CrossMark

Measuring Urban Agglomeration: A Refoundation of the Mean City-Population Size Index

Andre Lemelin · Fernando Rubiera-Morollón · Ana Gómez-Loscos

Abstract In this paper, we put forth the view that the potential for urbanization economies increases with interaction opportunities. From that premise follow three fundamental properties that an agglomeration index should possess: (1) to increase with the concentration of population and conform to the Pigou–Dalton transfer principle; (2) to increase with the absolute size of constituent population interaction zones; and (3) to be consistent in aggregation. Limiting our attention to pairwise interactions, and invoking the space-analytic foundations of local labor market area (LLMA) delineation, we develop an index of agglomeration based on the number of interaction opportunities per capita in a geographical area. This leads to Arriaga's mean city-population size, which is the mathematical expectation of the size of the LLMA in which a randomly chosen individual lives. The index has other important properties. It does not require an arbitrary population threshold to separate urban from non-urban areas. It is easily adapted to situations where an LLMA lies partly outside the geographical area for which agglomeration is measured. Finally, it can be satisfactorily approximated when data is truncated or aggregated into size-classes. We apply the index to the Spanish NUTS III regions, and evaluate its performance by examining its correlation with the location quotients of several knowledge intensive business services known to be highly sensitive to urbanization economies. The Arriaga index's correlations are clearly stronger than those of either the classical degree of urbanization or the Hirshman–Herfindahl concentration index.

A. Lemelin
INRS - Institut National du la Recherche Scientifique, Quebec, Canada
e-mail: andre.lemelin@ucs.inrs.ca

F. Rubiera-Morollón (✉)
REGIOlab, University of Oviedo, Oviedo, Asturias, Spain
e-mail: frubiera@uniovi.es

A. Gómez-Loscos (✉)
Banco de España, Madrid, Spain
e-mail: anlosc@gmail.com

1 Introduction

Typically, the relationship between development and urbanization is illustrated with a
graph plotting the degree of urbanization (fraction of population living in an urban area)
against GDP per capita or its rate of growth. This is true of international comparisons
(Henderson 2003a, p. 281; Spence et al. 2009, Annex 2), as well as regions (Crédit Suisse
2012, p. 16; Zhu et al. 2012, Fig. 2). Urban concentration has also been the subject of much
attention in relation to economic development. Indeed, about the urbanization process that
occurs with development, Henderson (2003b) writes: "There are two key aspects to the
process. One is urbanization itself and the other is urban concentration (…)". Brülhart and
Sbergami (2009) also measure agglomeration alternatively through urbanization shares and
through indexes of spatial concentration. In this paper, we develop a measure which
combines both aspects.

Specifically, the issue we address is how to measure urbanization, using readily available
data, in a way that reflects the potential for agglomeration economies of the urbanization
type. Our approach is founded on the view that agglomeration provides opportunities for
interactions between economic agents, a key mechanism by which urbanization economies
are generated. Our measure will therefore be closely related to the view of an "urban area" as
an integrated market. And as it turns out, we "rediscover" an index originally proposed by
demographer Eduardo Arriaga (1970, 1975) as a measure of urbanization. That index pos-
sesses three properties that are fundamental for a measure of agglomeration: (1) it increases
with the concentration of population and conforms to the Pigou–Dalton transfer principle; (2)
it increases with the absolute size of constituent population interaction zones; and (3) it is
consistent in aggregation. The index has other important properties: it does not require an
arbitrary population threshold to separate urban from non-urban areas; it is easily adapted to
situations where an population zone lies partly outside the geographical area for which
agglomeration is measured (boundary problem); and it can be satisfactorily approximated
when data is truncated or aggregated into size-classes.

The rest of this paper is organized as follows. The next section presents a view of
urbanization economies resulting from opportunities for interaction. Ensue the three fun-
damental properties that we believe an agglomeration index should possess. Concentration
measures, in particular, fail to meet these conditions, but Arriaga's index does. The next
section develops this index from our theoretical view of agglomeration and discusses its
properties. Then the index is computed for the Spanish NUTS III regions, and its per-
formance is compared to that of the degree of urbanization and the Hirshman–Herfindahl
concentration index. A concluding section completes the paper.

2 Measuring Urban Agglomeration as Opportunities for Interaction

A taylor, a physician and a sports coach all take measurements of the human body. But for
different purposes, they use different measures. How should we measure agglomeration for

the purpose of examining the potential for agglomeration economies of the urbanization type (urbanization economies, for short)? To answer that question, we need to have a theory, a model, or at least a general view of how urbanization economies arise.

The concept of *Agglomeration economies*, first proposed by Weber (1909), is central in regional and urban economics. Ohlin (1933), Hoover (1937), Isard (1956) clarify the idea and distinguish different types of *agglomeration economies*: (1) *large-scale* economies, (2) *localization* economies and (3) *urbanization* economies. Here we are concerned with urbanization economies, and more specifically with urbanization economies insofar as they are the result of interaction between economic agents. The concentration of population in an urban area multiplies opportunities for interaction; the greater the number of possible interactions, the greater the potential for urbanization economies, as proximity encourages formal and informal exchanges of ideas which nourish innovation and contribute to the diffusion of knowledge (knowledge spillovers). Consequently, the measure we are looking for is a measure of opportunities for interaction. This aspect of urbanization economies has been examined by Glaeser et al. (1992) in relation to city growth. The authors compare the evolution of employment in individual industries across cities, to confront the predictions of competing views of how knowledge spillovers stimulate growth (specialization vs. diversity; competition vs. monopoly and the internalization of externalities). In this paper, however, we deal with a different issue, namely, measuring opportunities for interaction in a geographical area (region, group of regions, country) that may comprise several cities, or a single city, or even none at all (a region of villages, for example).

Now, in the abstract, restricting our attention to interaction between pairs of individuals, the number of possible pairs in a group increases as the square of the number of individuals.[1] Not all links, however, are equally probable: distance (physical or social) may impede communication, and congestion may interfere with exchanges. But, practically speaking, it is impossible to weigh all pairs according to distance and congestion factors. Yet these can be taken into account indirectly, using labor mobility as an indicator of whether interactions are likely or not.

Local labor market areas (LLMAs) are delineated on the basis of labor mobility, so they are well-suited as basic territorial units for examining interaction-based urbanization economies. Labor mobility is also a key criterion used by statistical agencies to circumscribe metropolitan or urban areas. Consequently, an LLMA cannot in principle comprise only part of a metropolitan or urban area. LLMAs may (and do) however transcend municipal boundaries and include several population nuclei, as do metropolitan areas. But, contrary to metropolitan and urban areas, the set of LLMAs covers the whole territory, not just the main urban areas (it is a partition of the territory). It follows that any area that would not be comprised in any metropolitan or urban area is nonetheless included in some LLMA; so LLMAs containing metropolitan areas may also contain some additional areas. And, of course, not all LLMAs are metropolitan areas or even urban areas. The delineation of LLMAs has been implemented in several countries.[2] The partitioning of geographic space into LLMAs is an exercise in the spatial analysis of commuting patterns, in order to

---

[1] The number of distinct pairs in a group of n individuals is equal to $n(n–1)/2$. If n is sufficiently large, this can be approximated by $n^2/2$. More generally, combinatorics shows that the number of interactions involving two persons or more increases more than proportionally with the size of the group.

[2] Among which (in alphabetical order): Canada (Munro et al. 2011); France (DATAR-DARES-INSEE 2011); Italy (ISTAT 1997, 2005, 2006; Sforzi 2012); New Zealand (Statistics New Zealand 2009; Goodyear 2008; Papps and Newell 2002); Portugal (Alfonso and Venâncio 2013); Spain (Rubiera and Viñuela 2012; Boix and Galleto 2006); United Kingdom (Bond and Coombes 2007; ONS 2007); USA (Tolbert and Sizer 1996; USDA ERS 2012).

"define geographical units where the majority of the interaction between workers seeking jobs and employers recruiting labour occurs (i.e. to define boundaries across which relatively few people travel between home and work)" (Casado-Díaz 2000).

Now, population is often used as an indicator of potential urbanization economies for a metropolitan area. In such case, when the focus is on opportunities for interaction, the implicit simplifying assumption is made that the interactions which matter are the ones taking place within the limits of the metropolitan area. The same simplifying assumption can be made regarding LLMAs to estimate the number of interaction opportunities in each LLMA. And since LLMAs define a partition of geographic space, including zones that would be classified as non-urban, it becomes possible to estimate the number of interaction opportunities for a region by aggregating the number of interaction opportunities in the LLMAs that constitute the region, following the procedure detailed below. Finally, to relate the number of interaction opportunities to productivity (output per unit of input), it must be divided by population. By computing interaction opportunities at the LLMA level before aggregating, we use the labor mobility criterion which defines LLMAs to determine which interactions are likely or not. This is how we implicitly take into account distance decay.[3]

What does this interaction-based view of agglomeration tell us about how to measure the possibility of urbanization economies? First, since the number of opportunities for interaction increases more than proportionately with population, an agglomeration index should increase with concentration. More specifically, an agglomeration index should conform to (an inverted form of) the Pigou–Dalton transfer principle:[4] the number of opportunities created by moving one person to a larger LLMA (increasing concentration) is greater than the number of opportunities destroyed by his/her leaving the smaller LLMA. Second, a measure of concentration only is not a proper agglomeration index, because concentration is a property of the distribution of population between LLMAs, and it is insensitive to their sizes: there are more interaction opportunities *per capita* in a region consisting of two LLMAs with populations of 150,000 and 50,000 than in a region consisting of two LLMAs with populations of 15,000 and 5,000; or, to take an absurd example, a desert with a single inhabitant is 100 % concentrated, but offers no opportunities for interaction. Third, an agglomeration index should be consistent in aggregation: the same rule that is applied to aggregate individual LLMA agglomeration indexes into a regional index should also be valid to aggregate regional indexes into an agglomeration index of a group of regions.

There are other desirable properties, which will be discussed later. Since there are interaction opportunities even in the smallest of villages, an agglomeration index should be defined without reference to a population threshold below which an area is excluded from calculations. On the other hand, data is sometimes truncated, or aggregated into size-classes, so it would be advantageous to be able to adapt an agglomeration index to such circumstances. More generally, an index with modest data requirements is more susceptible of application. Another attractive property is the possibility of computing an agglomeration index for a region that includes only part of some LLMAs (more about the boundary problem below).

---

[3] However, our index does not take into account negative congestion externalities. Referring to Capello and Camagni (2000), it could be said that our index is one component of a "city effect indicator" of positive externalities, while leaving aside congestion externalities which could be accounted for in an "urban overload indicator".

[4] Dalton (1920, p. 351), Pigou (1912, p. 24). Also Cowell (2009).

Let us now examine measures of urbanization found in the literature to see whether they have the fundamental properties enunciated above, namely: (1) to increase with the concentration of population and conform to the transfer principle; (2) to increase with the absolute size of constituent LLMAs; and (3) to be consistent in aggregation. The *degree of urbanization* (percentage of population living in an urban area) fails as a measure of agglomeration with respect to all three criteria, in addition to raising the difficulty of drawing a line between urban and non-urban.

Much of the literature on urbanization and development, however, focuses on the relationship between urban concentration and economic development or productivity. Yet, by definition, measures of concentration do not satisfy our second criterion. As a matter of fact, according to Cowell (2009), a measure of concentration (which Cowell applies to measure income inequality) should satisfy the "income scale independence principle", which, transposed to the measurement of agglomeration, is the exact opposite of our second criterion of an interaction-based measure of agglomeration. It will nonetheless be useful to briefly review urban concentration measures. Wheaton and Shishido (1981), Henderson (1988) use the Hirshman–Herfindahl index, to which we shall return later. In the present context, it is defined as the sum of squared population shares of LLMAs.[5] Henderson (2003a, 2003b) uses *urban primacy*, the share of urban population that lives in the largest city. It does not satisfy the transfer principle, but Henderson argues that it is conveniently available for many years and many countries, and that it is closely correlated with Hirshman–Herfindahl indexes. Rosen and Resnick (1980)[6] measure urban concentration using the econometrically estimated exponent of the Pareto distribution for the city-size distributions of 44 countries. They find that it is quite sensitive to the definition of the city and the choice of city sample size (number of cities or city-population threshold). Brülhart and Sbergami, (2009) apply Theil entropy indexes developed in Brülhart and Traeger (2005). But, as mentioned earlier, these are all concentration measures which fail to satisfy our second criterion.

Uchida and Nelson (2010) look for a remedy to inconsistencies in published UN data on the degree of urbanization which are due to divergences between reporting countries in the way they delineate urban areas and in the population thresholds applied to define what is urban and what is not. They propose an "agglomeration index" which focuses on the key indicators of the sources of agglomeration economies: population density, the size of the population in a "large" urban centre, and travel time to that urban centre. Combining data from several sources (including GIS data on transport networks), and applying interpolation techniques, they map all three factors on the surface of the earth in 1-kilometer pixels. For each of the three factors, a threshold is defined, and the estimated population in pixel areas that meet all three criteria is classified as "urban". This yields an estimate of the degree of urbanization. The Uchida-Nelson agglomeration index approach is promising, but it remains tied to the urban/non-urban dichotomy and as such, it fails all three of the criteria mentioned above.

Reflecting on the weaknesses of the percentage of population living in "urban areas" as a measure of urbanization, demographer Eduardo Arriaga (1970, 1975) proposed an index which "takes into account the statistical concept of the expected value of the size of the locality where a person, randomly chosen, resides" (p. 208). In the following section, we develop an index based on interaction opportunities, which turns out to be Arriaga's index

---

[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.

[6] Mentioned by both Henderson (2003b), Wheaton and Shishido (1981).

applied to LLMAs. This index, as we shall see, satisfies all three basic conditions listed above.

3 Arriaga's Index Applied to LLMAs

3.1 Arriaga's Index as a Measure of Interaction Opportunities

We derive Arriaga's agglomeration index from our view of urbanization economies as the result of interaction between economic agents. In so doing, we make two simplifying assumptions. First, we limit our attention to *pairwise* interactions, so that the number of interaction opportunities among n persons is $n(n–1)/2$, which we approximate as $n^2/2$. Second, invoking the space-analytic foundations of LLMA delineation, we take into account interaction opportunities within LLMAs, while ignoring interactions that may take place across LLMAs.

Formally, let n_i be the population size of the ith LLMA in the geographical area under consideration. Then the total amount of interaction opportunities is approximated as $\frac{1}{2}\sum_i n_i^2$. We relate productivity (output per unit of input) to the number of interaction opportunities of the average individual, which leads to the following measure of agglomeration:

$$\text{Measure of agglomeration} = \frac{\sum_i n_i^2}{2\sum_i n_i} \tag{1}$$

Expression (1) is precisely one half of Arriaga's mean city-population size. Now let

$$f_i = \frac{n_i}{\sum_j n_j} \tag{2}$$

be the fraction of population in the ith LLMA. Dropping the division by 2, we have an index of agglomeration defined as

$$I = \frac{\sum_i n_i^2}{\sum_i n_i} = \sum_i \left(\frac{n_i}{\sum_j n_j}\right) n_i = \sum_i f_i n_i \tag{3}$$

This is Arriaga's (1970) index. His interpretation of the index stands out from formula (3)[7]: since f_i is the probability that a randomly chosen individual reside in the ith LLMA, then I is the mathematical expectation of the size of the LLMA in which a randomly chosen individual lives.[8] In other words, the average individual lives in an LLMA of size I.

Let us pursue the development, substituting from (2), we find:

---

[7] 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 LLMAs cover the whole territory. Lemelin et al. (2012, Appendix 1) present a version of the index that is based on information aggregated by size classes and therefore deals with truncation.

[8] We assume that each individual has an equal probability of being chosen, in which case relative frequencies are correctly interpreted as probabilities.

$$I = \sum_i f_i n_i = \left(\sum_j n_j\right) \sum_i f_i \left(\frac{n_i}{\sum_j n_j}\right) = \left(\sum_j n_j\right) \sum_i f_i^2 \tag{4}$$

where:

$$H = \sum_i f_i^2 \tag{5}$$

is the Hirschman–Herfindahl (HH) concentration index. So:

$$I = \left(\sum_j n_j\right) H \tag{6}$$

Our index is the HH concentration index, multiplied by the total population in the geographical area under consideration. This illustrates how it accounts for both concentration and size, which we consider as two aspects of the potential for economies of agglomeration of the urbanization type.

The reader may be surprised that the index takes the same value $I = n$ for a territory with a single LLMA of size n as it does for a territory with K LLMAs, all of size n. This is true, and it is as it should be. It can be seen from Eq. (6) that in the first case, $I = n \times 1$; in the second, using the numbers equivalent property[9] of H, $I = (K \times n)(1/1\ K \times K) = n$, as greater size makes up for less concentration. In both cases, the average individual lives in an LLMA of size n (following Arriaga's interpretation). Insofar as I measures the average individual's interaction opportunities, the potential increase in productivity associated with agglomeration economies is equal in both cases.

Finally, Arriaga's index has an interesting geometric interpretation: in a graph of the stepwise cumulative distribution of population according to LLMA sizes, it is the area above the curve. We illustrate this using fictitious data as an example. Table 1 gives the population of each of 5 LLMAs in some geographic area under investigation.

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 LLMAs with a population of at least 50,000. In our example this would be 92.4 % (425,000/460,000). Referring to the distribution of population (Table 1, column 2), 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 of the more detailed distribution of population, represented in Fig. 1. It is completely insensitive to the distribution of population among the LLMAs with more than 50,000 inhabitants.

Let K be the number of LLMAs in the geographical area under consideration (here 5). Then the area above the curve is equal to:

$$I = \sum_{i=1}^{K} (1 - F_{i-1})(n_i - n_{i-1}) \tag{7}$$

---

[9] Adelman (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 LLMAs which would exhibit a concentration level equal to H.

Table 1 Fictitious data

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

Fig. 1 Cumulative distribution of population according to LLMA sizes
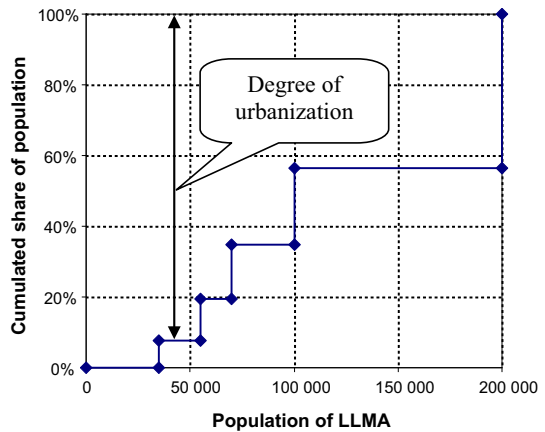
where LLMAs are assumed to be ordered from smallest to largest, and $F_i = \sum_{j=1}^{i} f_j$ is the cumulative distribution. In our example, this is equal to 0.0242. It is shown in the Appendix 1 that (7) is equivalent to (3).

Before moving on to examining the properties of the index, it should be pointed out that Arriaga's original presentation dealt not with LLMAs, but with traditional city-population data. Arriaga examined the sensitivity of his index to the choice of an urban threshold, a cut-off point in the distribution below which localities are classified as rural and excluded from the computation of the index; he found that it was pretty robust.

3.2 Properties of the Index

To start with, the domain of the Arriaga index is well defined. It is 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 autarkic villages (of, say, 100 inhabitants); the cumulative distribution curve would then be close to an upside-down «L», with the horizontal line at the 100 % level, and the vertical line at the 100 population level. The upper bound of the index is equal to the population of the largest LLMA in the geographical area under consideration, and would occur if all population were concentrated in that largest LLMA. The curve would then be a mirror-image of an «L», with the vertical bar to the right.

3.2.1 Axiomatic Properties

We now verify whether the Arriaga index possesses the fundamental properties of an agglomeration index. Considering Eq. (6), the value of the index clearly increases with concentration as measured by the HH index.[10] Moreover, it is demonstrated in the Appendix 3 that it conforms to the transfer principle. Second, once again turning to Eq. (6), the index clearly increases with the average size of LLMAs.

Let us examine whether the Arriaga index is consistent in aggregation. First, notice that, for a geographical area containing a single LLMA of population n, formula (3) shows that the value of the index is simply $n^2/n = n$. The same formula shows that for K LLMAs, the value of the index is a weighted average of individual LLMA indices n_i, where the weights f_i are the population shares of LLMAs. Now consider a geographical area of interest partitioned into two regions, defined by a pair of complementary sets A and $\bar{A}$: the ith LLMA belongs to one region if $i \in A$, and to the other if $i \notin A$. Now let

$$F_A = \left(\frac{\sum\limits_{i \in A} n_i}{\sum\limits_{i} n_i} \right) \text{ and } F_{\bar{A}} = \left(\frac{\sum\limits_{i \notin A} n_i}{\sum\limits_{i} n_i} \right) = 1 - F_A \qquad (8)$$

where FA is the fraction of population that lives in an LLMA belonging to the set A of LLMAs that constitute one region, while $F_{\bar{A}}$ is the fraction of population that lives in the other region. Also let

$$f_{i|A} = \frac{n_i}{\sum\limits_{j \in A} n_j} = \frac{n_i}{\sum\limits_{j} n_j} \frac{\sum\limits_{j} n_j}{\sum\limits_{j \in A} n_j} \quad \text{and likewise, } \frac{f_i}{F_A} f_{i|\bar{A}} = \frac{f_i}{F_{\bar{A}}} \qquad (9)$$

According to formula (3), we have

$$I_A = \sum_{i \notin A} f_{i|A} n_i \quad \text{and } I_{\bar{A}} = \sum_{i \notin A} f_{i|\bar{A}} n_i \qquad (10)$$

Recalling that n_i is the agglomeration index of the ith LLMA, formula (3) applied to the two regions translates as

$$I = F_A I_A + (1 - F_A) I_{\bar{A}} \qquad (11)$$

Substitute from (9) and (10), and use $F_{\bar{A}} = 1 - F_A$ to find

$$I = F_A \sum_{i \in A} \left(\frac{f_i}{F_A} \right) n_i + F_{\bar{A}} \sum_{i \notin A} \left(\frac{f_i}{F_{\bar{A}}} \right) n_i = \sum_{i} f_i n_i \qquad (12)$$

So indeed, the same rule that is applied to aggregate individual LLMA agglomeration indexes into a regional index is also valid to aggregate regional indexes into an agglomeration index of a group of regions.

---

[10] It does not, however, increase monotonically with the Pareto parameter interpreted as a measure of concentration. The reason for this is exposed in the Appendix 2.

3.2.2 Boundary Problem and Other Applicability Considerations

So far, the Arriaga index has been developed under the assumption that every LLMA is entirely contained in the geographical area for which the index is to be computed. But this cannot be guaranteed, since LLMAs are delineated without regard for administrative boundaries.[11] Let f_{ir} be the fraction of the population in region r residing in the ith LLMA. Then, bearing in mind Arriaga's (1970) interpretation, the average size of the LLMA where a randomly chosen individual lives is given by a formula slightly different from (3):

$$I_r = \sum_i f_{ir} n_i \tag{13}$$

where f_{ir} replaces f_i. Eq. (13) is the way to compute our index for regions when LLMAs extend across regional boundaries. Now, however, the tight relationship with the HH concentration index breaks down.

To sum up, the Arriaga index applied to LLMAs satisfies the three fundamental properties of an agglomeration index. It is less difficult to compute than the Uchida and Nelson (2010) agglomeration index, while implicitly taking into account their three criteria—population density, the size of a "large" urban centre, and travel time—through the spatial-analytic underpinnings of LLMA delineation, and without the requirement of defining an arbitrary urban threshold. On the other hand, if a delineation of LLMAs is not available, the Arriaga index is applicable to traditional city-population data, although it is preferable that "cities" be defined as functional areas as are metropolitan areas (in general, every metropolitan area is an LLMA). Finally, we have shown elsewhere[12] that it would also be possible to compute the index, and obtain similar results, when the underlying population data is available only in LLMA size-categories, rather than for individual LLMAs.

4 Empirical Application

In this section, we apply this index to the Spanish provinces (NUTS III regions), using LLMAs as basic units for the construction of the index. 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 levels, even though a different geography may be preferable. Most of the economic information provided by the Spanish National Statistical Institute (INE) (GDP, stock of capital, wages or employment data…) is available only for the whole country or at the NUTS III level. There is little data available at a finer level of geographical detail, such as municipalities. To find geographically disaggregated economic information, one has to look up some very specific databases or use the data provided by taxes or unemployment registers. This scarceness of finer information also prevails in many other countries. Fortunately, detailed population data is often available. And from detailed population data, it is possible to construct the Arriaga index and put the degree of agglomeration in relation with other key economic concepts such as regional productivity or growth.

---

[11] Such is the case in Spain, for example, where several LLMAs include spatial units that are located in more than one province.

[12] Lemelin et al. (2012, Appendix 1).

4.1 The Spanish Provinces

Administratively, Spain is divided into 8,105 municipalities that are aggregated into 52 provinces (NUTS III level) and 17 Autonomous Communities or NUTS II regions. 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, which have no real internal political or administrative meaning.

It is important to point out that municipalities are not the basic territorial units from which we construct our index. In Spain, a municipality is an administrative division of the territory which has not necessarily been defined with economic significance in mind. Indeed, in many cases, there is a high level of commuting between neighboring municipalities. And so municipalities have been aggregated into LLMAs which may transcend municipal boundaries, and eventually make up a metropolitan area, which might include several population nuclei surrounding a core one. To delineate LLMAs in Spain, Boix and Galleto (2006) have applied the regionalization method developed for Italy by Sforzi (ISTAT 1997, 2005, 2006; Sforzi 2012). The Spanish LLMAs have been delineated through a multi-stage process. Applying an algorithm that consists of four main stages and a fifth stage of fine-tuning, Boix and Galleto aggregate the 8,106 Spanish municipalities into 806 LLMAs. The algorithm starts with the municipal administrative unit and it generates the LLMAs using data on resident employed population, total employed population and home-to-work commuting, from the 2001 Spanish Population and Housing Census (INE).[13],[14]

The LLMA data is used to compute the Arriaga index of urban agglomeration for each province. Since there are LLMAs which straddle provincial boundaries, the actual formula used in the calculations is Eq. (13), developed above to deal with the boundary problem.[15] Finally, for ease of presentation, all index values were divided by the population of the Madrid LLMA, the largest in the country, so that their range of variation is from 0 to 1.

4.2 Comparison of the Indexes

Figure 2 represents the cumulative distribution of population according to LLMA size for the province of Asturias, 2001 (similar to Fig. 1). The area above the curve is equal to the Arriaga index, the value of which is 233,637, or 4.4 % of the population of the Madrid LLMA.

All the Spanish provinces are plotted in Fig. 3, ranked by the value of the index. Madrid is followed by Barcelona and, at a much greater distance by Vizcaya, Valencia and Seville,

---

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

[14] More details and applications of LLMAs are found in Fernández and Rubiera (2012). Rubiera and Viñuela (2012) perform an economic evaluation of this spatial unit.

[15] In Appendix 2 of Lemelin et al. (2012), this is compared with an index for which each LLMA 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.
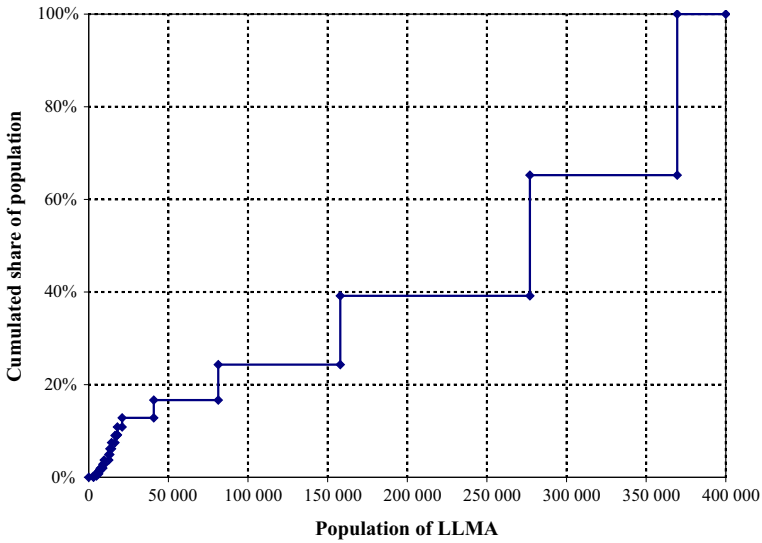
Fig. 2 Cumulative distribution of population according to LLMA size, Asturias (2001). *Source*: own based on INE (2001) data
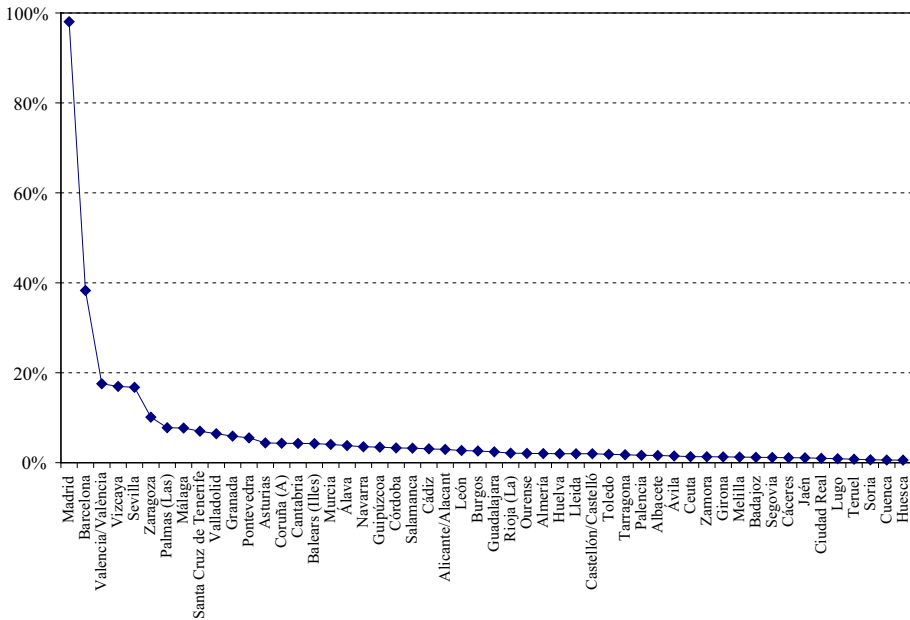

Fig. 3 Arriaga index for Spanish provinces, 2001 (% of Madrid LLMA population). *Source*: own based on INE (2001) data

which contain cities among the biggest in the country. At the opposite end are located the provinces with the lowest population densities of the country (Huesca, Cuenca, Soria and Teruel).
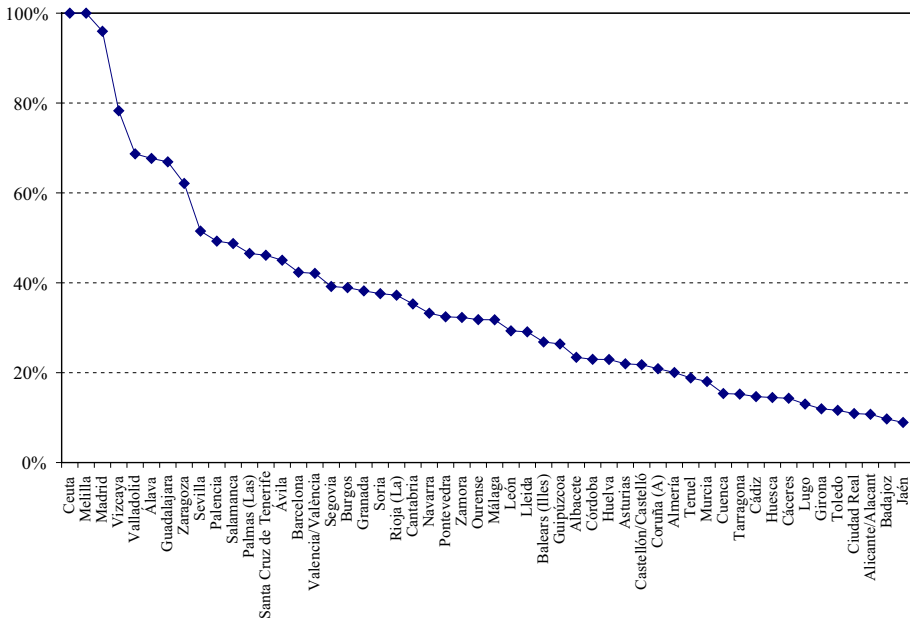
Fig. 4 HH concentration index for Spanish provinces, 2001 (%). *Source*: own based on INE (2001) data

Figure 4 plots the HH index of urban concentration and Fig. 5 the classical degree of urbanization (the percentage of population living in cities of more than 50,000 inhabitants). Both figures show how ranking provinces according to the HH index or the classical degree of urbanization leads to obvious aberrations if one wants to compare provinces with respect to their potential for interaction-based urbanization economies.[16] Ceuta and Melilla, two autonomous cities located in North Africa, with a joint population of less than 140,000 inhabitants in 2001, stand at the top of the hierarchy, with a 100 % degree of urbanization and a HH index of 1. The province of Teruel has no LLMA of 50,000 inhabitants or more, so its degree of urbanization is zero, and the tail-end of the curve in Fig. 5 drops abruptly. In spite of this, and although it is the fourth least populated province of Spain with the second lowest density, Teruel ranks significantly better, 39th, with respect to the HH index; the Arriaga index puts it in 49th position. There are other cases of misrepresentation with the HH index: for example, Alicante is in 50th position with respect to that index, but it occupies the 5th rank in Spain for population, and the 7th in terms of density; the Arriaga index, which takes into account both concentration and scale, ranks Alicante 24th. Barcelona, the province with the country's second largest city, is surprisingly ranked 15th according to the HH index. Aberrations also appear with the classical index: Sevilla, for instance, is the 4th largest province in terms of its population (1.7 million), 70 % of which live in the Sevilla metropolitan area, the 4th largest city in Spain; yet it falls to the 18th place according to its degree of urbanization, and to 10th place according to the HH index, behind Álava with a mere 300 thousand in population and a much lower density. We quote two final examples of violent changes in ranking between the three indexes. Málaga, the 6th largest city in Spain, occupies the 18th place according to the classical index, but the 8th with Arriaga's (9th with the HH

---

[16] Of course, the HH index and the classical degree of urbanization do not measure the same thing as our index (see Sect. 2 above), and it is normal that they differ from Fig. 1. But that is not our point.
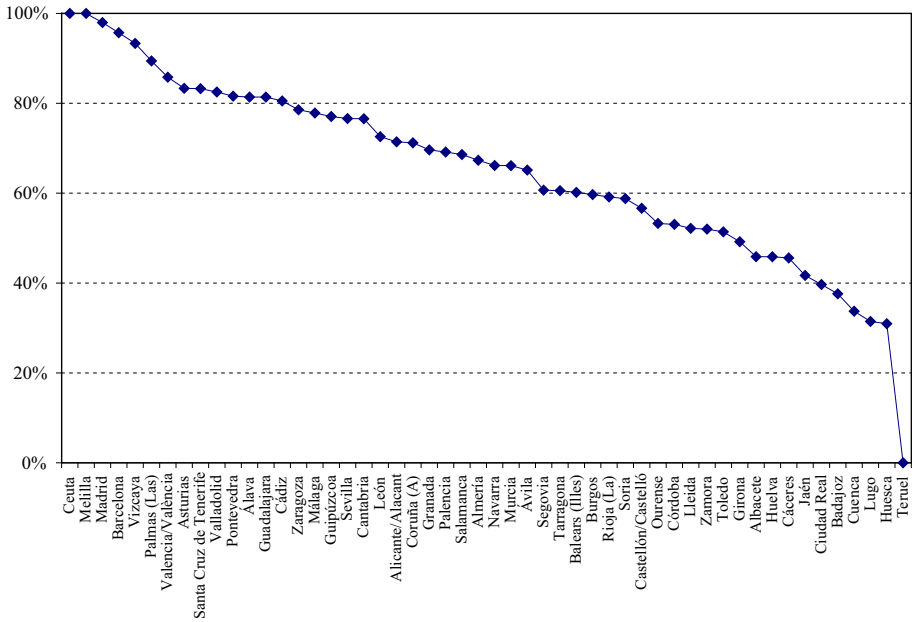
Fig. 5 Degree of urbanization of Spanish provinces, 2001 (% of population living in LLMAs of 50,000 hab. or more). *Source*: own based on INE (2001) data

index); Guadalajara, among the least populated provinces, with one of the lowest densities in the country, ranks 7th with the HH index and 13th with the classical index, while it is 27th with Arriaga's. We conclude that the Arriaga index displays a more suitable classification of the Spanish provinces, which better reflects potential agglomeration economies from urbanization, something the other two indexes are unable to capture.[17]

4.3 Index Performance Evaluation

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 of the urbanization type: high order producer services, also called knowledge intensive business services (henceforth KIBS). There are numerous empirical studies that use location quotients (alongside other measures) to confirm the tendency of these industries to concentrate in relation to agglomeration economies of the urbanization type.[18] Hence, we expect a good index of urban agglomeration to be highly correlated with

---

[17] In addition, we related the three measures of agglomeration to GDP per capita, and found a higher correlation with the Arriaga index. But the correlations were not spectacular, reflecting the fact that other determinants also play a major role.

[18] 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), Wernerheim and Sharpe (2003), among many others.

Table 2 Correlations between the location quotients of KIBS industries and the proposed index, the degree of urbanization and the HH index (2001)

	Arriaga index	Degree of urbanization	HH urban concentration index
All NUTS III regions			
Telecommunications	0.83	0.43	0.47
Financial services	0.76	0.29	0.29
Computing and information technologies	0.85	0.49	0.50
Research & development	0.79	0.65	0.54
Legal services and consulting	0.68	0.55	0.34
Technical consulting	0.61	0.36	0.36
Advertising	0.89	0.50	0.36
Audiovisual and press	0.84	0.55	0.48
Excluding outliers Ceuta and Melilla			
Telecommunications	0.84	0.55	0.72
Financial services	0.80	0.49	0.67
Computing and information technologies	0.86	0.59	0.74
Research & development	0.79	0.71	0.67
Legal services and consulting	0.69	0.62	0.46
Technical consulting	0.62	0.52	0.68
Advertising	0.91	0.64	0.62
Audiovisual and press	0.84	0.59	0.58

Source: Own

the location quotients of these services, and among competing indexes, we would prefer the one with the highest correlation.

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

$$LQ_{jp} = \frac{(e_{jp}/e_p)}{(E_j/E)} \tag{14}$$

where LQ_{jp} is the location quotient of sector j in province p; e_{jp} is employment in sector j in province p; $e_p = \sum_J e_{jp}$ is total employment in province p; $E_j = \sum_{i=1}^{n} e_{jp}$ 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.

Table 2 shows the correlations between each index (the degree of urbanization, the HH concentration index and our index), and the location quotients of eight high order producer-service industries. In all cases, our index is more closely correlated than the others with the location quotients. The second part of Table 2 shows the same correlations, but without the two outlier observations Ceuta and Melilla (see above). Interestingly, our index barely changes, while the two others improve substantially (but not to the point of becoming better than the our index); this may indicate that perhaps the Arriaga index deals more effectively with outliers.

Figure 6 illustrates the relationship between each of the three indexes and the location quotients of four of the KIBS industries. To make the graphs more legible, our proposed index is plotted against a logarithmic scale, which transforms the linear trend line into a curve.
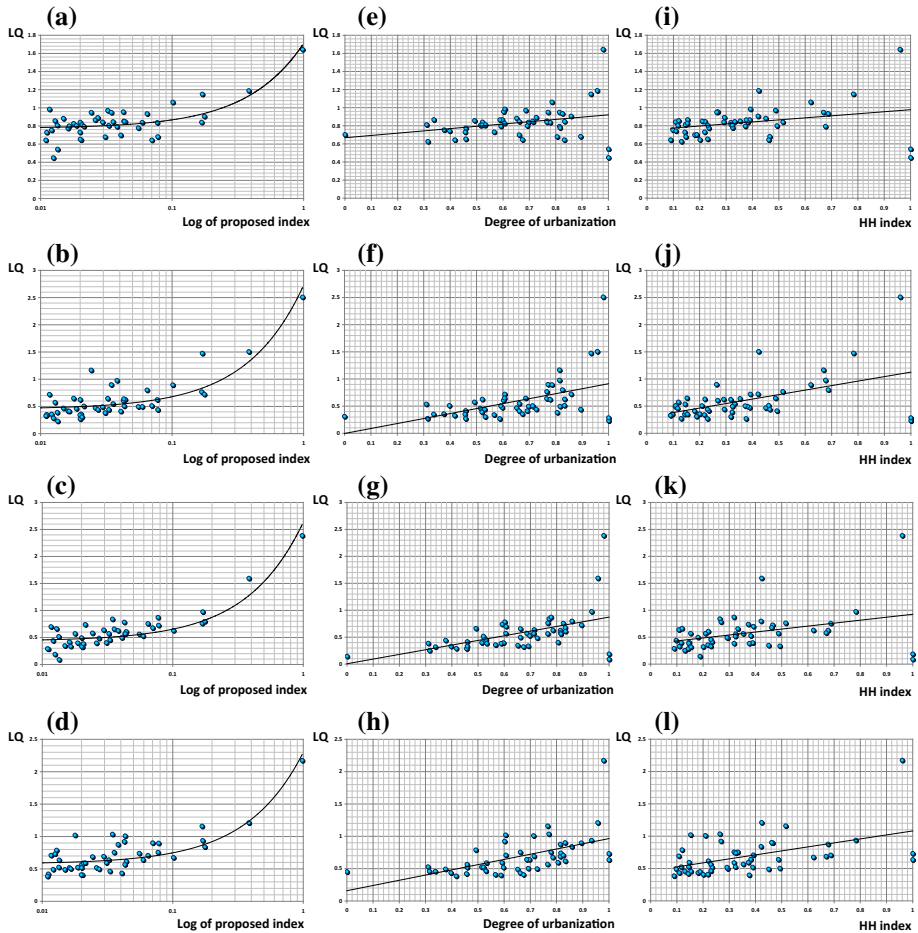
Fig. 6 Relation between the location quotients of selected KIBS industries and the proposed index, the degree of urbanization and the HH index (2001). Relation between **a** the proposed index and financial services LQs, **b** the proposed index and computing and information technologies LQs, **c** the proposed index and advertising services LQs, **d** the proposed index and audiovisual and entertainment services LQs, **e** the degree of urbanization and financial services LQs, **f** the degree of urbanization and computing and information technologies LQs, **g** the degree fo urbanization and advertising services LQs, **h** the degree of urbanization and audiovisual and entertainment services LQs, **i** the HH concentration index and financial services LQs, **j** the HH concentration index and computing and information technologies LQs, **k** the HH concentrationn index and advertising services LQs, **l** the HH concentration index and audiovisual and entertainment services LQs. *Note*: the curves in 6.1 a, b, c and d are not quadratic. They are straight lines displayed on a semi-logarithmic scale. The logarithmic abscissa is used to make the graphs legible: with a linear abscissa, the majority of data points would be crammed together near the origin. *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 first point to the left of Madrid. For the rest of the provinces, this index clearly displays a better fit between the location quotients of KIBS industries and the measure of urban agglomeration. As can be seen, the relation with both the degree of urbanization and the HH index, in many cases, shows apparent heteroskedasticity. The

largest deviations from trend appear for the higher values of the index, that is, for the provinces where the biggest cities are. In addition, both the degree of urbanization and the HH index are equal to 1 for Ceuta and Melilla, two autonomous cities, located in North Africa, with a joint population of less than 140,000 inhabitants in 2001. In general, the degree of urbanization and the HH index account for concentration, but not for size. As a consequence, they take on high values for provinces with a population that is substantial, but not so large, concentrated around a medium-sized city. In such cases, these two indicators clearly overstate the interaction opportunities and the potential for urbanization economies. The Arriaga index, which takes account of both size and concentration, does not suffer from the same distortion.

5 Summary and Conclusions

In this paper, we put forth the view that the potential for urbanization economies increases with interaction opportunities. From that premise follow three fundamental properties that an agglomeration index should possess: (1) to increase with the concentration of population and conform to the transfer principle; (2) to increase with the absolute size of constituent LLMAs; and (3) to be consistent in aggregation. Concentration measures, in particular, fail to meet condition (2).

We then develop an index of agglomeration based on the number of interaction opportunities per capita in a geographical area of interest. This is made possible thanks to two simplifying assumptions: (1) we limit our attention to *pairwise* interactions, and (2) invoking the space-analytic foundations of LLMA delineation, we take into account interaction opportunities within LLMAs, while ignoring interactions that may take place across LLMAs. This leads to Arriaga's mean city-population size, which is the mathematical expectation of the size of the LLMA in which a randomly chosen individual lives.

We apply the index to the Spanish provinces, and compare it to the degree of urbanization and the Hirshman–Herfindahl concentration index. We find that the three indexes rank the provinces quite differently. An examination of the more extreme cases of rank change shows that ranking according to the proposed index better reflects the geographical distribution of population, both with respect to size and concentration, and allows to correctly capture the potential for agglomeration economies from urbanization. Next, we correlate all three indexes with the location quotients of four knowledge intensive business services (KIBS) known to be highly sensitive to agglomeration economies of the urbanization type. We find that our index clearly displays a better fit between the location quotients of KIBS industries and the measure of urban agglomeration, as is confirmed by the much higher correlation coefficients.

The index has other advantages. It does not require to define an arbitrary population threshold which excludes areas classified as non-urban from calculations. It is easily extended to accommodate situations where an LLMA lies partly outside the geographical area for which agglomeration is measured. Finally, its already modest data requirements can be weakened if necessary to compute a satisfactory approximation of the index using data that is truncated or aggregated into size-classes. All these properties, together with the fact that the practice of delineating LLMAs is spreading among statistical agencies, make the index easily reproducible for different areas or countries and so, it will become increasingly convenient to use it.

This index, both in its original version and applied to LLMAs, is well rooted in a theoretical view of agglomeration economies, its data requirements are modest, and we

have shown that, at least in the case of Spain, it performs better than other commonly used agglomeration indicators. We look forward to seeing its use expand.

Appendix 1: Geometric Interpretation of the Agglomeration Index

Define $n_0 = 0$ and let K be the number of LLMAs in geographical area under consideration. Then

$$f_i = \frac{n_i}{\sum_{j=1}^{K} n_j} = \frac{n_i}{\sum_{j=0}^{K} n_j} \tag{15}$$

is the fraction of population residing in the ith LLMA (with $f_0 = 0$), and $F_i = \sum_{j=1}^{i} f_j = \sum_{j=0}^{i} f_j$ is the cumulative distribution (with $F_0 = 0$).[19] LLMAs are assumed to be ordered from smallest to largest. The area above the curve is computed as:

$$I = \sum_{i=1}^{K} (1 - F_{i-1}) (n_i - n_{i-1}) \tag{16}$$

In our example, this is equal to 0.0242. Note that the first term of formula (14) is the area above the curve to the left of the first LLMA in Fig. 1. This reflects the fact that LLMAs cover the whole territory, so that the threshold between urban and non-urban is irrelevant. The first term in (15) is equal to the size of the smallest LLMA:

$$(1 - F_0) (n_1 - n_0) = (1 - 0) (n_1 - 0) = n_1 \tag{17}$$

Equation (15) can be written as:

$$I = \sum_{i=1}^{K} \left(1 - \sum_{j=0}^{i-1} f_j \right) (n_i - n_{i-1}) \tag{18}$$

$$I = \sum_{i=1}^{K} \left(\sum_{j=i}^{K} f_j \right) (n_i - n_{i-1}) \tag{19}$$

$$I = \sum_{i=1}^{K} \left(\sum_{j=i}^{K} f_j \right) n_i - \sum_{i=1}^{K} \left(\sum_{j=i}^{K} f_j \right) n_{i-1} \tag{20}$$

Remembering that $n_0 = 0$,

$$I = \sum_{i=1}^{K} \left(\sum_{j=i}^{K} f_j \right) n_i - \sum_{i=2}^{K} \left(\sum_{j=i}^{K} f_j \right) n_{i-1} \tag{21}$$

$$I = \sum_{i=1}^{K} \left(\sum_{j=i}^{K} f_j \right) n_i - \sum_{i=1}^{K-1} \left(\sum_{j=i+1}^{K} f_j \right) n_i \tag{22}$$

---

[19] Note that the f_{ip} are independent of the scaling, since both the numerator and denominator are divided by the denominator in equation (1).

$$I = f_K n_K + \sum_{i=1}^{K-1} \left(\sum_{j=i}^{K} f_j \right) n_i - \sum_{i=1}^{K-1} \left(\sum_{j=i+1}^{K} f_j \right) n_i \qquad (23)$$

$$I = f_K n_K + \sum_{i=1}^{K-1} \left(\sum_{j=i}^{K} f_j - \sum_{j=i+1}^{K} f_j \right) n_i \qquad (24)$$

$$I = f_K n_K + \sum_{i=1}^{K-1} f_i n_i = \sum_{i=1}^{K} f_i n_i \qquad (25)$$

which is exactly Eq. (3).

Appendix 2: Transfer Principle

A key property of the index is that it correctly reflects the change in the potential for interactions and urbanization 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 Δn_i represent the change in the relative population size of the ith LLMA. A reallocation of population is restricted by the condition that $\sum_{i=1}^{k} \Delta n_i = 0$. Any reallocation can be represented as a series of reallocations between two LLMAs, and any reallocation between two LLMAs can be represented as a series of reallocations between an LLMA and the following or preceding one when LLMAs are ordered according to size. Therefore, we need only to consider a reallocation of population from the $(s-1)$th LLMA to the sth (from a LLMA to the next higher ranking one in terms of size):

$$\Delta n_s = -\Delta n_{s-1} > 0, \text{ and } \Delta n_i = 0 \text{ for } i \neq s, s-1 \qquad (26)$$

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

Following Eq. (2), define:[20]

$$\Delta f_i = \frac{\Delta n_i}{\sum_{j=1}^{K} n_j} = \frac{\Delta n_i}{\sum_{j=0}^{K} n_j} \qquad (27)$$

where, in view of (26),

$$\sum_{j=1}^{K} n_j - \Delta n_{s-1} + \Delta n_s = \sum_{j=1}^{K} n_j \qquad (28)$$

and

$$\Delta f_s = -\Delta f_{s-1} > 0 \qquad (29)$$

---

[20] The argument that follows can be generalized, albeit laboriously, to the version of the index that deals with the boundary problem. See Appendix 3 of Lemelin et al. (2012).

The value of the index after the reallocation is:

$$I' = \sum_{i \neq s-1, s} f_i n_i + (f_{s-1} - \Delta f_s)(n_{s-1} - \Delta n_s) + (f_s + \Delta f_s)(n_s + \Delta n_s) \tag{30}$$

$$I' = \sum_{i \neq s-1, s} f_i n_i + f_{s-1} n_{s-1} + f_s n_s + (f_s - f_{s-1}) \Delta n_s + \Delta f_s (n_s - n_{s-1}) + 2\Delta f_s \Delta n_s \tag{31}$$

$$I' = \sum_{i=1}^{K} f_i n_i + (f_s - f_{s-1}) \Delta n_s + \Delta f_s (n_s - n_{s-1}) + 2\Delta f_s \Delta n_s \tag{32}$$

$$I' = I + (f_s - f_{s-1}) \Delta n_s + \Delta f_s (n_s - n_{s-1}) + 2\Delta f_s \Delta n_s \tag{33}$$

Given that the LLMAs are ordered from the smallest to the largest, $ns > ns-1$, and $fs > fs-1$, so that $I' > I$.

Appendix 3: 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:

$$K + 1 - i = A n_i^{-a} \tag{34}$$

where K is the number of cities (ranked from the smallest to the largest), n_i is the size of city i,[21] and A and a are parameters. Parameter A can be calibrated from the size of the largest city:

$$K + 1 - K = 1 = A n_K^{-a} \tag{35}$$

$$A = n_K^a \tag{36}$$

Inverting (34), we obtain:

$$n_i^a = \frac{A}{K + 1 - i} \tag{37}$$

$$n_i = \left(\frac{A}{K + 1 - i} \right)^{1/a} \tag{38}$$

Total urban population is:

$$N = \sum_{i=1}^{K} n_i = \sum_{i=1}^{K} \left(\frac{A}{K + 1 - i} \right)^{1/a} \tag{39}$$

---

[21] It makes little difference whether city sizes are absolute or relative to some benchmark, such as Madrid above.
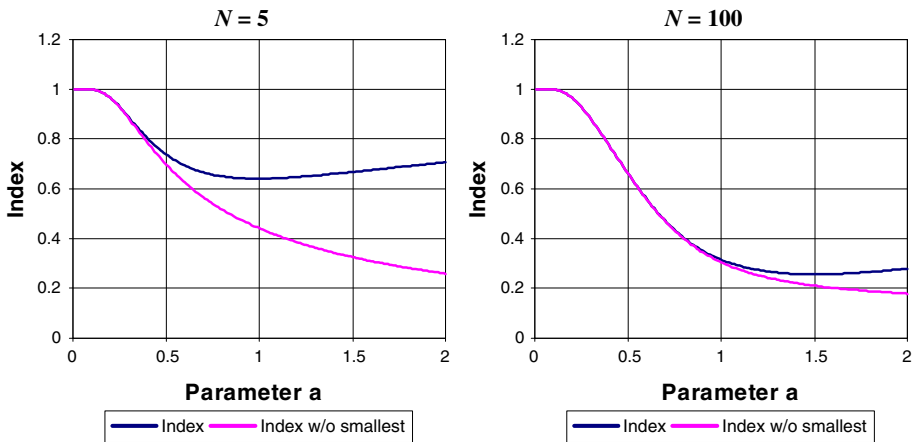
Fig. 7 Relationship of the proposed index to the Pareto elasticity parameter

And so it is quite straightforward to construct a cumulative distribution similar to the one in Fig. 1 reflecting a theoretical Pareto distribution. It is then possible to apply our proposed index to a theoretical Pareto distribution using formula (3). There results

$$I = \frac{\sum_{i=1}^{K} \left(\frac{A}{K+1-i}\right)^{2/a}}{\sum_{i=1}^{K} \left(\frac{A}{K+1-i}\right)^{1/a}} \tag{40}$$

where we exploit the identity in Eq. (3).[22] If we assume that the number of cities K and the size of the largest city n_K are fixed, then, using (36), (40) can be written as:[23]

$$I = \frac{\sum_{i=1}^{K} \left(\frac{n_K^a}{K+1-i}\right)^{2/a}}{\sum_{i=1}^{K} \left(\frac{n_K^a}{K+1-i}\right)^{1/a}} \tag{41}$$

$$I = \frac{\sum_{i=1}^{K} n_K^2 \left(\frac{1}{K+1-i}\right)^{2/a}}{\sum_{i=1}^{K} n_K \left(\frac{1}{K+1-i}\right)^{1/a}} \tag{42}$$

$$I = n_K \frac{\sum_{i=1}^{K} \left(\frac{1}{K+1-i}\right)^{2/a}}{\sum_{i=1}^{K} \left(\frac{1}{K+1-i}\right)^{1/a}} \tag{43}$$

---

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

[23] Note that the denominator of (40) is a CES aggregator function.

The derivative of the index relative to the Pareto parameter is [24]

$$\frac{\partial I}{\partial a} = n_K \left(\frac{1}{a}\right)^2 \frac{2 \sum_{j=1}^{K} \sum_{i=1}^{K} \left[j^{-1/a} i^{-2/a} \ln i \right] - \sum_{j=1}^{K} \sum_{i=1}^{K} \left[j^{-2/a} i^{-1/a} \ln i \right]}{\left(\sum_{i=1}^{K} i^{-1/a} \right)^2} \tag{44}$$

The sign of that derivative is the sign of its numerator, but we could not determine that sign analytically. Using numerical simulations,[25] 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, $n_1 = n_K K^{-1/a}$, increases with a, leaving a larger gap to the left of the first point on the cumulative distribution (see Fig. 1). Referring to index computation formula (7), it is easily verified that its first term is equal to n_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 Fig. 7).

References

Adelman, M. A. (1969). Comment on the "H" concentration measure as a numbers-equivalent. *The Review of Economics and Statistics, 51*(1), 99–101.

Alfonso, A., & Venâncio, A. (2013). *The relevance of commuting zones for regional spending efficiency*. Working Paper 17/2013/DE/UECE/ADVANCE, Department of Economics, School of Economics and Management, Technical University of Lisbon. http://pascal.iseg.utl.pt/~depeco/wp/wp172013.pdf 2014-01-10

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

Arriaga, E. E. (1975). "Selected measures of urbanization", Chap. II. In S. Goldstein, D. F. Sly (Eds.), *The measurement of urbanization and projection of urban population*, Working Paper 2, International Union for the scientific study of population. Committee on urbanization and population redistribution. Ordina Editions, Dolhain, Belgium.

Boix, R., & Galleto, V. (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.

Bond, S., & Coombes, M. (2007). *2001-based Travel-To-Work Areas Methodology*. Office for National Statistics. Retrieved from 13 January 2013. http://www.ons.gov.uk

Brülhart, M., & Sbergami, F. (2009). Agglomeration and growth: Cross-country evidence. *Journal of Urban Economics, 65*, 48–63.

Brülhart, M., & Traeger, R. (2005). An account of geographic concentration patterns in Europe. *Regional Science and Urban Economics, 35*, 597–624.

Capello, R., & Camagni, R. (2000). Beyond optimal city size: An evaluation of alternative growth patterns. *Urban Studies, 37*(9), 1479–1496.

Casado-Díaz, J. M. (2000). Local labour market areas in Spain: A case study. *Regional Studies, 34*, 843–856.

Cowell, F. A. (2009). *Measuring inequality, LSE perspectives on economic analysis*. Oxford: Oxford University Press.

Crédit Suisse Research Institute. (2012). *Opportunities in an Urbanizing World*, Zurich, Switzerland. Retrieved from 16 Janaury 2014. https://www.credit-suisse.com/ch/fr/news-and-expertise/research/credit-suisse-research-institute/publications.html

---

[24] The interested reader can find the proof in Lemelin et al. (2012, Appendix 4).

[25] Spreadsheet calculations were performed for values of K from 1 to 100, and a from 0.01 to 2 in increments of 0.01.

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal, 30*(119), 348–361.

Daniels, P. (1985). *Service industries: A geographical perspective*. New York: Methuen.

DATAR-DARES-INSEE (2011). *Atlas des zones d'emploi 2010*. Délégation interministérielle à l'Aménagement du Territoire et à l'Attractivité régionale (DATAR), Direction de l'Animation de la Recherche, des Études et des Statistiques (DARES) and Institut National de la Statistique et des Études Économiques (INSEE). Retrieved from 13 January 2014. http://www.insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=atlas-zone-emploi-2010

Excerpts from this document are downloadable from F. Sforzi's homepage. Retrieved from 23 January 2014. http://economia.unipr.it/DOCENTI/SFORZI/docs/files/SISTEMI_LOCALI.PDF

Fernández, E., & Rubiera, F. (2012). Defining the spatial scale in modern economic analysis: New challenges from data at local level. Advances in spatial science series. Berlin: Springer.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Schleifer, A. (1992). Growth in cities. *Journal of Political Economy, 100*(6), 1126–1152.

Goodyear, R. (2008). *Workforces on the move: An examination of commuting patterns to the cities of Auckland, Wellington and Christchurch*. Paper presented at NZAE conference, Wellington City, New Zealand, July 2008. Statistics New Zealand. Retrieved from 10 January 2014. http://www.stats.govt.nz/methods/research-papers/nzae/nzae-2008/workforces-on-the-move.aspx

Henderson, J. V. (1988). *Urban development: Theory, fact and illusion*. Oxford: Oxford University Press.

Henderson, J. V. (2003a). Urbanization and economic development. *Annals of Economics and Finance, 4*, 275–341.

Henderson, J. V. (2003b). The urbanization process and economic growth: The So-What question. *Journal of Economic Growth, 8*, 47–71.

Hoover, E. M. (1937). *Location theory and the shoe and leather industry*. Cambridge, MA: Harvard University Press.

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

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

Isard, W. (1956). *Location and space-economy*. Cambridge, MA: The Technology Press of Massachusetts, Institute of Technology.

ISTAT. (1997). *I Sistemi Locali del Lavoro 1991*. A cura di F. Sforzi, Collana Argomenti 10, Instituto Nazionale di Statistica, Roma.

ISTAT. (2005). *I Sistemi Locali del Lavoro. Censimento 2001. Dati definitivi*, Instituto Nazionale di Statistica. Retrieved from 10 January 2014. http://dawinci.istat.it/daWinci/jsp/MD/download/sll_comunicato.pdf

ISTAT. (2006). *8o Censimento generale dell'industria e dei servizi. Distretti industriali e sistemi locali del lavoro 2001*. Collana Censimenti, Instituto Nazionale di Statistica. Retrieved from 10 January 2014. http://www.istat.it/it/files/2011/01/Volume_Distretti1.pdf

Lemelin, A., Rubiera-Morrollón F., & Gómez-Loscos, A. (2012). « A territorial index of potential agglomeration economies from urbanization » , Montréal, INRS-UCS, coll. *Inédits*, 2012–03. http://www.ucs.inrs.ca/sites/default/files/centre_ucs/pdf/Inedit03-12.pdf

Munro, A., Alasia, A., & Bollman, R. D. (2011). "Self-contained labour areas: A proposed delineation and classification by degree of rurality", *Rural and Small Town Canada Analysis Bulletin*, 8(8), Catalogue no. 20-006-X, Statistics Canada. Retrieved from 16 January 2014. http://www.statcan.gc.ca/pub/21-006-x/21-006-x2008008-eng.htm

Ohlin, B. (1933). *Interregional and internal trade*. Cambridge, MA: Harvard University Press.

ONS. (2007). *Introduction to the 2001-based Travel-to-Work Areas*. ONS. Retrieved from 13 January 2013. http://www.ons.gov.uk

Papps, K. L., & Newell, J. O. (2002). *Identifying functional labour market areas in New Zealand: A reconnaissance study using travel-to-work data*. Discussion Paper No. 443, IZA (Institute for the study of labor), Bonn, Germany. Retrieved from 10 January 2014. http://ftp.iza.org/dp443.pdf

Pigou, A.C. (1912). *Wealth and Welfare*. MacMillan, London. Retrieved from 16 January 2014. https://archive.org/details/cu31924032613386

Polèse, M., Shearmur, R., & Rubiera, F. (2007). Observing regularities in location patters. An analysis of the spatial distribution of economic activity in Spain. *European Urban and Regional Studies, 14*(2), 157–180.

Rosen, K., & Resnick, M. (1980). The size distribution of cities: An examination of the pareto law and primacy. *Journal of Urban Economics, 8*, 165–186.

Rubiera, F., & Viñuela, A. (2012). "From funtional areas to analytical regions, where the agglomeration economies make sense", Chap. 2, p. 23–44. In: E. Fernández, F. Rubiera (Eds.) *Defining the spatial scale in modern economic analysis: New challenges from data at local level*. Springer.

Sforzi, F. (2012). "From administrative spatial units to local labour market areas", Chap. 1, p. 3–21. In: E. Fernández, F. Rubiera (Eds.) *Defining the spatial scale in modern economic analysis: New challenges from data at local level.* Springer.

Shearmur, R., & Doloreux, D. (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), 333–355.

Spence, M., Annez, P. C., & Buckley, R. M. (2009). *Urbanization and Growth.* Washington, D.C: Commission on Growth and Development, The World Bank.

Statistics New Zealand. (2009). *Workforces on the move: Commuting patterns in New Zealand.* Statistics New Zealand. Retrieved from 14 January 2014. http://www.stats.govt.nz/browse_for_stats/people_and_communities/Geographic-areas/commuting-patterns-in-nz-1996-2006.aspx

Tolbert, C. M.& Sizer, M. (1996). *U.S. Commuting Zones and Local Market Areas. A 1990 Update.* ERS Staff Paper, Rural Economy Division, Economic Research Service, U.S. Department of Agriculture. Retrieved from 10 January 2014. https://usa.ipums.org/usa/resources/volii/cmz90.pdf

Uchida, H., & Nelson, A. (2010). Agglomeration index: Towards a new measure of urban concentration. In J. Beall, B. Guha-Khasnobis, & R. Kanbur (Eds.), *Urbanization and development.* Oxford: Oxford University Press.

USDA ERS. (2012). "Commuting Zones and Labor Market Areas: Documentation" (Web document), Economic Research Department, U.S. Department of Agriculture. Retrieved from 14 January 2014. http://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas/documentation

Weber, A. (1909). *Über den Standort der Industrien.* Mohr, TÏbingen; translated by Friedrich, C. J. (1929) as *Alfred Weber's Theory of the Location of Industries*, Chicago, IL: University of Chicago Press.

Wernerheim, M., & Sharpe, C. (2003). High-order producer services in metropolitan Canada: How footloose are they? *Regional Studies, 37*, 469–490.

Wheaton, W., & Shishido, H. (1981). Urban concentration, agglomeration economies, and the level of economic development. *Economic Development and Cultural Change, 30*, 17–30.

Zhu, Nong, Xubei Luo and Heng-fu Zou (2012): *Regional differences in China's urbanization and its determinants*, CEMA Working Papers 535, China Economics and Management Academy, Central University of Finance and Economics. Retrieved from 21 January 2014. http://ideas.repec.org/p/cuf/wpaper/535.html