The Balassa Index meets the Theil Index: a Decomposition Methodology for Location Studies

Eleonora Cutrini

Faculty of Economics *'G. Fuà'* - Università Politecnica delle Marche
Piazzale Martelli 8, 60121, Ancona - Italy
cutrini@dea.unian.it

Abstract

This paper focuses on the decomposition of location measures according to a twofold geographical level of analysis. The methodology, based on the use of the Theil dissimilarity index, is potentially applicable to the assessment of the spatial distribution of several economic phenomena when different hierarchical levels of analysis are considered.
In the specific case, the entropy index considered is first defined as the weighed average of the log of regional Balassa indices. Thanks to the decomposability typical of the generalised entropy indices, the index constructed allows to disentangle the relative importance of country comparative advantages from regional competitiveness.
The decomposition of relative concentration indices provides a rigorous method to quantify the cross-country divergence in localisation from the inner-country agglomeration patterns. Similarly, the construction of specialisation measures is intended to capture the different constituents at the national and sub-national levels. The paper ends with the decomposition of typical entropy obtained condensing relative concentration and relative specialisation into a single index.

Keywords: Concentration, Specialisation, Balassa Index, Theil dissimilarity index, Typical entropy

JEL classification: C43, L16, 018, R12

Introduction

The use of entropy indices is common in the income distribution literature but they have been less used in spatial contexts. Although a number of authors have used the entropy measures to describe the spatial inequality in the aggregate income in Europe (e.g. De la Fuente and Vives (1995), Duro and Esteban (1998), Duro (2001), Akita (2000), Combes e Overman (2003)) their application to industrial concentration analysis has been rare (Brülhart e Traeger (2004), Aiginger e Pfaffermayr (2004), Aiginger e Davies (2004)). In this paper a decomposition methodology of sectoral concentration and typical entropy is presented.

Sala-I-Martin realised an analogous decomposition in order to distinguish between 'within country' and 'across country' components of global per capita income inequality. In the inequality literature the individual is taken as basic unit of analysis and the country represents the reference '*meso-level*' to disentangle the within and between components of the overall world income inequality (Sala-I-Martin (2002)). In regional studies instead, the income inequalities have been assessed relying on a spatial hierarchical structure (country-region levels) (Akita (2000), Brülhart e Traeger (2004)). Brülhart e Traeger (2004)'s empirical analysis is particularly connected to the decomposition that follows. Their work stands out with reference to the previous literature because of the specific use of entropy indices to evaluate geographic concentration together with Aiginger e Davies (2004)[1]. The major pitfall of their work is that the presentation of the method used is restricted to the decomposition property of general entropy measures by population subgroups without unrevealing the actual methodology used as Sala-I-Martin does (Sala-I-Martin (2002)).

The methodology contribution developed in this paper goes beyond their analysis in many respects. First, because it sheds light on the actual decomposition needed to separate the different parts of sectoral relative concentration - 'within country' and 'between country' components. It also gives straightforward indications to grasp the economic meaning of the decomposition analysis. Indeed, splitting the different components up allows to distinguish the magnitude of national and internal regional comparative advantages that may be the result of European economic integration process.

Secondly, because the entropy dissimilarity index is also used to assess relative specialisation. In particular, a region-based specialisation measure for each country is introduced which turn out to be decomposable by an inner specialisation component- a weighed average of regional specialisation indices relative to country- and a country bias.

Finally, because an index of typical entropy is constructed. The overall

[1] In Aiginger e Davies (2004) the entropy measure is used in the original version of the information of a direct message, while it is arguable that in Brülhart e Traeger (2004) the Theil index is used in the relative version, as a dissimilarity index.

entropy defines relative concentration and relative specialisation as two side of the same coin. A decomposition of the typical entropy index in 'between country' and 'within country' components is also presented.

1 Some basic insights from information theory

Theil index in the basic form is suitable to evaluate absolute concentration and absolute specialisation. It is a version strictly associated to the information content of a direct message in the mathematical communication theory (Shannon e Weaver (1949), Theil (1967)).

The information content of a *direct message* - the Theil entropy measure- is defined as:

$$I_{dm} = \sum_{i=1}^{n} p_i \ln(\frac{1}{p_i}) \tag{1}$$

Where p_i represents the probability that an event E of a distribution occurs which sums to 1.

$$\sum_{i=1}^{n} p_i = 1 \tag{2}$$

The maximum value which the entropy measure defined in Eq. 1 can take is $\ln(n)$, which define a situation in which the ex-ante ignorance about the qualitative outcome of an event E is maximum.

In order to find its maximum, it is necessary to maximize this function subject to $\sum_{i=1}^{n} p_i = 1$ to find its maximum. Therefore the Lagrangian expression is considered:

$$-\sum_{i=1}^{n} p_i * \ln(p_i - \lambda(\sum_{i=1}^{n} p_i - 1) \tag{3}$$

where λ is the Lagrangian multiplier. By differentiating with respect to p_i and putting the result equal to zero:

$$-1 - \ln(p_i) - \lambda = 0 \tag{4}$$

It is possible to conclude that $\ln(p_i) = -1 - \lambda$ for each i so that the probabilities should be all the same and hence equal to $\frac{1}{n}$. Thus the expected information takes its maximum value ($\ln(n)$) for the fully diversified case, in other words, when all events have the same chance (p. 25, Theil (1967)).

The *normalised Theil index* - a sort of deviation of the observed entropy from the uniform distribution (maximum entropy) - is defined as follow:

$$nT = \ln(n) - \sum_{i=1}^{n} p_i \ln(\frac{1}{p_i}) \tag{5}$$

$$nT = \ln(n) + \sum_{i=1}^{n} p_i \ln(p_i) \tag{6}$$

A competing measure of normalised entropy can be constructed as follows (Troutt e Acar (2004)):

$$nT = -(\ln(n))^{-1} \sum_{i=1}^{n} p_i \ln(p_i) \tag{7}$$

The measure of the information content of an *indirect message* as defined in the Information Theory (Shannon e Weaver (1949)) provided economists with a particularly valuable tools of descriptive analysis. The dissimilarity index introduced by Theil into economics (Theil (1967)) refers to a measure of the information content of an indirect message:

$$I_{im} = \sum_{i=1} \sum_{j=1} q_i \ln(\frac{q_i}{p_i}) \tag{8}$$

The indirect message is a generalisation of the direct message. In the latter case, the rationale of the message is noticing that one of the events has actually occurred while in the former case the message modify the probabilities of each event of the distribution to take place. p_i represents the *ex-ante* probability and q_i corresponds to the *ex-post* probability that the event i will occur.

The direct message index may be categorized as a dissimilarity index between the actual distribution and the theoretic uniform distribution.

The entropy index in the general form of indirect message is interpretable as a dissimilarity index and it can be used to assess the difference in two contemporary structures. It may be considered as a 'distance index' (Maasoumi (1993)), as it does not satisfy some properties necessary to be defined as a proper distance or metrics[2].

The main advantage in the use of the Theil index is its sensitiveness to minimal structural changes which allow for a throughout analysis of both sectoral concentration and regional diversification.

[2]A metrics or a distance function, defined on X*X (X is a not empty set), with $x, y, z \subseteq X$, satisfies the following axiomatic principles (See Lipschutz (1994)):
1)$d(x, y) \geq 0$;
2)$d(x, y) = d(y, x)$;
3)If $d(x, y) = 0$ then $x = y$;
4) $d(x, y) + d(y, z) \geq d(x, z)$. Theil index satisfies only the principles 1) and 3).

2 Absolute and relative entropy indices

2.1 Notations and definitions

Let us first define the variables and identify the subscripts used in the analysis of sectoral concentration and regional specialisation. L is the employment, value added would also be suitable as activity indicator in specialisation analysis but employment is preferred for it allows cross-country comparisons and a higher data availability.

The following notation is used. The subscript i, j and k identify the country, the region and the sector respectively. r_i is the number of regions inside country i, m is the number of countries, n is the number of sectors.

For instance:

- L_{ijk} is the employment of region j in country i in manufacturing sector k
- L_{ij} is total manufacturing employment of region j in country i ($L_{ij} = \sum_{k=1}^{n} L_{ijk}$)
- L_{ik} is the employment of country i in manufacturing sector k ($L_{ik} = \sum_{j=1}^{r_i} L_{ijk}$)
- L_i is total manufacturing employment of country i($L_i = \sum_{k=1}^{n} \sum_{j=1}^{r_i} L_{ijk}$)
- L_k is the employment of all regions in the area in manufacturing sector k ($L_k = \sum_{i=1}^{m} \sum_{j=1}^{r_i} L_{ijk}$)
- L is the total manufacturing employment of all regions ($L = \sum_{i=1}^{m} \sum_{j=1}^{r_i} L_{ij}$)

Sectoral shares are defined as follows:

- $v_{ijk} = \frac{L_{ijk}}{L_{ij}}$
- $v_{ik} = \frac{L_{ik}}{L_i}$
- $v_k = \frac{L_k}{L}$
- $v^*_{ijk} = \frac{L_{ijk}}{L}$
- $v^*_{ik} = \frac{L_{ik}}{L}$

and country and regional shares are defined as follows:

- $s_i = \frac{L_i}{L}$
- $s_{ij} = \frac{L_{ij}}{L}$

- $s_{ik} = \frac{L_{ik}}{L_k}$

- $s_{ijk} = \frac{L_{ijk}}{L_k}$

- $s^*_{ij} = \frac{L_{ij}}{L_i}$

- $s^*_{ijk} = \frac{L_{ijk}}{L_{ik}}$

Finally, Location Quotients or Balassa Indices are defined as follows:

$$LQ_{ik} = BI_{ik} = \frac{v_{ik}}{v_k} = \frac{s_{ik}}{s_i} \tag{9}$$

$$LQ_{ijk} = BI_{ijk} = \frac{v_{ijk}}{v_{ik}} = \frac{s^*_{ijk}}{s^*_{ij}} \tag{10}$$

$$LQ^*_{ijk} = BI^*_{ijk} = \frac{v_{ijk}}{v_k} = \frac{s_{ijk}}{s_{ij}} \tag{11}$$

Location quotient is a particularly useful piece of information to evaluate both regional specialisation and sectoral concentration. It has been used to assess country structure of comparative advantages, with a $BI_{ik} > 1$ referring to a comparative advantage of country i in sector k or, in sectoral concentration terms, a localisation of sector k in country i.

In our analysis it has been used in a version suitable to account for both overall and *internal* regional localisation. If $LQ^*_{ijk} > 1$ then region ij exhibits a comparative advantage in sector k with respect to the EU area considered as a benchmark. Specularly, turning to the concentration side, this means that sector k is more localised in region ij relative to EU.

LQ_{ijk} are related to the internal geography of regional competitive advantages so that for example if BI_{ijk} takes a value higher than 1, this informs of a regional comparative advantage of region ij in sector k relative to the reference country. In other words, sector k is localised in region ij compared with the country.

2.2 Absolute indices

Absolute concentration refers to the distribution of country (region) shares in production in a specific industry, without comparing the distribution with that of country (region) shares in total manufacturing (as relative concentration does).

If one substitutes s_{ijk} (the regional shares of the European-wide sectoral employment) for p_i (the probabilities that an event E will take place in the information theory) in equation 5 the absolute concentration index is obtained:

$$T_k^{AC} = \ln(m * r_i) - \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln(\frac{1}{s_{ijk}}) \quad (12)$$

The index defined in equation 12 informs about the extent to which a sector k spread across the regions considered in the analysis. The higher is the value of the index the more sector k is concentrated across regions. A decreasing value of the index over time is an indicator of a process of interrregional dispersion.

For the regional specialisation analysis the Theil index may be used in the 'direct message' version as a *specialisation* summary measure of region j in country i. It evaluates the degree of diversification in the internal economic structure. Substituting v_{ijk} (the sectoral shares of the region considered) for p_i in equation 5 we obtain:

$$T_{ij}^{AS} = \ln(n) - \sum_{k=1}^{n} v_{ijk} \ln(\frac{1}{v_{ijk}}) \quad (13)$$

The higher is the value of the index of absolute specialisation (AS) the higher is the level of specialisation of the regional economy.

Both absolute concentration and absolute specialisation indices measure how different the distribution of employment shares is from a uniform distribution (across country or across sectors respectively).

2.3 Relative indices

The dissimilarity version of the Theil index is particularly useful to the study of the spreading of comparative advantages across Europe. It may be used in two different versions: as a measure of relative concentration, and as a measure of relative specialisation.

2.3.1 *Relative concentration*

If one substitutes s_{ij} for p_i and s_{ijk} for q_i in equation 8, equation 14 is obtained.

$$T_k^{RC} = \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln(\frac{s_{ijk}}{s_{ij}}) = \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln(BI_{ijk}^*) \quad (14)$$

A first intuition of the meaning of this summary statistic is related to its nature of dissimilarity index. I_k^{RC} gives information about the different regional spreading of the manufacturing sector k with respect to the spatial location of the manufacturing as a whole. If the regional allocation of sector k overlaps exactly the regional distribution of the aggregate then I_k^{RC} will be 0.

2.3.2 *Relative specialisation*

If one substitutes v_k for p_i and v_{ijk} for q_i in equation 8, the following equation is obtained:

$$T_{ij} = \sum_{k=1}^{n} v_{ijk} \ln(\frac{v_{ijk}}{v_k}) = \sum_{k=1}^{n} v_{ijk} \ln(BI^*_{ijk}) \tag{15}$$

Analogously, if one takes the country as the benchmark geographical unit -instead of the European area as a whole - the relative specialisation index of region j of country i becomes:

$$T^c_{ij} = \sum_{k=1}^{n} v_{ijk} \ln(\frac{v_{ijk}}{v_{ik}}) \tag{16}$$

Finally, if one aims to assess country i sectoral diversification relative to European sectoral structure:

$$T_i = \sum_{k=1}^{n} v_{ik} \ln(\frac{v_{ik}}{v_k}) \tag{17}$$

To conclude, the relative specialisation index informs about the dissimilarity in the sectoral composition of each region compared with the structure of the selected benchmark. Three different indices of relative specialisation are identified because it is well known that the degree of diversification of a region with respect of a benchmark is strictly related to the benchmark one refers to. Clearly the choice of the benchmark geographical unit depends upon the scope of the analysis. If T_{ij} takes the value 0 then region j located in country i has a manufacturing structure which is identical to the EU-15 average. Similarly, if $T^c_{ij} = 0$ then the manufacturing distribution across sectors of region j located in country i mirrors the manufacturing structure of country i. Similarly, if $T_i = 0$ then the manufacturing distribution of country i matches the EU distribution across sectors. The higher the index, the more dissimilar the regional manufacturing structure is from the geographical unit choosen as a benchmark.

Bearing in mind equations 9, 10 and 11 it can be noticed that:

$$T_{ij} = \sum_{k=1}^{n} v_{ijk} \ln(\frac{v_{ijk}}{v_k}) = \sum_{k=1}^{n} v_{ijk} \ln(\frac{s_{ijk}}{s_{ij}}) = \sum_{k=1}^{n} v_{ijk} \ln(BI^*_{ijk}) \tag{18}$$

$$T^c_{ij} = \sum_{k=1}^{n} v_{ijk} \ln(\frac{v_{ijk}}{v_{ik}}) = \sum_{k=1}^{n} v_{ijk} \ln(\frac{s^*_{ijk}}{s^*_{ij}}) = \sum_{k=1}^{n} v_{ijk} \ln(BI_{ijk}) \tag{19}$$

$$T_i = \sum_{k=1}^{n} v_{ik} \ln(\frac{v_{ik}}{v_k}) = \sum_{k=1}^{n} v_{ik} \ln(\frac{s_{ik}}{s_i}) = \sum_{k=1}^{n} v_{ik} \ln(BI_{ik}) \tag{20}$$

This means that each relative specialisation index is interpretable as a weighed average of the log of the location quotients in each sector, weighed by the importance of each sector in the specific region analysed.

3 Decomposing the relative concentration index

3.1 Decomposition methodology

In the previous section it has been stated that relative concentration refers to the dissimilarity in the localisation of each sector k with respect to the spreading of the overall manufacturing sector across the spatial units considered (countries, regions). If a sector k spreads exactly proportionally to total manufacturing employment the relative concentration index will exhibits a nil value.

$$T_k = \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln \frac{v_{ijk}}{v_k} \tag{21}$$

Adding and subtracting the term $\sum_{i=1}^{m} s_{ik} \ln(v_{ik})$ to equation 21 the following equation is obtained:

$$T_k = \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln \frac{v_{ijk}}{v_k} + \sum_{i=1}^{m} s_{ik} \ln(v_{ik}) - \sum_{i=1}^{m} s_{ik} \ln(v_{ik}) \tag{22}$$

$$T_k = \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln(v_{ijk}) - \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln(v_k) + \sum_{i=1}^{m} s_{ik} \ln(v_{ik}) - \sum_{i=1}^{m} s_{ik} \ln(v_{ik}) \tag{23}$$

and because $\sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} = \sum_{i=1}^{m} s_{ik}$

$$T_k = \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln(v_{ijk}) - \sum_{i=1}^{m} s_{ik} \ln(v_k) + \sum_{i=1}^{m} s_{ik} \ln(v_{ik}) - \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln(v_{ik}) \tag{24}$$

Combining the second and the third elements the *between country* component is obtained:

$$T_k^{bc} = \sum_{i=1}^{m} s_{ik} \ln \frac{v_{ik}}{v_k} \tag{25}$$

instead, the *within country* component is obtained putting together the first element of 24 with the forth one:

$$T_k^{wc} = \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln(v_{ijk}) - \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln(v_{ik}) \tag{26}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln \frac{v_{ijk}}{v_{ik}} \tag{27}$$

so that

$$T_k = T_k^{bc} + T_k^{wc} \tag{28}$$

$$T_k = \sum_{i=1}^{m} s_{ik} \ln \frac{v_{ik}}{v_k} + \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ijk} \ln \frac{v_{ijk}}{v_{ik}} \tag{29}$$

The decomposition can also be envisaged in a different manner. Particularly, the Theil within countries (T_k^{wc}) is interpretable as a weighed average of the Theil index between regions inside each country considered in the analysis.

The relative concentration index of sector k -RC index- is defined as[3]:

$$T_k = \sum_{i=1}^{m} \sum_{j=1}^{r_i} \frac{L_{ijk}}{L_k} \ln \frac{\frac{L_{ijk}}{L_k}}{\frac{L_{ij}}{L}} \tag{30}$$

Then, Theil index, defined in equation 30, can be decomposed as follows

$$T_k = T_k^{wc} + T_k^{bc} \tag{31}$$

Where

$$T_k^{wc} = \sum_{i=1}^{m} \frac{L_{ik}}{L_k} T_{ik}^{br} = \tag{32}$$

$$= \sum_{i=1}^{m} \frac{L_{ik}}{L_k} \sum_{j=1}^{r_i} \frac{L_{ijk}}{L_{ik}} \ln \frac{\frac{L_{ijk}}{L_{ik}}}{\frac{L_{ij}}{L_i}} = \tag{33}$$

$$= \sum_{i=1}^{m} s_{ik} \sum_{j=1}^{r_i} s_{ijk}^{*} \ln \frac{s_{ijk}^{*}}{s_{ij}^{*}} \tag{34}$$

and

$$T_k^{bc} = \sum_{i=1}^{m} \frac{L_{ik}}{L_k} \ln \frac{\frac{L_{ik}}{L_k}}{\frac{L_i}{L}} = \tag{35}$$

$$= \sum_{i=1}^{m} s_{ik} \ln \frac{s_{ik}}{s_i} \tag{36}$$

Note that equation 34 corresponds to equation 27 and equation 36 corresponds to equation 25.

[3]Note that equation 30, equation 21 and equation 41 are equivalent.

The first term (defined in equation 32) is the Theil measure of 'within-country relative concentration'. It is a weighed average of the Theils within each country, where the weights are the share of the country in overall EU sector *k*. Conversely, the second term (defined in equation 35) measures the usual 'across-country relative concentration'.

3.2 Relative concentration measure as a weighed average of the log of regional Balassa indices

One of the widespread way to assess regional specialisation and geographical concentration of sectors is using Location Quotient or Balassa Indices. Different approaches have been adopted to measuring specialisation, a first wave of empirical studies used trade data on exports and imports (Aquino (1978), Sapir (1996)) while an alternative branch of research has adopted a more descriptive approach less directly connected to theory (Kim (1995), Amiti (1999), Brülhart (2001), Haaland *et al.* (1999)). Here, it has been referred to this second method which relies on production-based measure of specialisation (Location Quotient) instead of the trade-based measures of Revealed Comparative Advantage (RCA).

Theil dissimilarity index of relative concentration introduced in the previous paragraph is a summary statistic of Balassa Indices weighted by sectoral regions' shares (s_{ijk}) to get a measure of industry specialisation. The aim of this section is to show how the Theil dissimilarity index presented in the previous section is related to the Balassa Index or Location Quotient. This relationship is particularly interesting. Indeed, all the Generalized Entropy Indices -namely the Mean Logarithmic Deviation (GEI(0)), the Theil's Index (GEI(1)), one half the square of the coefficient of variation (GEI(2))- satisfy a set of desirable axiomatic principles: the Pigou-Dalton Transfer Principle, the Scale Independence Principle, the Principle of Population and the Principle of Decomposability (Cowell (1995)).

In particular, thanks to the decomposition property of the Theil index in between and within groups components, it is possible to distinguish between country comparative advantages and within country regional comparative advantages in each sector k.

The Location Quotients or Balassa Indices defined in equations 9, 10, 11 are considered. The following relation holds:

$$BI^*_{ijk} = BI_{ik}BI_{ijk} \tag{37}$$

Thus:

$$\sum_{i=1}^{m}\sum_{j=1}^{r_i} s_{ijk}\ln(BI^*_{ijk}) = \sum_{i=1}^{m}\sum_{j=1}^{r_i} s_{ijk}\ln(BI_{ijk}BI_{ik}) = \tag{38}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{r_i} s_{ijk}\ln(BI_{ijk}) + \sum_{i=1}^{m}\sum_{j=1}^{r_i} s_{ijk}\ln(BI_{ik}) = \tag{39}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{r_i} s_{ijk}\ln(BI_{ijk}) + \sum_{i=1}^{m} s_{ik}\ln(BI_{ik}) \tag{40}$$

In conclusion:

$$T_k = \sum_{i=1}^{m}\sum_{j=1}^{r_i} s_{ijk}\ln(BI^*_{ijk}) \tag{41}$$

$$T_k^{wc} = \sum_{i=1}^{m}\sum_{j=1}^{r_i} s_{ijk}\ln(BI_{ijk}) \tag{42}$$

$$T_k^{bc} = \sum_{i=1}^{m} s_{ik}\ln(BI_{ik}) \tag{43}$$

No upper bound exists for Balassa Index whose lower limit is 0. As a consequence, relative concentration Index has no upper bound and the lower limit is 0.

- If $T_k = 0$ it must be that $BI^*_{ijk} = 1$ for each regions in the area; in this case the localisation of manufacturing sector k overlaps the distribution of the overall manufacturing sector so that in sector k no region shows neither a comparative advantage nor a comparative disadvantage with respect to the overall area (in our case the European Union), $T_k = 0$ occurs when $s_{ijk} = s_{ij}$ for each region, so that the sector k is distributed across the European regions in the same way as the total manufacturing sector span across the same regions. $T_k = 0$ is thus a sort of no RC-benchmark.

- If $T_k^{wc} = 0$ then $BI_{ijk} = 1$ for each region; no internal region exhibits a comparative advantage (disadvantage) in sector k compared with the national counterpart($s^*_{ijk} = s^*_{ij}$), $T_k^{wc} = 0$ defines a benchmark of *no-relative concentration within countries*;

- If $T_k^{bc} = 0$ it must be that $BI_{ik} = 1$ for each i; no country has a comparative advantage (disadvantage) in sector k with respect to the overall area ($s_{ik} = s_i$), $T_k^{bc} = 0$ defines a benchmark of *no-relative concentration between countries*;

- $T_k = 0$ implies that both $T_k^{wc} = 0$ and $T_k^{bc} = 0$

- Either $T_k^{wc} = 0$ or $T_k^{bc} = 0$ does not imply that $T_k = 0$.

- If $\ln(BI_{ijk}) < 0$ then region j of country i is despecialised in sector k;
- If $\ln(BI_{ijk}) > 0$ then region j of country i is specialised in sector k;
- If $\ln(BI_{ijk}) = 0$ then region j of country i is neither specialised nor despecialised in sector k.

The higher the value of total relative concentration index (T_k) is, the more the allocation of regional comparative advantages in sector k compared with the EU is uneven. An increasing total relative concentration index over time denotes a process of regional specialisation in that sector somewhere in Europe.

If T_k^{wc} is 0 then sector k is proportionally distributed to total manufacturing employment in the internal regions of each country. Put it differently, a nil value of 'within' relative concentration suggests that no internal region comparative advantage with respect to the reference country exists. The higher the domestic component is, the more the inner allocation of comparative advantages of each country is uneven. An increasing value of the 'within country' factor is related to a process of rising diversification internal to the countries.

The 'across country' factor embodies the importance of national comparative advantages with respect to the whole area considered as a benchmark (Europe in our case). A nil value of T_k^{bc} implies that the across countries distribution of sector k overlaps perfectly the allocation across countries of manufacturing as a whole. In other words, countries reveal neither a comparative advantage nor a comparative advantage in the specific sector k analysed. Accordingly, the higher the 'between country' component is, the more the allocation of national comparative advantages in sector k is unbalanced. An increase in the 'between country' component of relative concentration indicates an increasing unequal allocation of comparative advantages, associated to a process of country specialisation.

4 A region-based definition of country specialisation

The aim of this section is to change the usual perspective and to demonstrate how country specialisation (T_i) can be seen as a residual of the averaged regional specialisation to EU in each country once the dissimilarity of the regional manufacturing structures with reference to the country has been accounted for. The relationship between patterns of country specialisation and regional ones. It is worth noting that regional specialisation does not necessarily go hand in hand with country specialisation. For example, despite Spanish regions experienced a very small changes in specialisation relative to the country (Paluzie *et al.* (2001)), Spain is one of the country which become more specialised with respect to other European countries (Midelfart-Knarvik *et al.* (2002)). In contrast, Italian internal landscape seems to replicate the pattern of specialisation of the country with respect to the EU. Indeed, regional specialisation slightly increased in a majority of regions with respect to the country (de Robertis (2001)) as a whole and Italy is becoming more specialised with respect to the EU (Midelfart-Knarvik *et al.* (2002)). As Combes e Overman (2003) suggested, decomposing changes into within and between nations components, should be appropriate to investigate EU specialisation patterns and disentangle the mixed trend of change in specialisation.

So far the literature lack of a rigorous methodology to account for such a separate analysis. To this end, a way of taking into account two hierarchical levels (regions and countries) is now presented.

Let us first define the weighed average of the regional specialisation indices for each country as follows:

$$avRS_i^{EU} = \sum_{j=1}^{r_i} T_{ij} s_{ij}^* \tag{44}$$

$$avRS_i^{c} = \sum_{j=1}^{r_i} T_{ij}^{c} s_{ij}^* \tag{45}$$

Country relative specialisation to EU (T_i) is nothing more than the difference between the two country-based averaged regional specialisation measures:

$$T_i = \sum_{j=1}^{r_i} (T_{ij} - T_{ij}^{c}) s_{ij}^* \tag{46}$$

$$= \sum_{j=1}^{r_i} (\sum_{k=1}^{n} v_{ijk} \ln(\frac{v_{ik}}{v_k})) s_{ij}^* \tag{47}$$

$$= \sum_{j=1}^{r_i} \sum_{k=1}^{n} v_{ijk} s_{ij}^* \ln(\frac{v_{ik}}{v_k}) \tag{48}$$

$$= \sum_{j=1}^{r_i} \sum_{k=1}^{n} \frac{L_{ij}}{L_i} \frac{L_{ijk}}{L_{ij}} \ln(\frac{v_{ik}}{v_k}) \tag{49}$$

$$= \sum_{j=1}^{r_i} \sum_{k=1}^{n} \frac{L_{ijk}}{L_i} \ln(\frac{v_{ik}}{v_k}) \tag{50}$$

$$= \sum_{k=1}^{n} \frac{L_{ik}}{L_i} \ln(\frac{v_{ik}}{v_k}) \tag{51}$$

Changing the usual viewpoint, the actual country specialisation is defined in a region-based way. It is the weighed average of the regional specialisation indices relative to EU ($avRS_i^{EU}$) and it turns out to be constituted by two elements: the inner country component ($avRS_i^c$), which accounts for the internal regional specialisation relative to the country, and the country bias, in other words the country specialisation relative to EU (T_i).

$$avRS_i^{EU} = avRS_i^c + T_i \tag{52}$$

To conclude, also the weighed average of regional specialisation indices of each country ($avRS_i^{EU}$) is subject to an interesting decomposition to disentangle the within country component of regional specialisation from the purely country one.

5 Typical entropy: relative concentration and relative specialisation as two side of the same coin

5.1 Definition of typical entropy

As a fist step in the analysis of the relationship between RS and RC, it is worth noting that equations 18, 19 and 20 differ from equations 41, 42, 43 only for the weights applied.

Let us first remind that:

$$v^*_{ijk} = v_k s_{ijk} = s_{ij} v_{ijk} = \frac{L_{ijk}}{L} \tag{53}$$

Where v_k are the industry shares of EU aggregate manufacturing and s_{ijk} represent the regional shares of EU aggregate manufacturing.

and

$$v^*_{ik} = v_k s_{ik} = s_i v_{ik} = \frac{L_{ik}}{L} \tag{54}$$

Typical entropy can be defined as follows, using Balassa indices weighted by the sectoral regional shares of EU aggregate manufacturing (v^*_{ijk}):

$$E = \sum_{k=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{r_i} v^*_{ijk} \ln(BI^*_{ijk}) \tag{55}$$

$$= \sum_{k=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ij} v_{ijk} \ln(BI^*_{ijk}) = \sum_{k=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{r_i} v_k s_{ijk} \ln(BI^*_{ijk}) \tag{56}$$

Substituting equations 18 and 41 into equation 56:

$$E = \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ij} T_{ij} = \sum_{k=1}^{n} v_k T_k \tag{57}$$

Clearly, typical entropy is interpretable as a summary statistics of either relative concentration indices or relative specialisation indices weighted by regional shares (s_{ij}) and EU sectoral shares (v_k) of EU aggregate manufacturing, respectively.

5.2 Decomposing typical entropy

$$E = \sum_{k=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{r_i} v^*_{ijk} \ln(BI^*_{ijk}) \tag{58}$$

Substituting equation 37 into equation 58:

$$\sum_{k=1}^{n} \sum_{i=1}^{m} \sum_{j=1}^{r_i} v^*_{ijk} \ln(BI_{ijk} BI_{ik}) \tag{59}$$

$$\sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} v_{ijk}^* \ln(BI_{ijk}^*) = \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} v_{ijk}^* \ln(BI_{ijk}) + \sum_{k=1}^{n}\sum_{i=1}^{m} v_{ik}^* \ln(BI_{ik}) \tag{60}$$

The same result is obtained by adding and subtracting the term $\sum_{k=1}^{n}\sum_{i=1}^{m} v_{ik}^* \ln(v_{ik})$ to equation 58 as follows:

$$\sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} v_{ijk}^* \ln(BI_{ijk}^*) + \sum_{k=1}^{n}\sum_{i=1}^{m} v_{ik}^* \ln(v_{ik}) - \sum_{k=1}^{n}\sum_{i=1}^{m} v_{ik}^* \ln(v_{ik}) \tag{61}$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} \frac{L_{ijk}}{L} \ln(v_{ijk}) - \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} \frac{L_{ijk}}{L} \ln(v_k) + \sum_{k=1}^{n}\sum_{i=1}^{m} \frac{L_{ik}}{L} \ln(v_{ik}) - \sum_{k=1}^{n}\sum_{i=1}^{m} \frac{L_{ik}}{L} \ln(v_{ik}) \tag{62}$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} \frac{L_{ijk}}{L} \ln(v_{ijk}) - \sum_{k=1}^{n}\sum_{i=1}^{m} \frac{L_{ik}}{L} \ln(v_k) + \sum_{k=1}^{n}\sum_{i=1}^{m} \frac{L_{ik}}{L} \ln(v_{ik}) - \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} \frac{L_{ijk}}{L} \ln(v_{ik}) \tag{63}$$

Combining the first with the forth terms and the second with the third:

$$= \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} \frac{L_{ijk}}{L} \ln(\frac{v_{ijk}}{v_{ik}}) + \sum_{k=1}^{n}\sum_{i=1}^{m} \frac{L_{ik}}{L} \ln(\frac{v_{ik}}{v_k}) \tag{64}$$

$$\sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} v_{ijk}^* \ln(BI_{ijk}^*) = \sum_{k=1}^{n}\sum_{i=1}^{m} v_{ik}^* \ln(BI_{ik}) + \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} v_{ijk}^* \ln(BI_{ijk}) \tag{65}$$

The 'between country' component of typical entropy is:

$$E^{bc} = \sum_{k=1}^{n}\sum_{i=1}^{m} v_{ik}^* \ln(BI_{ik}) \tag{66}$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{m} s_i v_{ik} \ln(BI_{ik}) = \sum_{k=1}^{n}\sum_{i=1}^{m} v_k s_{ik} \ln(BI_{ik}) \tag{67}$$

$$= \sum_{i=1}^{m} s_i T_i = \sum_{k=1}^{n} v_k T_k^{bc} \tag{68}$$

The 'within country' component of total entropy is:

$$E^{wc} = \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} v_{ijk}^* \ln(BI_{ijk}) \tag{69}$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} s_{ij} v_{ijk} \ln(BI_{ijk}) = \sum_{k=1}^{n}\sum_{i=1}^{m}\sum_{j=1}^{r_i} v_k s_{ijk} \ln(BI_{ijk}) \tag{70}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ij} T_{ij}^{c} = \sum_{k=1}^{n} v_k T_k^{wc} \tag{71}$$

Indeed, total entropy is interpretable as a summary statistics of relative specialisation indices (T_{ij}) weighted by the manufacturing regional shares (s_{ij}):

$$\sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ij} T_{ij} = \sum_{i=1}^{m} s_i T_i + \sum_{i=1}^{m} \sum_{j=1}^{r_i} s_{ij} T_{ij}^{c} \tag{72}$$

Similarly, total entropy can be seen as a summary statistics of relative concentration Theil indices (T_k) weighted by the EU industry shares (v_k):

$$\sum_{k=1}^{n} v_k T_k = \sum_{k=1}^{n} v_k T_k^{bc} + \sum_{k=1}^{n} v_k T_k^{wc} \tag{73}$$

Referring to the previous paragraph where a decomposition between averaged regional specialisation measures by country has been presented, it is worth noting that, the different components of typical entropy in that context can be expressed as follows:

$$E = \sum_{i=1}^{m} avRS_i^{EU} * s_i \tag{74}$$

$$E^{wc} = \sum_{i=1}^{m} avRS_i^{c} * s_i \tag{75}$$

$$E^{bc} = \sum_{i=1}^{m} T_i * s_i \tag{76}$$

To conclude, total entropy lets us to see relative specialisation and relative concentration as two sides of the same coin. On one side typical entropy is a sort of averaged dissimilarity index obtained by applying to each RC index (T_k) the corresponding industry shares of EU aggregate manufacturing (v_k). On the other side, typical entropy is a weighed dissimilarity index of relative specialisation indices (T_{ij}) weighted by the regional shares of EU manufacturing (s_{ij}).

In terms of decomposition properties, the conglomeration of relative concentration and relative specialisation into a unique entropy index allows to operate the decomposition in geographical units ('within' and 'between country' components) at this level which was also feasible for the RC side, but not for the RS side[4].

[4]Relative specialisation index is decomposable by sectors groups: e.g. high-skilled and low-skilled.

Besides, and more interestingly, it allows to have an idea of the level of average RS ad RC respectively. As both RS and RC indices have no upper bound, typical entropy represents an average value of RS and RC. Typical RS informs about the average dissimilarity between the distribution across geographical units of sectors and the location across geographical units of overall EU manufacturing. Similarly, typical RC gives an idea of the average dissimilarity between the regional manufacturing structure across sectors and the EU manufacturing structure across sectors.

6 Concluding remarks

We have provided a methodology based on the use of additively decomposable entropy indices in location studies. The decomposition introduced allows to overcome a typical shortcoming of the existing empirical literature, namely the focus on a single geographical level of analysis. As a matter of fact, by implementing the methodology it is possible to handle two relevant hierarchical partition grids assessing the evolution of their relative importance in the patterns of localisation of economic activities. The distinction between absolute and relative location measures is presented to introduce their different connotation. While the former drawn on the merely statistical meaning of concentration, the latter are subject to a more direct economic interpretation. They denote competitiveness and the concept of comparative advantage. On one side, each relative concentration measure is divided into within and between country components. On the other side, each country specialisation index is defined as the difference between the weighted average of the regional specialisation indicators relative to the EU and the regional specialisation measures with respect to the country.

From a theoretical viewpoint, the decomposition allows to gauge the relative importance of complementary theoretical frameworks which place emphasis either on region or on country, respectively.

References

AIGINGER, K. E DAVIES, S. W. (2004). *Industrial specialisation and geographic concentration: two sides of the same coin? not for the European Union*. Forthcoming in the Journal of Applied Economics.

AIGINGER, K. E PFAFFERMAYR, M. (2004). *The Single Market and geographic concentration in Europe*. Review of International Economics, 12(1): 1–11.

AKITA, T. (2000). *Decomposing regional income inequality using two-stage nested Theil decomposition method*. Paper presented at the 6th World Congress of the Regional Science Association International, Lugano, May.

AMITI, M. (1999). *Specialisation patterns in Europe*. Weltwirtschaftliches Archiv, (135): 573–593.

AQUINO, A. (1978). *Intra-industry trade and inter-industry specialization as concurrent sources of international trade in manufactures*. Weltwirtschaftliches Archiv, (114): 275–296.

BRÜLHART, M. (2001). *Evolving geographical concentration of European Union*. Weltwirtschaftliches Archiv, 137(2): 215–243.

BRÜLHART, M. E TRAEGER, R. (2004). *An account of geographic concentration patterns in Europe*. Forthcoming in Regional Science and Urban Economics.

COMBES, P.-P. E OVERMAN, H. (2003). *The spatial distribution of economic activities in the European Union*, vol. Handbook of Urban and Regional Economics. Elsevier-North Holland, Amsterdam.

COWELL, F. A. (1995). *Measuring Inequality*. Prentice Hall, Hemel Hempstead, 2a ed.

DE ROBERTIS, G. (2001). *European integration and internal economic geography: The case of the italian manufacturing industry 1971-1991*. The International Trade Journal, XV(3): 345–371.

HAALAND, J. I., KIND, H. J., MIDELFART-KNARVIK, K. H. E TORSTENSSON, J. (1999). *What determines the economic geography of Europe?* CEPR Discussion Paper N. 2072.

KIM, S. (1995). *Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure, 1860-1987*. The Quarterly Journal of Economics, pp. 881–908.

LIPSCHUTZ, S. (1994). *Topologia*. Milano.

MAASOUMI, E. (1993). *A compendium to information theory in economics and econometrics*. Econometric reviews, 12(2): 137–181.

MIDELFART-KNARVIK, K. H., OVERMAN, H. G., REDDING, S. J. e VENABLES, A. J. (2002). *Integration and industrial specialisation in the European Union*. Revue Economique, 53: 469–481.

PALUZIE, E., PONS, J. e TIRADO, D. A. (2001). *Regional integration and specialization patterns in Spain*. Regional studies, 35(4): 285–296.

SALA-I-MARTIN, X. (2002). *The disturbing 'rise' of global income inequality*. NBER Working Paper N° 8904.

SAPIR, A. (1996). *The effects of Europe's internal market program on production and trade: a first assessment*. Weltwirtschaftliches Archiv, (132): 457–475.

SHANNON, C. E. e WEAVER, W. (1949). *The Matematical theory of Communication*. Urbana.

THEIL, H. (1967). *Economics and Information Theory*. North-Holland, Amsterdam.

TROUTT, M. D. e ACAR, W. (2004). *A Lorenz-Pareto measure of pure diversification*. European Journal of Operational Research.