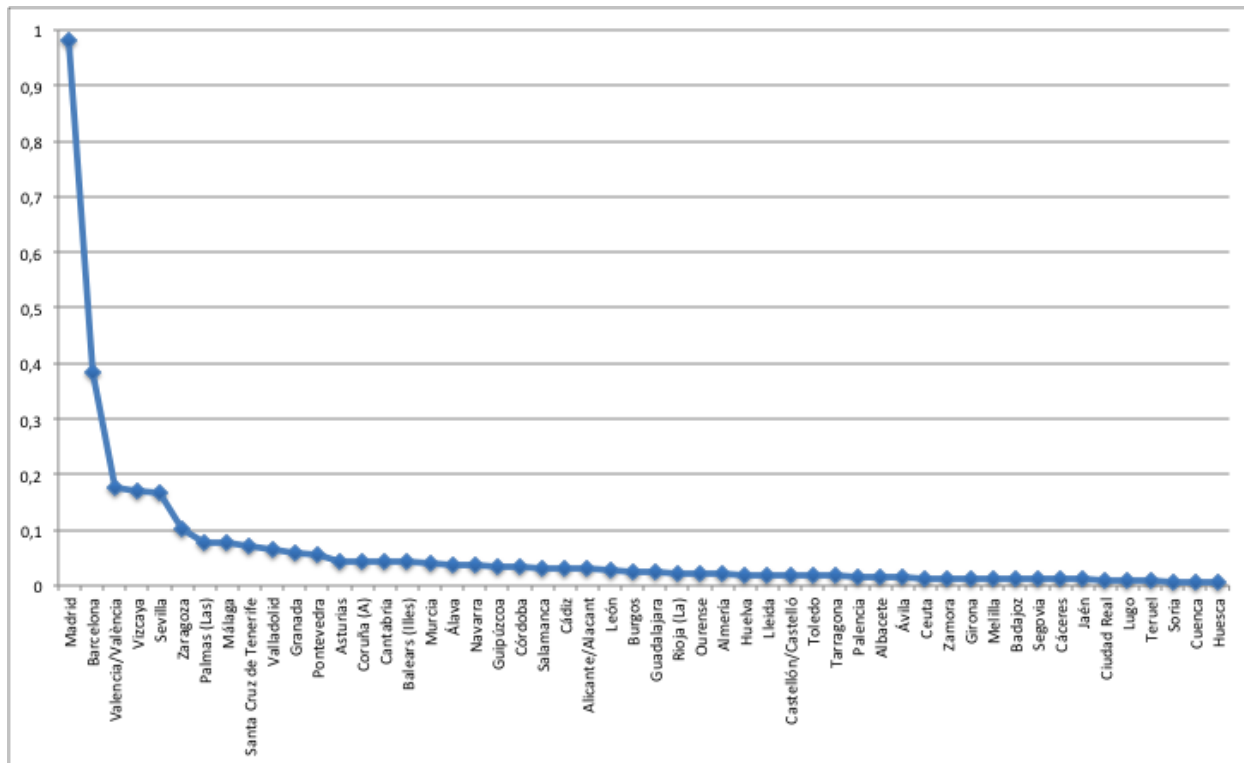


Source: Own based on INE (2001) data.

**Index performance evaluation**

Figure 12 plots the classical and HH measures of urbanization/concentration, the percentage of population living in cities of more than 50 000 inhabitants. Clearly, the values of our index (Figure 11) are very different and so, the ranking derived. Our index offers greater contrast, better reflecting the position of cities such as Barcelona or, less noticeably, Vizcaya, Valencia and Sevilla.

To illustrate how our index is able to better capture economic patterns, we correlate it with the *location quotients* of some of the activities known to be highly sensitive to agglomeration economies: high order producer services, also called knowledge intensive business services (henceforth KIBS). There are numerous empirical studies that confirm the relationship between agglomeration economies and this kind of services[20]. In this literature, the relationship between agglomeration economies and the concentration of these knowledge intensive services is clearly confirmed. Hence, we expect that the index which presents the highest correlation with the location quotients of these services is the one that better captures the potential for agglomeration economies.

The location quotient (*LQ*) that we use is the simplest one, defined as follows:

$$LQ_{ji} = \frac{\left( e_{ji}/e_i \right)}{\left( E_j/E \right)} \qquad [44]$$

where $LQ_{ji}$ is the location quotient of sector $x$ in area $a$; $e_{ji}$ is employment in sector $j$ in province $i$; $e_i = \sum_j e_{ji}$ is total employment in spatial unit $i$; $E_j = \sum_{i=1}^{n} e_{ji}$ is the total employment in sector $j$ in Spain (*n* is the number of spatial units: 52 provinces). Finally, $E = \sum_j E_j$ is the total employment in Spain.

Results of the correlations are computed for three indexes (the classical, the HH and our proposed one) and four examples of high order producer services: (i) financial services, (ii) computing and information technologies, (iii) advertising services and (iv) audiovisual and entertainment services. Nevertheless, the results are almost the same for all kinds of high order producer services. Our proposed index is presented in logarithms to obtain a better picture of the distribution among the provinces.

---

[20] The reasons for the concentration of such services in large metropolitan areas are strongly connected with the presence of different types of effects directly derived from the existence of agglomeration economies. The diversity and rapidly changing nature of talents and know-how mean that only the largest cities will provide the necessary specialized labor pool. Such industries are, in other words, dependent on a constant stream of face-to-face meetings with a wide (and changing) range of individuals that only can occur in cities, but better in large cities. See Daniels (1985), Illeris (1996), Shearmur and Doloreux (2008), Polèse *et al.* (2007) or Wernerheim and Sharpe (2003), among many others.

**Figure 12 – Classical and HH indexes ranked in decreasing order, Spain (2001).**
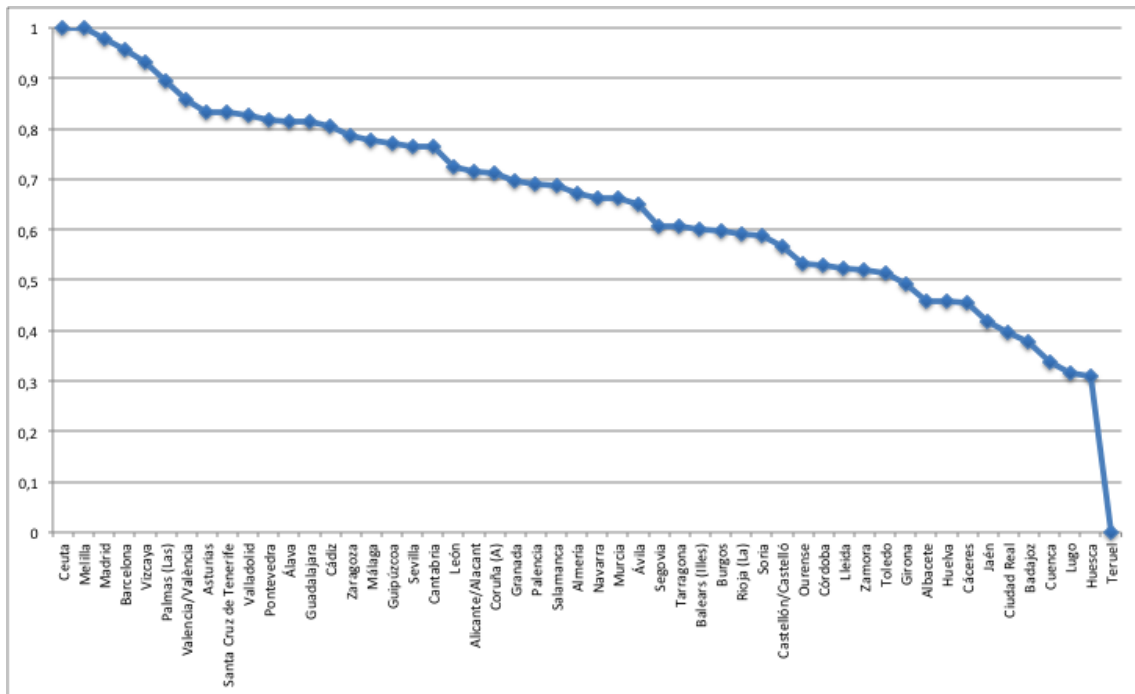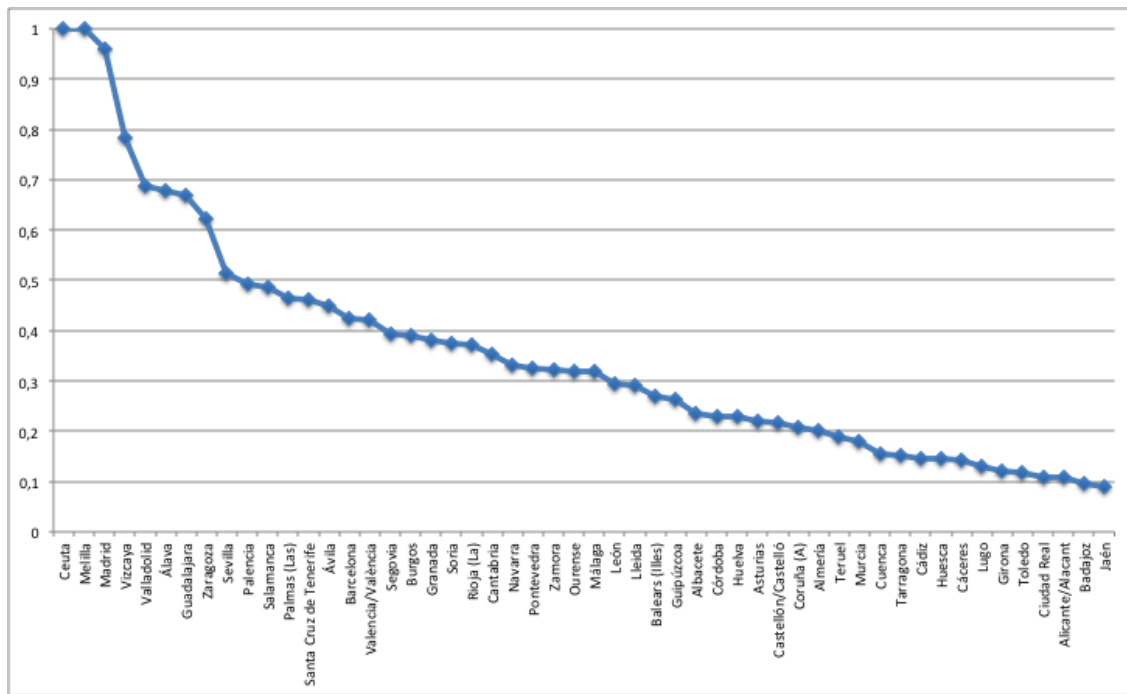
Figure 12a – Classical Index



Figure 12b – HH Index



Source: Own based on INE (2001) data.

For all these activities, the proposed index captures much better the effect of the main metropolitan areas of the country: Madrid, represented by the top right-hand point in the trend lines, and Barcelona, the point in the middle of the trend lines. For the rest of the cities, this index clearly identifies the relationship between urbanization and agglomeration better than others. Table 2 shows the correlations among our proposed index and the classical way of measuring urbanization and the HH concentration index (as examples of other common procedures to capture urbanization or agglomeration effects). In all the cases, the highest correlations are obtained with our index.

Figure 13 illustrates these results. As can be seen, the classical index, in many cases, presents too much dispersion (not to mention heteroskedasticity) but, essentially, the biggest mistakes appear in the higher values of the index, that is, in the provinces where the biggest cities are. It gives almost the same value to Madrid, Barcelona, Ceuta and Melilla21. It is not capable of distinguishing between the biggest cities, especially Madrid but also Barcelona, and other cities of substantial but smaller size, such as Bilbao. The HH index works similarly to the classical one. It makes significant errors in the middle of the distribution in which there are cases of important concentration without a clear process of agglomeration, and vice versa. It also presents the problem found in the classical index for the top of the distribution.

The correlation of KIBS with our proposed index let us to clearly observe how it is able to perform well not only in those provinces with large cities but also in all those with small or medium sized cities. Indeed, the relationship between the location of KIBS and our index takes a clear curvilinear shape.

**Table 2 - Correlations between the proposed, classical and HH indexes and KIBS (2001)**

|  | Telecommuni-cations | Financial services | Informatics and new technologies | Research and development | Advertising | Audiovisual and entertainment |
|---|---|---|---|---|---|---|
| **Proposed index** | 0.83 | 0.76 | 0.85 | 0.79 | 0.89 | 0.84 |
| **Classical index** | 0.43 | 0.29 | 0.49 | 0.65 | 0.50 | 0.55 |
| **HH Index** | 0.47 | 0.29 | 0.50 | 0.54 | 0.36 | 0.48 |

Source: Own.

---

[21] Two autonomous cities, located in the North of Africa, with a joint population of less than 140 000 inhabitants in 2001.

**Figure 13 – Correlations between the proposed, the classical and the HH indexes
and KIBS LQs for the Spanish provinces (2001)** [1]

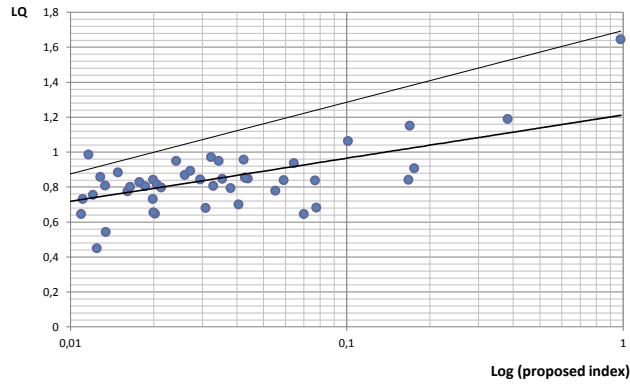Figure 13.1a – Correlation between the proposed index and financial services LQs

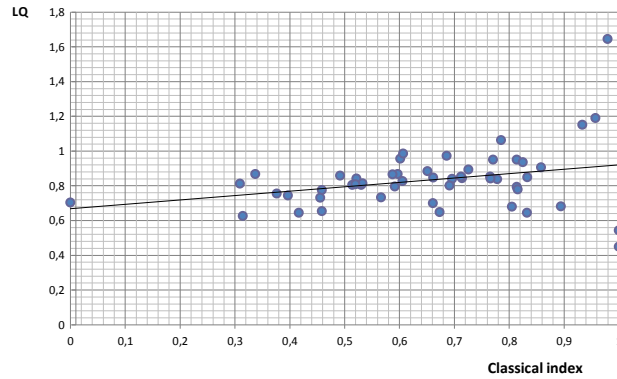Figure 13.2a – Correlation between the classical urbanization index and financial services LQs

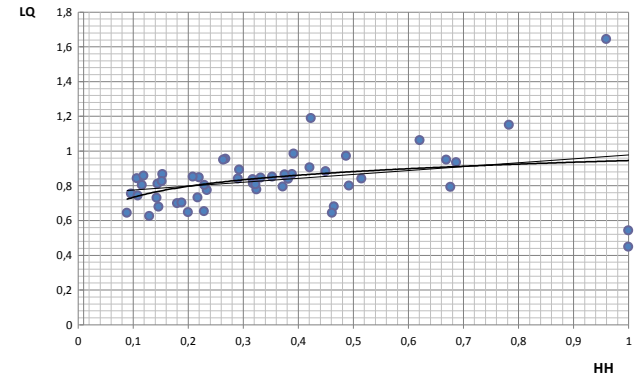Figure 13.3a – Correlation between the HH concentration index and financial services LQs



Figure 13.1b – Correlation between the proposed index and computing and information technologies LQs
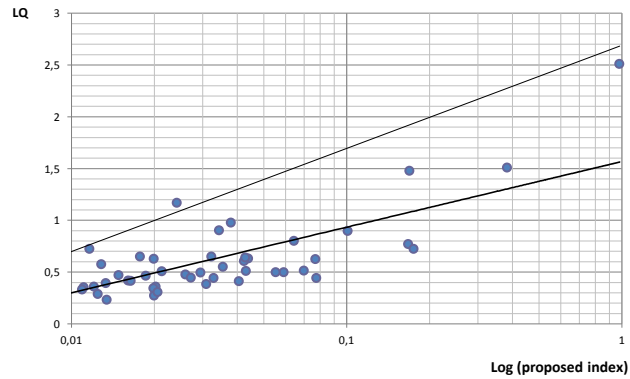
Figure 13.2b – Correlation between the classical urbanization index and computing and information technologies LQs
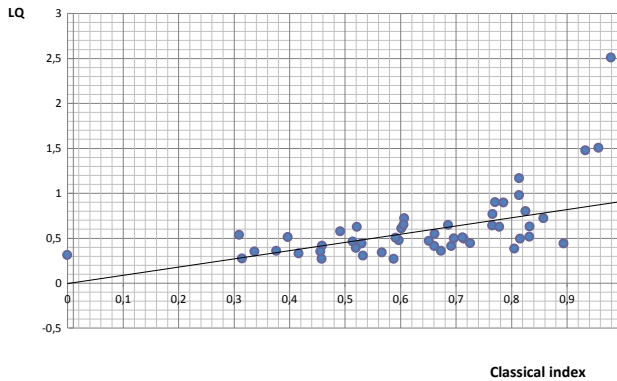
Figure 13.3b – Correlation between the HH concentration index and computing and information technologies LQs
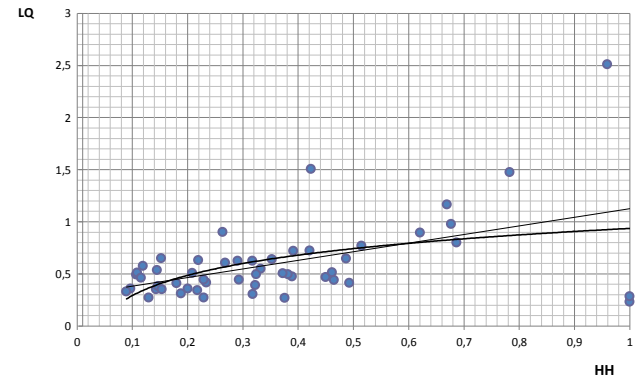
Figure 13.1c − Correlation between the proposed index and advertising services LQs
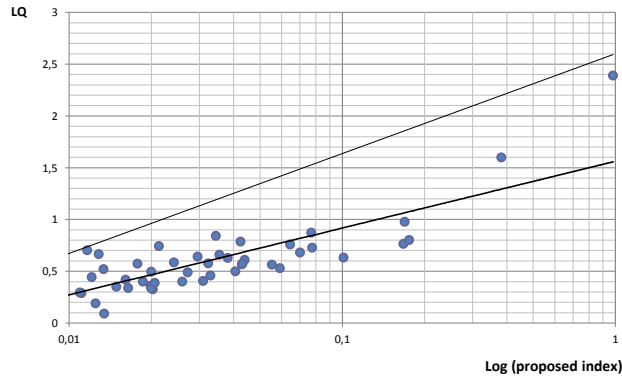


Figure 13.2c − Correlation between the classical urbanization index and advertising services LQs
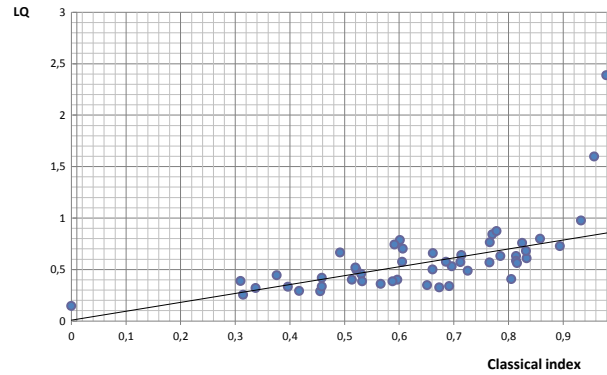


Figure 13.3c − Correlation between the HH concentrationn index and advertising services LQs



Figure 13.1d − Correlation between the proposed index and audiovisual and entertainment services LQs



Figure 13.2d − Correlation between the classical urbanization index and audiovisual and entertainment services LQs
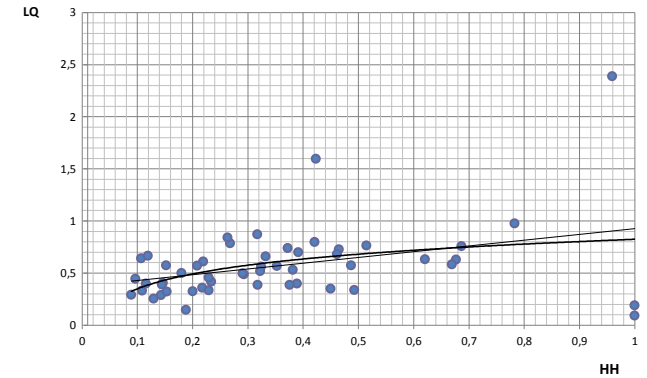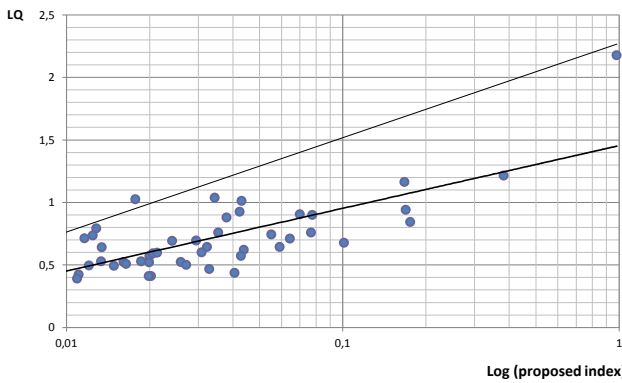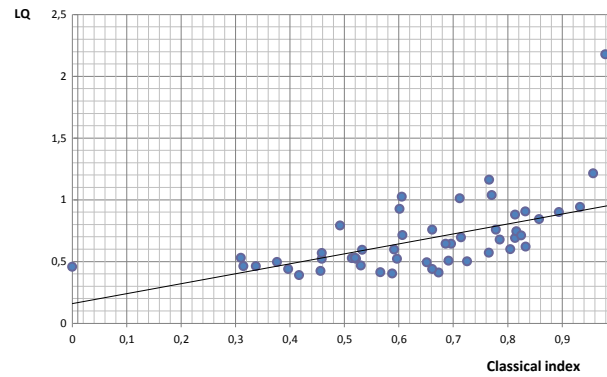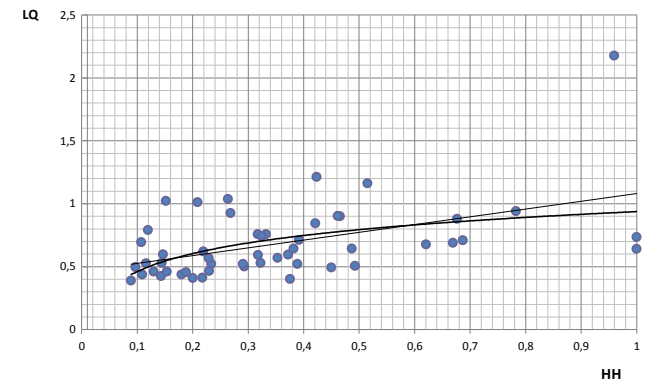


Figure 13.3d − Correlation between the HH concentration index and audiovisual and entertainment services LQs



Source: Own based on INE (2001) data.

[1] The proposed index is expressed in logarithms.

## SUMMARY AND CONCLUSIONS

Urbanization is, undoubtedly, a crucial phenomenon in the processes of economic and social development and industry specialization of a territory. It is commonplace to say that cities form and grow to exploit economies of agglomeration. Consequently, there is a need in empirical spatial analysis for some type of variable that informs about agglomeration economies. What is typically measured is agglomeration itself (for instance, when the unit of analysis is the individual urban area, city population, or employment). A measure of agglomeration is not a measure of agglomeration economies, but rather a measure of the potential for agglomeration economies. However, the usual ways of measuring agglomeration in a territory are not entirely satisfactory.

In this paper, we develop an index with several desirable properties. It has a graphical representation that is closely analogous to the development of the Gini coefficient from the Lorenz curve and can be formulated as an extension of the HH concentration index. It varies between 0 and 1, conforms to a version of the Pigou-Dalton transfer principle, and lends itself to territorial aggregation and disaggregation. It can be related to the elasticity parameter of the Pareto distribution, a generalization of Zipf's rank-size rule.

In our empirical illustration, we apply our index to the Spanish provinces (NUTS III regions) using local labor markets as the basic units for its construction, instead of administrative territorial divisions such as municipalities. Our index offers a much more credible and contrasted ranking than other classical ways of measuring the potential agglomeration economies of a territory. We look forward to applying our proposed index to other countries, such as the US and its counties.

The index could be very valuable for any descriptive or econometric analysis for a single country or for cross-country comparisons. Furthermore, it is a useful tool that could be used to enhance a more suitable design of economic policies.

# REFERENCES

ADELMAN, M. A. (1969) "Comment on the "H" Concentration Measure as a Numbers-Equivalent", *The Review of Economics and Statistics*, 51(1), pp. 99-101.

ADES, F. and E. L. GLAESER (1995) "Trade and circuses: explaining urban giants", *Quarterly Journal of Economics*, 110, pp. 195-227.

ARBIA, G. (1989) Spatial data configuration in statistical analysis of regional economic and related problems. Dordrecht, Netherlands: Kluwer.

ARRIAGA, E. E. (1970) "A new approach to the measurements of urbanization", *Economic Development and Cultural Change*, 18(2), pp. 206-218.

BEARDSELL, M. and V. HENDERSON (1999) "Spatial evolution of the computer industry in the USA", *European Economic Review*, 43, pp. 431-56.

BOIX, R. and V. GALLETO (2006) *Identificación de Sistemas Locales de Trabajo y Distritos Industriales en España*. Dirección General de Política de la Pequeña y Mediana Empresa, Ministerio de Industria, Comercio y Turismo.

BRÜLHART, M. and F. SBERGAMI (2009) "Agglomeration and growth: cross-country evidence", *Journal of Urban Economics*, 65, pp. 48-63.

CICCONE, A. and R. E. HALL (1976) "Productivity and the density of economic activity", *American Economic Review*, 86(1), pp. 54-70.

DANIELS, P. (1985) Service Industries: A geographical Perspective. New York, Methuen.

DURANTON, G. and D. PUGA (2002) "Diversity Specialization in Cities: Why, Where and Does it Matter", in P. MCCANN (Ed.): *Industrial Localization Economics*, pp. 151-186. Cheltenham-Edward Elgar.

FAY, M. and C. OPAL (2000) *Urbanization without growth: a not so uncommon phenomenon*. Working Paper No. 2412, The World Bank, Washington, DC.

FERNÁNDEZ, E. and F. RUBIERA (2012) Defining the Spatial Scale in Modern Economic Analysis: New Challenges from Data at Local Level. Advances in Spatial Science Series. Springer.

FOTHERINGHAM, A. S. and D. W. S. WONG (1991) "The modifiable areal unit problem in multivariate statistical-analysis", *Environment and Planning A,* 23, pp. 1025-1044.

GLAESER, E. L. (1994) "Cities, information, and economic growth", *Cityscape*, 1(1), pp. 9-77.

GLAESER, E.L. (1998) "Are Cities Dying?", *Journal of Economics Perspectives*, 12(2), pp. 139-160.

HALL, P. (2000) "Creative cities and economic development", *Urban Studies*, 37(4), pp. 639-649.

HENDERSON, V. (1988) *Urban Development: Theory, Fact and Illusion*. New York: Oxford University Press.

HENDERSON, V. (2003) "Marshall's scale economies", *Journal of Urban Economics*, 53, pp. 1-28.

INE (2001) *Censo de Población, 2001*. Instituto Nacional de Estadística (www.ine.es).

ILLERIS, S. (1996) *The service economy: A geographical approach*. Chichester, U.K., Wiley.

ISTAT (2006) Distretti Industriali e Sistemi Locali del Lavoro 2001. Collana Censimenti.

JACOBS, J. (1969) *The Economy of Cities*, Vintage, New York.

JACOBS, J. (1984) *Cities and the wealth of nations*, Vintage, New York.

JONES, B. and S. KONÉ (1996) "An exploration of relationships between urbanization and per capita income: United States and countries of the world", *Papers in Regional Science*, 75(2), pp. 135-153.

KRUGMAN, P.R. (1991) "Increasing Returns and Economic Geography", *Journal of Political Economy*, 99, pp. 483-499.

LEMELIN, A. and M. POLÈSE (1995) "What about the bell-shaped relationship between primacy and development?" *International Regional Science Review*, 18, pp. 313-330.

LONG, L. and A. NUCCI (1997) "The Hoover index of population concentration: a correction and update", *The Professional Geographer*, 49, pp. 431- 440.

OPENSHAW, S., and P. J. TAYLOR (1981) "The modifiable areal unit problem", in N. WRIGLEY and R. J. BENNETT (Eds.) *Quantitative geography*, pp. 60-70. London: Routledge.

PAELINCK, J. H. P (2000) "On aggregation in spatial econometric modeling". *Journal of Geographical Systems* 2, pp. 157–165.

PETERSEN, G. (1991) *Urban economies and national development*. Office of Housing and Urban Programs, USAID, Washington, DC.

POLÈSE M., R. SHEARMUR and F. RUBIERA (2007) "Observing regularities in location patters. An analysis of the spatial distribution of economic activity in Spain", *European Urban & Regional Studies*, 14(2), pp. 157-180.

PORTER, M. (1990) *The competitive advantage of nations*, Free Press, New York.

PRUD´HOMME, R. (1997) "Urban transportation and economic development", *Région et Développement*, 5, pp. 40–53.

QUIGLEY, J.M. (1998) "Urban Diversity and Economic Growth". *Journal of Economic Perspectives*, 12(2), pp. 127-138.

RICHARDSON S. (1992) "Statistical methods for geographical correlation studies". In: ELLIOT P, CUZICK J., ENGLISH D., STERN R. (Eds.) *Geographical and environmental epidemiology: methods for small area studies.* Oxford University Press, New York, pp. 181-204.

ROBINSON W.S. (1950) "Ecological correlations and the behaviour of individuals", *American. Sociological Review*, 15, pp. 351-357.

ROSEN, K. and M. RESNICK (1980) "The size distribution of cities: an examination of the Pareto Law and Primacy", *Journal of Urban Economics*, 8, pp. 165-186.

ROSENTHAL, S. S. and C. W. STRANGE (2001) "The determinants of agglomeration", *Journal of Urban Economics*, 50(2), pp. 191-229.

RUBIERA, F. and A. VIÑUELA (2012) "From funtional areas to analytical regions, where the agglomeration economies make sense", in FERNÁNDEZ, E. and F. RUBIERA, (Eds.) *Defining the Spatial Scale in Modern Economic Analysis: New Challenges from Data at Local Level.* Springer.

SFORZI, F. and F. LORENZINI (2002) "I distretti industriali", *L´esperienza Italiana dei Distretti Industriali*. Instituto per la Promozione Indistriale (IPI).

SHEARMUR, R. and D. DOLOREUX (2008) "Urban hierarchy or local milieu? High-order producer service and (or) knowledge-intensive business service location in Canada, 1991-2001". *Professional Geographer*, 60 (3), pp. 333-355.

UCHIDA, H. and A. NELSON (2010) "Agglomeration Index: Towards a New Measure of Urban Concentration", in J. BEALL, B. GUHA-KHASNOBIS, and R. KANBUR (Eds.) *Urbanization and Development*, Oxford University Press.

WERNERHEIM, M. and C. SHARPE (2003) "High-order producer services in Metropolitan Canada: Howe footloose are they?", *Regional Studies,* 37, pp. 469-90.

WHEATON, W. and H. SHISHIDO (1981) "Urban concentration, agglomeration economies, and the level of economic development", *Economic Development and Cultural Change*, 30, pp. 17-30.

WORLD BANK (1991) Urban policy and economic development: an agenda for the 1990s. The World Bank, Washington, DC.

# Appendix I – Computation of the index with information aggregated by size classes

Some of the main strengths of the index are that it is simple to compute it and has only modest data requirements. An additional advantage is that it is possible to obtain almost the same results with less detailed information. Many official statistical institutes offer the information aggregated by levels and, as we will show, our proposal presents almost the same quality results with detailed data than with data aggregated by levels.

Suppose that the data for the Spanish case are available only by levels like these ones:

− Population living in rural areas or in urban areas with a population of less than 50 000

− Population living in urban areas with a population of at least 50 000, but less than 100 000

− Population living in urban areas with a population of at least 100 000, but less than 200 000

− Population living in urban areas with a population of at least 200 000, but less than 400 000

− Population living in urban areas with a population of more than 400 000

− Population living in the largest city in the province

We use the province of Barcelona (Spain) to illustrate how our proposed index can be adapted to this way of presenting the information. The province has a total population of 4 636 632 inhabitants according to the 2001 Spanish Census. The main city is Barcelona with a population of 3 136 758. Table A1.1 shows the distribution of the population of the province presented by the levels previously proposed.

**Table A1.1 – Distribution of population by city size in the province of Barcelona (2001).**

|  | Population | % Total population | Cumulative distribution | % Cumulative distribution | "Official" division (thousands) |
|---|---|---|---|---|---|
| **<50 000** | 190 326 | 4.0% | 190 326 | 4.0% | 50 |
| **50 000-100 000** | 161 793 | 3.4% | 352 119 | 7.3% | 100 |
| **100 000-200 000** | 366 084 | 7.6% | 718 203 | 14.9% | 200 |
| **200 000-400 000** | 476 033 | 9.9% | 1 194 236 | 24.8% | 400 |
| **>400 000 (*)** | 564 735 | 11.8% | 1 758 971 | 36.6% | 565 |
| **Largest city (Barcelona)** | 3 046 956 | 63.4% | 4 805 927 | 100.0% | 3 047 |
| **Total** | 4 805 927 |  |  |  |  |

(*) This consists of population living in *urban* areas with at least 400 000 inhabitants minus the population living in the largest city (Barcelona).

Source: own from INE (2001).

To calculate our index using this less detailed information, the first step is to obtain a cumulative distribution graph approximating the continuous distribution by using the category boundaries of the official statistics, as in the last column of Table A1.1. Usually, official statistics divide the distribution of population into discrete categories up to some maximum size (400 000 inhabitants in our simulations), beyond which the last category has no upper bound. Here, we take advantage of the fact that the size of the largest city is usually known. First, we subtract the population of the largest city (Barcelona) from the population in cities of the same size-class (here, 400 000 inhabitants or more). Then, we set the boundary of that size-class as the minimum of two values: (i) the total population in that size-class minus that of the largest city or (ii) the population of the largest city. This rule derives from the fact that, by definition, no other city can be larger than the largest, and that the second largest city cannot be larger than the total population in the category of more than 400 000 inhabitants minus the largest city. In Table A1.1, for the province of Barcelona, the population in the 400 000 plus size-class is 3 611 691; subtract the population of Barcelona (3 046 956), and the remainder (564 735) is the upper limit of the size-class. The resulting graph for Barcelona is displayed in Figure A1.1.

**Figure A1.1. - Cumulative distribution with city sizes on the X-axis, province of Barcelona (2001).**
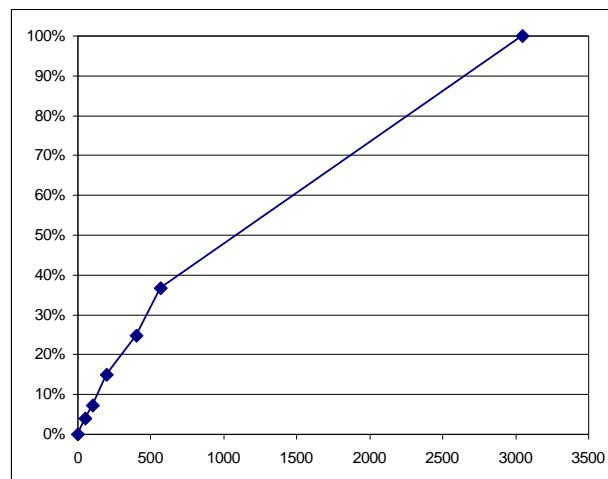


Source: Own based on INE (2001) data.

At that stage, the area above the curve is easily computed as a sum of trapezoidal areas, as illustrated in Figure A1.2, where data for the province of Asturias, 2001, is used for better readability. The area above the curve is the sum of trapezoids ABJK, BCIJ, CDHI, DEGH, and of triangle EFG.

**Table A1.2 - Distribution of population by city size in the province of Asturias (2001).**

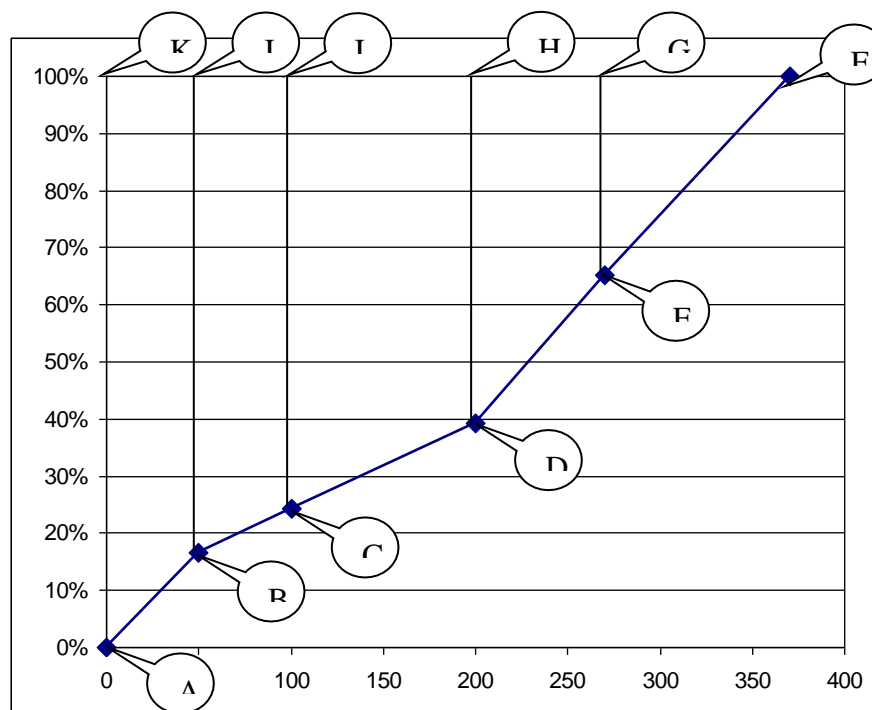| | Population | % Total population | Cumulative distribution | % Cumulative distribution | "Official" division (thousands) |
|---|---|---|---|---|---|
| **<50 000** | 177 250 | 16.7% | 177 250 | 16.7% | 50 |
| **50 000-100 000** | 81 340 | 7.7% | 258 590 | 24.3% | 100 |
| **100 000-200 000** | 157 948 | 14.9% | 416 538 | 39.2% | 200 |
| **200 000-400 000 (\*)** | 276 983 | 26.1% | 693 521 | 65.2% | 277 |
| **Largest city (Oviedo)** | 369 477 | 34.8% | 1 062 998 | 100.0% | 369 |
| **Total** | 1 062 998 | | | | |

(\*) This consists of population living in *urban* areas with at least 200 000 inhabitants minus the population living in the largest city (Oviedo).

Source: own from INE (2001).

**Fig. A1.2 – Cumulative distribution by city-size class, Asturias (2001)**



Let $k_i$ be the upper limit of size-class $i$ (by convention, $k_0 = 0$). Let $f_i$ be the fraction of population residing

in the $i^{\text{th}}$ size-class, and let $F_i = \sum_{j=1}^{i} f_j$ be the cumulative distribution, with $F_0 = f_0 = 0$. The area above

the curve is

$$A = \sum_{i=1}^{N} \frac{(1 - F_i) + (1 - F_{i-1})}{2} (k_i - k_{i-1})$$ [A1.01]

which is quite close to equation [5] in the main text. Develop

$$A = \sum_{i=1}^{N} \frac{\left(1 - \sum_{j=0}^{i} f_j\right) + \left(1 - \sum_{j=0}^{i-1} f_j\right)}{2} \left(k_i - k_{i-1}\right) \qquad [A1.02]$$

$$A = \sum_{i=1}^{N} \frac{2 - 2\sum_{j=0}^{i-1} f_j - f_i}{2} \left(k_i - k_{i-1}\right) \qquad [A1.03]$$

$$A = \sum_{i=1}^{N} \left(k_i - k_{i-1}\right) - \sum_{i=1}^{N} \left(\sum_{j=0}^{i-1} f_j\right)\left(k_i - k_{i-1}\right) - \sum_{i=1}^{N} \frac{f_i}{2}\left(k_i - k_{i-1}\right) \qquad [A1.04]$$

$$A = \sum_{i=1}^{N} k_i - \sum_{i=1}^{N} k_{i-1} - \sum_{i=1}^{N}\left(\sum_{j=0}^{i-1} f_j\right) k_i + \sum_{i=1}^{N}\left(\sum_{j=0}^{i-1} f_j\right) k_{i-1} - \sum_{i=1}^{N} \frac{f_i}{2} k_i + \sum_{i=1}^{N} \frac{f_i}{2} k_{i-1} \qquad [A1.05]$$

$$A = \sum_{i=1}^{N} k_i - \sum_{i=1}^{N-1} k_i - \sum_{i=1}^{N}\left(\sum_{j=0}^{i-1} f_j\right) k_i + \sum_{i=1}^{N-1}\left(\sum_{j=0}^{i} f_j\right) k_i - \sum_{i=1}^{N} \frac{f_i}{2} k_i + \sum_{i=1}^{N} \frac{f_i}{2} k_{i-1} \qquad [A1.06]$$

$$A = k_N - \left(\sum_{j=0}^{N-1} f_j\right) k_N - \sum_{i=1}^{N-1}\left(\sum_{j=0}^{i-1} f_j\right) k_i + \sum_{i=1}^{N-1}\left(\sum_{j=0}^{i} f_j\right) k_i$$
$$- \frac{f_N}{2} k_N - \sum_{i=1}^{N-1} \frac{f_i}{2} k_i + \sum_{i=1}^{N} \frac{f_i}{2} k_{i-1} \qquad [A1.07]$$

$$A = f_N k_N - \sum_{i=1}^{N-1}\left[\left(\sum_{j=0}^{i-1} f_j\right) - \left(\sum_{j=0}^{i} f_j\right)\right] k_i - \sum_{i=1}^{N} \frac{f_i}{2} k_i + \sum_{i=1}^{N} \frac{f_i}{2} k_{i-1} \qquad [A1.08]$$

$$A = \sum_{i=1}^{N} f_i k_i - \sum_{i=1}^{N} \frac{f_i}{2} k_i + \sum_{i=1}^{N} \frac{f_i}{2} k_{i-1} \qquad [A1.09]$$

$$A = \sum_{i=1}^{N} \frac{f_i}{2} k_i + \sum_{i=1}^{N} \frac{f_i}{2} k_{i-1} \qquad [A1.10]$$

$$A = \sum_{i=1}^{N} f_i \frac{k_i + k_{i-1}}{2} \qquad [A1.11]$$

Formula [A1.11] is closely analogous to formula [12] in the main text. But here, rather than actual urban area sizes, each size class is represented by the mid-point of its size-interval. Moreover, this calculation is almost identical to Arriaga's (1970, p. 213) proposed $U_b$ index for cases where there is no information on individual cities but the urban population has been classified by city-size categories. There are three differences. The first is that we define the sizes of urban areas in relative terms, as a fraction of the

population of the largest city in the universe under consideration, in order to obtain an index that varies between zero and one. Second, Arriaga's formula excludes the first term of [A1.11], equal to $f_1 k_1/2$; it applies to truncated data, where there is a size-threshold below which areas are not considered « urban ». In our view, it can be argued that there is some potential for agglomeration economies even in small villages (see above, after equation [4]). Finally, Arriaga does not assume that the size of the largest city is known.

However, the cumulative distributions in Figures A1.1 and A1.2 are not very good approximations of continuous distributions because there can be no cities in the interval between the size of the largest city and the next highest point on the X-axis. In some cases, such as the province of Barcelona, that interval can be very broad and, therefore, the representation of the cumulative distribution in Figures A1.1 and A1.2 above may be misleading. Introducing an extra point in the curve allows us to obtain a more accurate representation. This is shown in Figure A1.3a. Note that when all the city sizes are known, the whole cumulative distribution takes the form of a broken line with stepwise increases, just like the right tail of the graph in Figure A1.3a (see Figure 3 in the main text). Moreover, in some cases, such as the province of Barcelona, the size of the second largest city can be deduced. Indeed, there can only be one city with a population of at least 400 000 in the interval between 400 000 and 496,329. In those cases, it would be possible to replace the diagonal between points A and B in Figure A1.3a by a right-angled step. But such deductions are not so easy to automate, and ignoring them is less misleading than ignoring the absence of cities between the largest and the second largest. Finally, in order to obtain convenient measure which varies between zero and 1, or 100%, we apply the same normalization rule as before: we rescale the horizontal axis of the cumulative distribution in terms of the size of the largest city in the urban system under study (in our application, Madrid). With this transformation, the cumulative distribution for Barcelona takes the form of Figure A1.3b.

**Figure A1.3 – Cumulative distribution with no city in the last interval, province of Barcelona (2001).**

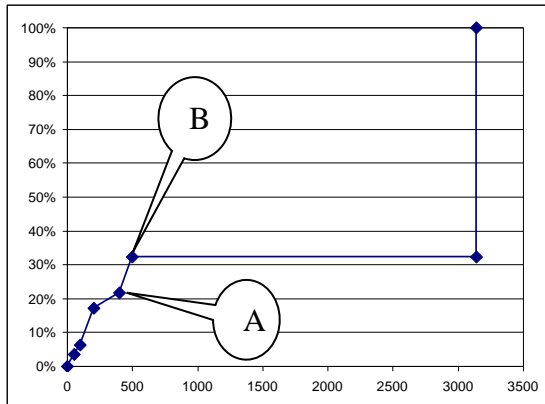Figure A1.3a – Cumulative distribution

Figure A1.3b – Cumulative distribution relative to the largest city (Madrid)
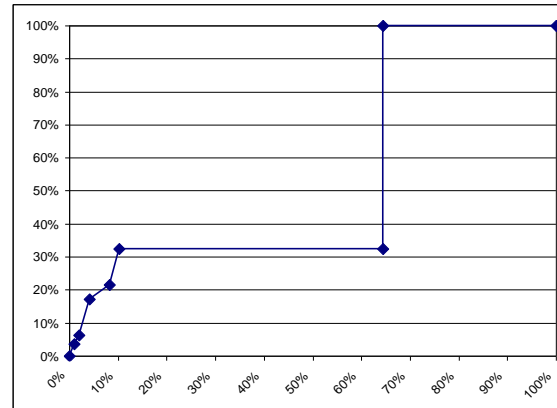


Source: Own based on INE (2001) data.

In Figure A1.4, we compare the results for the index using discrete data with those of the index using continuous data presented in the main text. The values for Madrid (0.98217 with the discrete data, and 0.98215 with the continuous data) and Barcelona (0.3840 and 0.3843) are beyond the boundaries of the graph. The differences are minimal. Indeed, the correlation between them is 0.99998 and the average absolute difference is 0.00077.

**Figure A1.4 – Comparison of the proposed index computed with discrete, and continuous data, Spain (2001).**
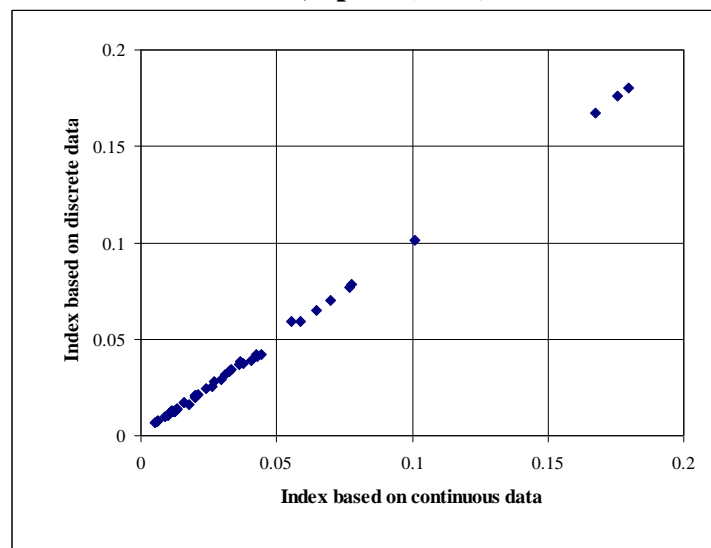


Source: Own based on INE (2001) data.

# Appendix II – Comparison of two solutions to the boundary problem

In the main text, we have computed our index for the Spanish provinces taking into account the fact that some LLMs include municipalities that are located in a different province. Another way of dealing with this problem, as we have mentioned, is to allocate each LLM as a whole to the province where its centroid is located. In effect, this amounts to redrawing provincial boundaries. In this Appendix, we compare the two approaches and find that, at least in the case of Spain, the results are very similar. This is encouraging, given that the more accurate approach described in Section 2.3 requires more detailed data to be applied. But, of course, what we found for LLMs in Spanish provinces may not be true in all applications.

In Figure A2.1, we compare the results for the index allocating each LLM as a whole to the province where its centroid is located with the index developed in the main text. The values for Madrid (0.9822 for the LLM centroid-based index, and 0.9805 for the index based on the location of individual municipalities) and Barcelona (0.3843 and 0.3830) are beyond the boundaries of the graph. The differences are minimal. Indeed, the correlation between them is 0.9999 and the average absolute difference is 0.00077.

**Figure A2.1 – Comparison between an index allocating each LLM as a whole to the province where its centroid is located and an index allocating every municipality of each LLM to the province where it is located, Spain (2001).**
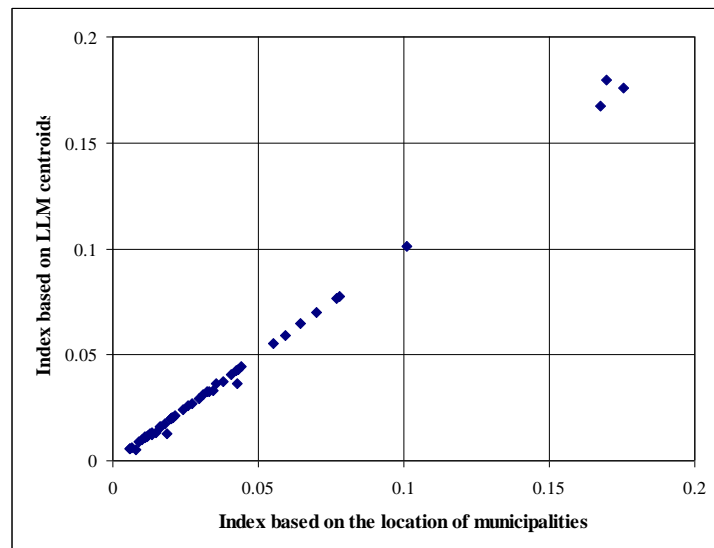


Source: Own based on INE (2001) data.

# Appendix III – The Pigou-Dalton principle with the boundary problem

We define an urban area *fraction* as the part of a given urban area which belongs or does not belong to the territory under consideration. Let $x_{i,1}$ be the population in urban area $i$ which actually resides in the territory under consideration, and $x_{i,0}$ be the population in urban area $i$ which resides outside of the territory under consideration, in the neighboring province. Naturally,

$$x_i = x_{i,1} + x_{i,0} \qquad\qquad [\text{A3.01}]$$

Then equation [3] must be rewritten as

$$f_i = \frac{x_{i,1}}{\displaystyle\sum_{j=1}^{N} x_{j,1}} = \frac{x_{i,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1}} \qquad\qquad [\text{A3.02}]$$

Let $\Delta x_{i,k}$ represent the change in the relative size of the population in urban area fraction $i,k$. A reallocation

of population is restricted by the condition that $\displaystyle\sum_{k=0}^{1} \sum_{j=0}^{N} \Delta x_{j,k} = \sum_{j=0}^{N} \Delta x_j = 0$. Any reallocation can be

represented as a series of reallocations between two urban area fractions. Furthermore, any reallocation between two urban area fractions $i,k$ and $j,h$ can be represented as a reallocation between urban area fractions $i,k$ and $j,k$, followed by a reallocation between urban area fractions $j,k$ and $j,h$. Finally, any reallocation between the corresponding fractions of two urban areas (such as between $i,k$ and $j,k$) can be represented by a series of reallocations between the corresponding fractions of an urban area and the following or preceding one when urban areas are ordered according to size[22]. Therefore, we need only to consider three cases:

Case A: a reallocation of population from the fraction of urban area $s-1$ which is included in the territory, to the corresponding fraction of urban area $s$ (from an urban area to the next higher ranking one in terms of size):

4) $\Delta x_{s,1} = -\Delta x_{s-1,1} > 0$, and $\Delta x_{i,k} = 0$ for $[i,k] \notin \{[s,1],[s-1,1]\}$      [A3.03]

Case B: a reallocation of population from the fraction of urban area $s-1$ which is outside the territory, to the corresponding fraction of urban area $s$ (from an urban area to the next higher ranking one in terms of size):

---

[22] The order is not necessarily the same for the series of urban area fractions belonging to the territory or for the series of fractions outside the territory.

$$\Delta x_{s,0} = -\Delta x_{s-1,0} > 0, \text{ and } \Delta x_{i,k} = 0 \text{ for } [i,k] \notin \{[s,0],[s-1,0]\} \tag{A3.04}$$

Case C: a reallocation of population from the fraction of urban area $s$ which is included in the territory, to the fraction of urban area $s$ which is outside the territory:

$$\Delta x_{s,0} = -\Delta x_{s,1} > 0, \text{ and } \Delta x_{i,k} = 0 \text{ for } i \neq s \tag{A3.05}$$

**CASE C**

Case C is the simplest. In view of

$$\Delta x_{s,0} = -\Delta x_{s,1} > 0, \text{ and } \Delta x_{i,k} = 0 \text{ for } i \neq s \tag{A3.05}$$

let

$$x'_{i,1} = \begin{cases} x_{s,1} + \Delta x_{s,1}, \text{ for } i = s \\ x_{i,1}, \text{ otherwise} \end{cases}, \tag{A3.06}$$

and

$$f'_i = \frac{x'_{i,1}}{\displaystyle\sum_{j=0}^{N} x'_{j,1}} = \frac{x'_{i,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} \tag{A3.07}$$

$$f'_i = \begin{cases} \dfrac{x_{s,1} + \Delta x_{s,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} = f_s \dfrac{\displaystyle\sum_{j=0}^{N} x_{j,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} + \dfrac{\Delta x_{s,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}}, \text{ for } i = s \\[4ex] \dfrac{x_{i,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} = f_i \dfrac{\displaystyle\sum_{j=0}^{N} x_{j,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}}, \text{ otherwise} \end{cases} \tag{A3.08}$$

We have

$$I' = \sum_{i=1}^{N} f'_i x_i \tag{A3.09}$$

$$I' = \sum_{i \neq s} f_i x_i \frac{\sum_{j=0}^{N} x_{j,1}}{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}}$$
$$+ \left( f_s \frac{\sum_{j=0}^{N} x_{j,1}}{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} + \frac{\Delta x_{s,1}}{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} \right) x_s \qquad \text{[A3.10]}$$

since [A3.05] implies $\Delta x_s = 0$.

$$I' = \sum_{i=1}^{N} f_i x_i \frac{\sum_{j=0}^{N} x_{j,1}}{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} + x_s \frac{\Delta x_{s,1}}{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} \qquad \text{[A3.11]}$$

$$I' = I \frac{\sum_{j=0}^{N} x_{j,1}}{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} + x_s \frac{\Delta x_{s,1}}{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} \qquad \text{[A3.12]}$$

If $x_s = I$, , then

$$I' = I \frac{\sum_{j=0}^{N} x_{j,1}}{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} + I \frac{\Delta x_{s,1}}{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} = I \frac{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}}{\sum_{j=0}^{N} x_{j,1} + \Delta x_{s,1}} = I \qquad \text{[A3.13]}$$

It follows, by construction [A3.05]

$$\Delta x_{s,0} = -\Delta x_{s,1} > 0, \text{ and } \Delta x_{i,k} = 0 \text{ for } i \neq s \qquad \text{[A3.14]}$$

that $I' < \text{or} > I$ according to whether $x_s > \text{or} < I$. In words, if there is a population shift out of the territory under consideration within a given urban area, the index will decrease or increase according to whether the shift occurs in an urban area that is larger or smaller than the mathematical expectation of the urban area size as given by our index $I$. This is perfectly consistent with the Pigou-Dalton principle.

## CASE B

Case B is quite straightforward too. In view of

$$\Delta x_{s,0} = -\Delta x_{s-1,0} > 0, \text{ and } \Delta x_{i,k} = 0 \text{ for } [i,k] \notin \{[s,0],[s-1,0]\} \tag{A3.04}$$

the distribution $\{f_i\}$ remains unchanged. Let

$$x_i' = \begin{cases} x_s + \Delta x_{s,0}, \text{ for } i = s \\ x_s - \Delta x_{s,0}, \text{ for } i = s-1, \\ x_i, \text{ otherwise} \end{cases} \tag{A3.15}$$

We have

$$I' = \sum_{i=1}^{N} f_i x_i' \tag{A3.16}$$

$$I' = \sum_{i \neq s, s-1} f_i x_i + f_s \left( x_s + \Delta x_{s,0} \right) + f_{s-1} \left( x_{s-1} - \Delta x_{s,0} \right) \tag{A3.17}$$

$$I' = \sum_{i=1}^{N} f_i x_i + f_s \Delta x_{s,0} - f_{s-1} \Delta x_{s,0} \tag{A3.18}$$

$$I' = I + \left( f_s - f_{s-1} \right) \Delta x_{s,0} \tag{A3.19}$$

If $\Delta x_{s,0} > 0$, then $I' < \text{or} > I$ according to whether $f_s < \text{or} > f_{s-1}$[23]. Again, this is consistent with the Pigou-Dalton principle. Given that the distribution $\{f_i\}$ remains unchanged, it follows that a reallocation of population that takes place outside the territory increases the potential for agglomeration economies if the shift is towards an urban area where there is a greater fraction of the territory's population. That is because the gain or loss of potential for each individual resident is assumed to be proportional to the change in the size of the urban area, whatever its size. So a given change increases the territory's potential if the number of people whose individual potential rises is greater than the number of people for which it falls.

### CASE A

Case A would appear to be similar to the case where there is no boundary problem. But it is more complicated, because the ranking of urban areas when they are ordered according to size is likely to be different from their ranking when they are ranked according to the fraction of the territory's population they contain. For example, nearly 22% of the population in the province of Toledo live in the Toledo urban area (LLM), and less than 1% live in the Madrid urban area; But Madrid is a much larger city than Toledo.

---

[23] Recall the ranking of urban areas when they are ordered according to size is likely to be different from their ranking when they are ranked according to the fraction of the territory's population they contain.

In view of

$$\Delta x_{s,1} = -\Delta x_{s-1,1} > 0, \text{ and } \Delta x_{i,k} = 0 \text{ for } [i,k] \notin \{[s,1],[s-1,1]\} \quad \text{[A3.03]}$$

let

$$x'_{i,1} = \begin{cases} x_{s,1} + \Delta x_{s,1}, & \text{for } i = s \\ x_{s-1,1} - \Delta x_{s,1}, & \text{for } i = s-1, \\ x_{i,1}, & \text{otherwise} \end{cases} \quad \text{[A3.20]}$$

and

$$x'_{i} = \begin{cases} x_{s} + \Delta x_{s,1}, & \text{for } i = s \\ x_{s-1} - \Delta x_{s,1}, & \text{for } i = s-1, \\ x_{i}, & \text{otherwise} \end{cases} \quad \text{[A3.21]}$$

Note that $\displaystyle\sum_{j=0}^{N} x'_{j,1} = \sum_{j=0}^{N} x_{j,1}$ and $\displaystyle\sum_{j=0}^{N} x'_{j} = \sum_{j=0}^{N} x_{j}$

Let

$$f'_{i} = \frac{x'_{i,1}}{\displaystyle\sum_{j=0}^{N} x'_{j,1}} = \frac{x'_{i,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1}} = \begin{cases} \dfrac{x_{s,1} + \Delta x_{s,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1}} = f_{s} + \dfrac{\Delta x_{s,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1}}, & \text{for } i = s \\[4mm] \dfrac{x_{s-1,1} - \Delta x_{s,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1}} = f_{s-1} - \dfrac{\Delta x_{s,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1}}, & \text{for } i = s-1 \\[4mm] \dfrac{x_{i,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1}} = f_{i}, & \text{otherwise} \end{cases} \quad \text{[A3.22]}$$

We have

$$I' = \sum_{i=1}^{N} f'_{i} x'_{i} \quad \text{[A3.23]}$$

$$I' = \sum_{i \neq s, s-1} f_{i} x_{i} + \left( f_{s} + \frac{\Delta x_{s,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1}} \right)(x_{s} + \Delta x_{s,1}) + \left( f_{s-1} - \frac{\Delta x_{s,1}}{\displaystyle\sum_{j=0}^{N} x_{j,1}} \right)(x_{s-1} - \Delta x_{s,1}) \quad \text{[A3.24]}$$

$$I' = \sum_{i \neq s, s-1} f_i x_i + f_s x_s + f_s \Delta x_{s,1} + \left( \frac{\frac{\Delta x_{s,1}}{N}}{\sum_{j=0}^{N} x_{j,1}} \right) x_s + \left( \frac{\frac{\Delta x_{s,1}}{N}}{\sum_{j=0}^{N} x_{j,1}} \right) \Delta x_{s,1}$$

$$+ f_{s-1} x_{s-1} - f_{s-1} \Delta x_{s,1} - \left( \frac{\frac{\Delta x_{s,1}}{N}}{\sum_{j=0}^{N} x_{j,1}} \right) x_{s-1} + \left( \frac{\frac{\Delta x_{s,1}}{N}}{\sum_{j=0}^{N} x_{j,1}} \right) \Delta x_{s,1}$$

[A3.25]

$$I' = \sum_{i=1}^{N} f_i x_i + f_s \Delta x_{s,1} + \left( \frac{\frac{\Delta x_{s,1}}{N}}{\sum_{j=0}^{N} x_{j,1}} \right) x_s + \left( \frac{\frac{\Delta x_{s,1}}{N}}{\sum_{j=0}^{N} x_{j,1}} \right) \Delta x_{s,1}$$

$$- f_{s-1} \Delta x_{s,1} - \left( \frac{\frac{\Delta x_{s,1}}{N}}{\sum_{j=0}^{N} x_{j,1}} \right) x_{s-1} + \left( \frac{\frac{\Delta x_{s,1}}{N}}{\sum_{j=0}^{N} x_{j,1}} \right) \Delta x_{s,1}$$

[A3.26]

$$I' = I + \left( f_s - f_{s-1} \right) \Delta x_{s,1} + \left( \frac{\frac{\Delta x_{s,1}}{N}}{\sum_{j=0}^{N} x_{j,1}} \right) \left( x_s - x_{s-1} \right) + 2 \frac{\left( \Delta x_{s,1} \right)^2}{\frac{N}{\sum_{j=0}^{N} x_{j,1}}}$$

[A3.27]

Given that the urban areas are ordered from the smallest to the largest, $x_s > x_{s-1}$. But it cannot be assumed that $f_s > f_{s-1}$. If it is, then $I' > I$, in agreement with the Pigou-Dalton principle: the potential for agglomeration economies increases when population shifts from a smaller urban area which contains a smaller fraction of the territory's population to a larger urban area with a larger fraction of the territory's population. Otherwise, if $f_s < f_{s-1}$, the result is ambiguous, as the negative effect described in case B may possibly overwhelm the positive effect of the increase in the size of the larger urban area. Such ambiguity, however, in no way contradicts the Pigou-Dalton principle.

# Appendix IV – Derivative of the index relative to the Pareto parameter

Begin with

$$I = x_N \frac{\displaystyle\sum_{i=1}^{N} \left(\frac{1}{N+1-i}\right)^{2/a}}{\displaystyle\sum_{i=1}^{N} \left(\frac{1}{N+1-i}\right)^{1/a}} \qquad\qquad [42]$$

Note that

$$\sum_{i=1}^{N} \left(\frac{1}{N+1-i}\right)^{1/a} = \sum_{i=1}^{N} \left(\frac{1}{i}\right)^{1/a} \quad \text{and} \quad \sum_{i=1}^{N} \left(\frac{1}{N+1-i}\right)^{2/a} = \sum_{i=1}^{N} \left(\frac{1}{i}\right)^{2/a} \qquad [A4.01]$$

So

$$I = x_N \frac{\displaystyle\sum_{i=1}^{N} \left(\frac{1}{i}\right)^{2/a}}{\displaystyle\sum_{i=1}^{N} \left(\frac{1}{i}\right)^{1/a}} = x_N \frac{\displaystyle\sum_{i=1}^{N} i^{-2/a}}{\displaystyle\sum_{i=1}^{N} i^{-1/a}} \qquad\qquad [A4.02]$$

We have

$$\frac{\partial I}{\partial a} = x_N \frac{\partial}{\partial a} \frac{\displaystyle\sum_{i=1}^{N} i^{-2/a}}{\displaystyle\sum_{i=1}^{N} i^{-1/a}} \qquad\qquad [A4.03]$$

$$\frac{\partial I}{\partial a} = x_N \frac{\displaystyle\sum_{i=1}^{N} i^{-1/a} \sum_{i=1}^{N} \frac{\partial}{\partial a} i^{-2/a} - \sum_{i=1}^{N} i^{-2/a} \sum_{i=1}^{N} \frac{\partial}{\partial a} i^{-1/a}}{\left(\displaystyle\sum_{i=1}^{N} i^{-1/a}\right)^2} \qquad\qquad [A4.04]$$

where

$$\frac{\partial}{\partial a} i^{-1/a} = i^{-1/a} \ln i \left(\frac{1}{a}\right)^2 \quad \text{and} \quad \frac{\partial}{\partial a} i^{2/a} = 2i^{2/a} \ln i \left(\frac{1}{a}\right)^2 \qquad [A4.05]$$

so that

$$\frac{\partial I}{\partial a} = x_N \frac{\sum_{i=1}^{N} i^{-1/a} \sum_{i=1}^{N} \left[ 2i^{2/a} \ln i \left(\frac{1}{a}\right)^2 \right] - \sum_{i=1}^{N} i^{-2/a} \sum_{i=1}^{N} \left[ i^{-1/a} \ln i \left(\frac{1}{a}\right)^2 \right]}{\left( \sum_{i=1}^{N} i^{-1/a} \right)^2}$$

[A4.06]

$$\frac{\partial I}{\partial a} = x_N \left(\frac{1}{a}\right)^2 \frac{2\sum_{i=1}^{N} i^{-1/a} \sum_{i=1}^{N} \left[ i^{2/a} \ln i \right] - \sum_{i=1}^{N} i^{-2/a} \sum_{i=1}^{N} \left[ i^{-1/a} \ln i \right]}{\left( \sum_{i=1}^{N} i^{-1/a} \right)^2}$$

[A4.07]

$$\frac{\partial I}{\partial a} = x_N \left(\frac{1}{a}\right)^2 \frac{2\sum_{j=1}^{N} j^{-1/a} \sum_{i=1}^{N} \left[ i^{2/a} \ln i \right] - \sum_{j=1}^{N} j^{-2/a} \sum_{i=1}^{N} \left[ i^{-1/a} \ln i \right]}{\left( \sum_{i=1}^{N} i^{-1/a} \right)^2}$$

[A4.08]

$$\frac{\partial I}{\partial a} = x_N \left(\frac{1}{a}\right)^2 \frac{2\sum_{j=1}^{N} \sum_{i=1}^{N} \left[ j^{-1/a} i^{-2/a} \ln i \right] - \sum_{j=1}^{N} \sum_{i=1}^{N} \left[ j^{-2/a} i^{-1/a} \ln i \right]}{\left( \sum_{i=1}^{N} i^{-1/a} \right)^2}$$

[A4.09]

where the sign of the derivative depends on the sign of its numerator.