ARE RELATED AND UNRELATED VARIETY SUITABLE MEASURES AT ANALYZING INDUSTRIAL SECTORAL COMPOSITION? A CRITICAL REVIEW

GIANNINI, Massimo[1]

MARTINI, Barbara[2]

FIORELLI, Cristiana[3]

Abstract

In this paper, we review some characteristics of the entropy measures. Regional sciences, particularly the Evolutionary Economic Geography approach, use such a method for investigating how knowledge spills within the industrial sectoral composition. In this approach, the total entropy (variety) is decomposed in Related and Unrelated variety. We argue that total entropy should be instead decomposed into alpha and beta components that are easier to interpret and more coherent with the mathematical foundations. Moreover, the beta entropy measures the local entity's divergenceconcerning the entire economy. This is particularly useful in the context of a spatial transmission of knowledge.

Keywords: Entropy, Variety, Industrial sectoral composition, Knowledge diffusion.

JELCode: B52, J21, L16, O33

1. Introduction

The literature on firms' sectoral composition exploits measures inherited by the entropy definition. The latter comes from thermodynamics, which measures the degree of disorder in a dynamic system. In 1948, Claude Shannon applied entropy to measure microscopic uncertainty in the random losses of information in telecommunication signals, giving birth to the Information Theory.

The entropy measures are primarily employed in the Evolutionary Economic Geography (EEG); Frenken et al. (2007) proposed a particular decomposition of the Shannon entropy index, which they called "Variety" aiming at investigating how industries sectoral composition affects the spreading of knowledge spillovers and the growth paths. Exploiting the decomposability of the index, they produced two sub-indexes, called "Related Variety" (RV) and "Unrelated Variety" (UV). The idea provides a framework for analyzing two different channels of knowledge transmission. As we will see in detail in section 5, the knowledge spills among related industries (i.e., industries whose production type is close for technology, skills, markets, organization, etc. and strongly localized) differently concerning unrelated industries, which are different for types of produced goods, technologies, skills, and markets. The intense proximity, both from a geographic and economic point of view, among related industries allows a faster specific knowledge diffusion, as happens in the Marshall-Arrow-Romer (Mars) theory of externalities (agglomeration) and Jacob's theory, closer to the Schumpeterian idea of creative-destruction.On the other hand, when industries are more heterogeneous and diffuse, the knowledge spills slowly. The RV and UV indexes are a way to investigate these two knowledge transmission channels. Frenken et al. (2007) argued that spillovers within a region are expected to occur "among related sectors primarily, and only to a limited extent among unrelated sectors (page 688)". Moreover, the unrelatedness of an economy provides a sort of "risk-immunization" to asymmetric sector-specific shocks, much like in the portfolio theory, making it more resilient (Boschma and Iammarino, 2009).

The practical measure of RV and UV requires two different levels of industry aggregation. UV works at an upper level of aggregation (a sort of macrocosm), typically the two-digit level of the Standard Industry Classification (SIC) tree, as it involves the entire economic tissue of a geographic entity.
RV needs to zoom inside (microcosm) each sector of the two-digit tree; researchers have used a three-digit or even five-digit level of the SIC.In other words, RV is a measure of the entropy of a lower tree level within a higher one. We will come back with more details in section 5.

[1]Full- Professor, University of Rome Tor Vergata, Dept. of Business Engineering, E-mail: massimo.giannini@uniroma2.it – Contact author

[2]Aggregate Professor, University of Rome Tor Vergata, Dept. of Business Engineering

[3]Research Fellow, University of Rome Tor Vergata, Dept. of Business Engineering

This paper will review how an entropy measure should be used appropriately by borrowing from other disciplines with a long and consolidated experience in utilizing the entropy as a measure of composition or diversification, particularly biology and ecology. Section 2 reviews some of these measures. Sections 3 and 4 provide more details on some properties of the HCDT entropy family. Section 5 comes back to RV and UV with more pieces. Finally, section 6 tries to reconcile sections 3, 4, and 5. Some doubts emerge on the proper interpretation of the RV measure in the EEG literature. We will use the Italian Industry Classification tree (Ateco) as a case study, elaborated by the Italian National Institute for Statistics (Istat) in 2018. Conclusions follow.

2. A Brief Review of Entropy Measures and Their Mathematical Characteristics

The entropy measure owns many functional mathematical properties, particularly the additive one, which allows the index decomposition into sub-dimensions and/or sub-categories. Moreover, an entire family of entropy measures has been adapted to several disciplines. The Theil index, used mainly to measure inequality in the personal income distribution, is one of the most known. Today there are alternative definitions of entropy stemming from different disciplines: diversity, variety, evenness, and disparity are examples. The common thread is to build an index that must have some mathematical properties. In a discrete random variable, if p_i is the probability that an event occurs, which is usually estimated by its frequency in a sample (as examples: the share of a given animal species, the share of income of a given group, the share of workers in a given industry and so on), the entropy measure of the system must have these properties:

1) Positivity: $H(P) \geq 0$
2) Expansibility: Expansion of P by a new component equal to 0 does not change $H(P)$
3) Symmetry: $H(P)$ is invariant under permutations of $p_1, \ldots, p_n$
4) Continuity: $H(P)$ is a continuous function of P (for fixed n)
5) Additivity: $H(P \times Q) = H(P) + H(Q)$
6) Subadditivity: $H(X, Y) \leq H(X) + H(Y)$
7) Strong additivity: $H(X, Y) = H(X) + H(Y|X)$
8) Recursivity: $H(p_1, \ldots, p_n) = H(p_1 + p_2, p_3, \ldots, p_n) + (p_1 + p_2)H\left(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2}\right)$
9) Sum property: $H(P) = \sum_{i=1}^{n} g(p_i)$, for some continuous function g

The logarithmic law owns such properties. The Shannon entropy index:

$$H(P) = -\sum_{i=1}^{n} p_i \, ln(p_i)$$

is an example of entropy measure, as it follows the fundamental axioms. This entropy has a maximum when events are equiprobable (or uniformly distributed) $p_1 = p_2 = \cdots = p_n = 1/n$ so that $H(P) = ln(n)$.
Henry Theil, in 1967, proposed an inequality index based on the Shannon entropy measure; by exploiting the decomposability of the index, he divided the total inequality in a “within” component (income shares or groups) and a “between” one (among income shares or groups)

However, the entropy measures have a prominent role in biology and ecology, aiming at measuring the richness, or scarcity, of species in a population, their distribution in the habitat, their tree classification characteristics. There is a plethora of indexes, which are related to each other, although with some specificity. A review is beyond the scope of this paper; the interested reader is advised to read the public text by Daly et al. (2018). However, these indexes share some common questions, such as:

1. How many categories (or species) does an entity (a territory, a country, and so on) have?
2. How much of each category does an entity have?
3. How distinct are the categories of an entity?

Entropy, variety, diversity are strictly synonymous; they share the idea of measuring how species (often grouped into categories) are distributed in a territory. Stirling (1998, 2007) categorized these three questions as variety, balance, and disparity. There are many of indexes; the following Figure 1 summarizes some of them (Guevara et al., 2016).

ID	Measure	Formula	Reference
v	Variety	$v = \sum_i (p_i^0)$	
hhi	Herfindahl–Hirschman Index	$HHI = \sum_i (p_i^2)$	Rhoades (1993)
b, gs	Blau Index, Gini-Simpson	$B = 1 - \sum_i (p_i^2) = 1 - HHI$	Blau (1977); Gini (1912)
s	Simpson	$D_S = \sum_i n_i(n_i - 1)/N_t(N_t - 1)$	Simpson (1949)
bp	Berger-Parker	$D_{BP} = \max_i (p_i)$	Berger and Parker (1970)
e	Shannon Entropy	$H = -\sum_i (p_i \log p_i)$	Shannon (1948)
ev	Pielou Evenness	$J = -\sum_i (p_i \log p_i) \,/\, \log v$	Pielou (1970)
re	Rényi-Entropy	${}^qH = (1-q)^{-1} \log \left(\sum_i p_i^q\right)$	Rényi (1961)
hcdt	HCDT Entropy	${}^qH = (q-1)^{-1} \left(1 - \sum_i p_i^q\right)$	Havrda and Charvát (1967); Tsallis (1988)
hn	Hill Numbers	${}^qD_{HN} = \left(\sum_i p_i^q\right)^{1/(1-q)}$	Hill (1973)
d	Disparity	$DIS = \sum_{ij} d_{ij}/N$	
rao	Rao	$D_{RAO} = \sum_{ij} d_{ij} p_i p_j$	Rao (1982)
rs	Rao-Stirling	$\Delta = \sum_{ij} d_{ij}{}^{\alpha} (p_i p_j)^{\beta}$	Stirling (2007)

Figure1:Guevara et al., 2016.

The first block of measures is associated mainly with the dimension's "variety" and "balance" of the diversity, while the second block presents measures that also usethe dimension "disparity". C is the set of categories present in the entity. $i, j \in C, i \neq j$ and $i, j \neq j, i$; n_i is the value of abundance and p_i the proportion of the category i in the entity. $v = n(C)$ is the number of categories present in the entity - the variety. $N = \sum_i n_i$. Here log is the logarithm usually natural, and $q, \alpha, \beta \geq 0$. When $q \rightarrow 1$, HCDT and Rény entropies converge to the Shannon entropy. Additionally, for Hill numbers, when $q \rightarrow 1$, it results in the exponential of Shannon Entropy.

Variety measures how many categories or types an entity (habitat, country, territory, and so on) has. Variety is useful as a first approach to the diversity of an entity since the number of categories (e.g., species, industries, or any type of classification) is easy to understand and calculate. Being related to the concept of variety, it is helpful in some cases to know the ubiquity or, conversely, the rareness of each category, by considering its presence in all entities. Ubiquity could also be considered as the variety of entities that each category has (Hidalgo et al., 2007). Balance measures how much of each category the entity has. The raw indicators of balance are the values of abundance or the relative abundance values, e.g. the share p_i of each i-th category. The word balance is used when the abundance values are more equally distributed across the categories. For a given variety, a more balanced system is considered more diverse. Extreme cases are those where the quantity of elements for each category is the same (i.e. perfect balance) or conversely, where all the elements are concentrated in just one category (i.e.total concentration). The analogy with industry composition is clear; when each sector is composed of the same percentage of industries, the economy is balanced.

Diversity measures related to "balance" property could be understood as statistical dispersion and are mainly a function of p_i. While some measure the evenness or heterogeneity of the distribution, such as the Blau Index, others emphasize the concentration, such as the Herfindahl-Hirschman Index (HHI). The latter, for example, computes the probability that two individuals, randomly drawn, belong to the same category. This probability is calculated with replacement; after taking the first individual into account, it is replaced with an identical one, so as neither affecting the total number of individuals in that category (n_i) nor the total amount of individuals in the entity (N_t). HHI is used in economics, for instance, to estimate the concentration of markets or wealth (Ceriani and Verme, 2012).

Considering that balance is the opposite of concentration, the Gini-Simpson Index subtracts HHI from 1 to estimate balance. The same idea is behind the Blau Index. The Blau Index was created to measure the heterogeneity of social communities and its use is very common in sociology and other social sciences.

Like HHI, Simpson measure (Ds) has the same probabilistic idea of measuring concentration, but it computes the probability without replacement - meaning that the values of N_t and n_i decreases in 1 after the first probability is calculated. This measure of concentration and its equivalent balance or index of diversity (1 - Ds) are widespread in ecology. Moreover, the reciprocal index (RS = 1/DS) can be calculated.

Shannon Entropy is a frequently used measure of balance and diversity. As pointed by Hidalgo (2015, p.17), entropy is a measure of the multiplicity of states. A high value of multiplicity of states (categories) implies more evenness and less concentration: as a consequence, the higher the variety and the balance, the higher the entropy. A generalization of the Shannon Entropy is Rényi's entropy. Rényi's entropy allows the users to give more or less relative importance to rare categories through the parameter q.

Another parameterized entropy is the HCDT entropy (Havrda and Charvát, 1967; Daróczy, 1970; Tsallis, 1988). It is noteworthy that Variety, Shannon, Blau's and Berger-Parker's indexes are special cases of HCDT (respectively with the parameter $q = 0,1,2$ and infinity).

Finally, Hill numbers (Hill, 1973) are a mathematically unified family of diversity indexes that differ only by a parameter q and that take the effective number of categories into account, i.e. the number of equally abundant species that would be needed to give the same value of a diversity measure (Chao et al., 2014b). Moreover, several widely used diversity indexes, like variety/richness, Shannon entropy, Gini-Simpson Index, Rényi's or HDCT entropy, can be obtained from Hill numbers (Chao et al., 2014a).

When $q = 0$, variety, HCDT entropy, and Hill numbers are the same. Rényi entropy is equal to log(variety). When $q = 1$, Rényi entropy and HCDT entropy are equal to Shannon entropy (H), while Hill numbers are equal to the exponential of H. When $q = 2$, HCDT entropy is equal to Gini-Simpson, while the Hill numbers index is equal to the reciprocal of the index of concentration of Gini (or Herfindahl-Hirschman Index).

Some measures of diversity can capture variety, balance, and disparity at the same time. These measures are Rao and Rao-Stirling, where the former is widespread in ecology, while the latter is more commonly applied in social sciences and scientometrics (Rafols, 2014;Wang et al., 2015).

Concluding, there are several ways at measuring the composition of a population. They are closely related and share the same goal: to synthesize in an index the distribution of the categories of a given community (habitat, society, economy, or whatever), analyzing their concentration or dispersion between and within groups. The Shannon entropy, or the Frenken's variety, is just one of this large framework; under this point of view, RV and UV decomposition must be part of the framework and to be coherent with it.

3. Entropy Decomposition

As said, the evolutionary ecologists, biologists, and geneticists largely use the entropy indices, and its monotone transformation in the diversity index, to analyze how a population (meta-community) is distributed in sub-samples (communities). Often, this partition is further complicated by the presence of taxonomy or classification; in our case the SIC tree.

Just to have a practical case study, in the following we are going to apply the measures to the Italian workers' population distributed in the Italian Provinces (NUTS3); into each province, workers are distributed according to the SIC tree classification elaborated for Italy by Istat and called Ateco, in the last available year, 2018.

A meta-community (Italy) is partitioned into several local communities (provinces - indexed by $i = 1,2,\dots,I$). n_i individuals (workers) are sampled in community i. Let $s = 1,2,\dots,S$ denote the species (five-digit Ateco industries) that compose the meta-community, $n_{s,i}$ the number of individuals (workers) of species s (five-digit industry) sampled in the local community (province) $i, n_s = \sum_i n_{s,i}$ the total number of individuals of species s (i.e. the total number in Italy of workers in a given five-digit industry), $n = \sum_s \sum_i n_{s,i}$ the total number of sampled individuals (total workers in Italy). Within eachlocal community i, the probability $p_{s,i}$ for an individual to belong to species s is estimated by $\hat{p}_{s,i} = n_{s,i}/n_i$. The same probability for the meta-community is $p_s = n_s/n$. Communities have a weight w_i, satisfying $p_s = \sum_i w_i p_{s,i}$. The commonly used $w_i = n_i/n$ is a possible weight; in this case $\hat{p}_{s,i} = n_{s,i}/n$.

As said, entropy is a family of functions. This family is usually called HCDT entropy, in honor of the pioneering contributions by Havrda & Charvát 1967, Daróczy 1970 and Tsallis 1988. HCDT entropy is also known as Tsallis entropy. It is a parametrized family of entropy:

$$H(q) = -\sum_{s} p_s^q \, ln_q\left(p_s^q\right)$$

where $H(q)$ is the entropy of order q. The HCDT family uses the deformed logaritm; the logarithm of order q is defined as:

$$ln_q \, x = \frac{x^{1-q} - 1}{1-q}$$

and consequently, the Tsallis entropy is calculated as:

$$H(q) = \frac{1 - \sum_s p_s^q}{q-1} = -\sum_{s} p_s^q \, ln_q\left(p_s^q\right)$$

As shown in the previous section, by varying q we have alternative entropy measures: $q = 0$ produces Hill numbers. When $q = 1$, the Shannon entropy is obtained while $q = 2$, measures the Simpson one. $q$is hence the degree of entropy or diversity. By varying q we put more or less weight on rare species. $q = 0$ measures the "richness" of a meta-community.

Up to now, we have used entropy and diversity as strictly synonymous. Although they are, their meaning is not the same. Entropy is intimately related to the idea of uncertainty or disorder embodied in a metacommunity, diversity to the idea of how species are represented into a metacommunity. The two concepts are obviously related; Marcon and Hérault (2015a) generalized the duality of entropy and diversity: entropy can be converted promptly into diversity (Hill 1973; Jost 2006), which is easy to interpret and compare. Diversity is the deformed exponential of entropy:

$$D(q) = e_q^{qH(q)}$$

As the last word, diversity provides a "number" which is interpreted as the number of equal probability species needed to provide the diversity measured in the metacommunity; or, in other words, the number of equally frequent species that would give the same level of diversity of the data. For such a reason this type of approach to diversity is called the "species neutral approach".

In our data, the diversity must be interpreted as the number of uniformly distributed five-digit Ateco industries needed to match the measured diversity of our metacommunity. If the diversity is low, this means that we need few equally frequent industries to represent our data (concentration) and vice versa. The equally frequent species are also called the "effective species".

In the species-neutral approach, the diversity of the metacommunity is called γ and the one in a single community as α.

In our data, the Hill numbers ($q = 0$) is 817; having in sample 819 five-digit industries, we can conclude that "species" (five-digit sectors) are well represented in the metacommunity; in other words, Italy is rich of five-digit industries. Shannon entropy ($q = 1$) is 5.63 and Simpson one ($q = 2$) is 0.99, both confirming that the meta-community is rather diversified.

As said, it is interesting to calculate q-entropy, or diversity, at the metacommunity (γentropy) or species level, (α entropy). The gamma q-entropy is a weighted average of the q-alpha ones. The difference between γ and α provides a "residual" term, called βq-entropy. It is the generalized Jensen-Shannon divergence between the species distribution of the meta-community and those of communities (Marcon et al. 2014). The Jensen-Shannon is a method of measuring the statistical similarity between two sample distributions; it is strictly related to the Kullback-Leibler (KL) divergence or relative entropy, which measures how distant two distributions are from each other. For discrete probability distributions P(x) and Q(x), KL is given by:

$$KL(P||Q) = \sum_{x} P(x) \, ln \frac{P(x)}{Q(x)}$$

where || stands for "divergence".

Marcon et al. (2014) derived the decomposition of HCDT entropy, generalizing Shannon entropy partitioning (Rao & Nayak 1985; Marcon et al. 2012), based on Patil and Taillie's concept of diversity of a mixture (Patil & Taillie 1982)

$$H_\gamma(q) = H_\alpha(q) + H_\beta(q) = \sum_i w_i H_\alpha^i(q) + \sum_i w_i H_\beta^i(q)$$
$$H_\gamma(q) = \sum_s p_s^q \, ln_q(1/p_s)$$
$$H_\alpha^i(q) = \sum_s p_{s,i}^q \, ln_q(1/p_{s,i})$$
$$H_\beta^i(q) = \sum_s p_{s,i}^q \, ln_q(p_{s,i}/p_s)$$
$$w_i = \frac{n_i}{n}$$

s is the species (five-digit) in a single (local) community i. From the formula, it is clear that beta entropy $H_\beta^i(q)$is a measure of divergence between the local community and the meta-community in each species s.

For $q = 1$, and a single community ($i = 1$), this partition is quite like the ones that Frenken et al. (2007) call "related and unrelated variety", as we will see. In particular, the alpha-entropy corresponds to the variety. The unrelated is the entropy on the second level of the Ateco tree and the related comes from the difference (beta-entropy).

In our data the average alpha-entropy ($q = 1$) is 5.25 and the beta 0.38; hence the total Shannon entropy for the meta-community is 5.63. The decomposition confirms that our meta-community is well balanced, as the total entropy is almost totally due to the alpha-communities.

From entropy is possible to calculate diversity; in this case, gamma-diversity is given by the product of the alpha and beta diversity. Gamma diversity is 279 (effective species), alpha 191, and beta 1.46.

4. Related and Unrelated Variety

As said in the introduction, in the EEG, the concept of variety, and its decomposition between related and unrelated variety, is largely known, since the contribution by Frenken et al. (2007). It applies to a single community, not to the meta-community. They apply the Shannon entropy index (called "variety"[4]) to the three or five-digit level of the SIC tree; then variety is decomposed into UV (the entropy at two-digit level) and RV, obtained by averaging over entropies at five-digit industries belonging to the same two-digit sector. We recall that the Shannon entropy is a special case of HRST entropy for $q = 1$. Other contributions on the topic in Economics are Hidalgo et al., 2007, Rafols et al., 2010, Chavarro et al., 2014, Guevara et al., 2016, Eagle et al., 2010, and Farchy and Ranaivoson, 2011.

The total entropy (variety) at the lowest tree's slice (five-digit as an example) is hence decomposed by the following formula:

$$H(p_i) = Var = \sum_{s=1}^{n} p_s \, ln(1/p_s) = \sum_{s=1}^{n} p_{si} \frac{p_g}{p_g} ln\left(\frac{1}{p_s}\frac{p_g}{p_g}\right) = RV + UV$$
$$RV = \sum_{g=1}^{G} p_g \left(\sum_{s \in S_g} \frac{p_s}{p_g} ln\left(\frac{p_g}{p_s}\right)\right) = -\sum_{g=1}^{G} \left[\sum_{s \in S_g} p_s \, ln\left(\frac{p_s}{p_g}\right)\right]$$
$$UV = \sum_{g=1}^{G} p_g \, ln(1/p_g)$$

For a given local community or entity (the province in our case), n is the number of workers; the index compares two-digit ($p_g = n_g/n$) and five-digit ($p_s = n_s/n$) employment shares ; $s \in S_g$ are the species (five-digit industries) belonging to the same g sector (two-digit) where $g = 1, \ldots, G$. Hence $p_g = \sum_{i \in S_g} p_i$; it is the share of employment of a two-digit level w.r.t. total workers in the entity.

[4]Although this terminology is misleading, as variety means the number of groups of each entity.

The UV component is simply the entropy calculated at the first level of the tree (two-digit, or slice 1) in a given territory. The UV measures the degree to which employment shares are evenly distributed across unrelated (in this case two-digit) sectors. The values of UV can vary from 0 (when all employment is concentrated in only one two-digit sector) up to $ln(G)$ when all sectors employ an equal number of employees.

In the following, we refer to two-digit classification as "sector" and five-digit as "sub-groups" or "industries" belonging to the same sector. A given sector (two-digit) is hence composed of industries (five-digit) close to each other for type of production, or "related" according to the Frenken's definition. The RV can be obtained as the difference between Var and UV. Following what we saw in the previous sections, RV is a measure of divergence between two and five-digit distributions; mathematically, it is the Jensen-Shannon divergence. More precisely, from equation (RV), it is the sum of the G KL divergences (the term inside the square brackets); it is always a positive term, as $p_i < p_g = \sum_{i \in S_g} p_i$. RV measures the relative entropy between the two distributions; the lower RV, the more similar the two distributions are. In words, the distribution of workers of each sector G is mirrored within the sector itself and vice versa.

In the EEG, RV and UV play two different roles. When the productive tissue is widespread in a high number of different sectors (UV is high), the territory is less vulnerable to sector-specific shocks; the diversification operates as in the standard portfolio approach. Nonetheless, this implies that sectors are far from each other (unrelatedness) which in some sense contrast the knowledge spillovers which arise in more homogenous sectors. Conversely, when the economy is concentrated in few sectors, both at two and five-digit, it is highly specialized; firms are very close for the type of production, technical progress, human capital, markets and are highly localized. In this case knowledge spills like in the Mars agglomeration theory. When the economy is concentrated but at a more moderate level, in the sense that industries are close but not strictly related, Jacob externalities could emerge. This is a sort of intermediate case between a highly specialized economy (few sectors both in five and two-digits) and a diffuse economy (high number of heterogeneous industries both at two and five-digit). In this particular case, industries share again knowledge and skills but to a less extent w.r.t. the Mars case. Their "cognitive proximity" (Boschma 2005) allows firms to share knowledge, ideas, and practices.

It is clear, hence, that RV and UV can tell different stories about the debate around the relationship between sectoral composition and knowledge transmission. According to Frenken et al. (2007), "Unrelated variety protects a region best against external asymmetric shocks in demand and thus against rising unemployment. By contrast, related variety in a sector is expected to be beneficial for Jacobs externalities in the form of knowledge spillovers, thus enhancing growth and employment (page 688)". And again "the present authors (Frenken et al., 2007, author's note) **c**onsider related variety to be the indicator for Jacobs externalities because it measures the variety within each of two-digit classes. It is expected that the economies arising from variety are especially strong between subsectors, as knowledge spills over primarily between firms selling related products. By contrast, unrelated variety measures the extent to which a region is diversified in very different types of activity. This type of variety is expected to be instrumental in avoiding unemployment"(page 687). According to Aarstad et al. (2016), "In a region with a high level of related variety, enterprises operate in different industries that share several similarities, whereas in a region with a high level of unrelated variety, enterprises operate in different industries that share few or limited similarities ... Following this line of reasoning, regions with unrelated industrial variety will conversely experience less resource sharing because the cognitive distance between the enterprises is too great. ... we emphasize that industrial specialization is a two-dimensional construct in which a low level of specialization can indicate a region with a high level of related or unrelated industrial variety"(Aarstad et al., 2016, page 845). Aarstad et al. (2016) summarize their analysis in the following matrix (page 854):

		Related variety	
		Low	*High*
Unrelated variety	*High*	*1) Regions with a low level of related variety but a high level of unrelated variety:* • A low level of related variety **constrains innovation** • A high level of unrelated variety **constrains productivity**	*2) Regions with high levels of both related and unrelated variety:* • A high level of related variety **fosters innovation** • A high level of unrelated variety **constrains productivity**
	Low	*3) Specialized regions (with low levels of both related and unrelated variety):* • A low level of related variety **constrains innovation** • A low level of unrelated variety **fosters productivity**	*4) Regions with a high level of related variety but a low level of unrelated variety:* • A high level of related variety **fosters innovation** • A low level of unrelated variety **fosters productivity**

Figure 2:Aarstad et al. (2016).

These long citations stress the use that RV and UV play in this literature. Authors treat these two measures in an independent way to each other; each of them is responsible for its own interpretation. We have some doubts about this use of the entropy measures. Moreover, the literature treats each geographic entity on its own, completely detached from the economic context. But knowledge spills both over time and space. The literature on spatial econometrics underlines the importance of spillovers among spatially related geographical entities; a given region cannot be analyzed "in vitro". The sectoral composition of a geographic entity could affect, or be affected, by what happens in their neighborhood and in the whole economy (meta-community). We will back on the point in section 6.

Just as a case stud**y,** Figure 3 **s**hows the provinces' map according to whether RV and/or UV are upper (H) or lower (L) than the national average.

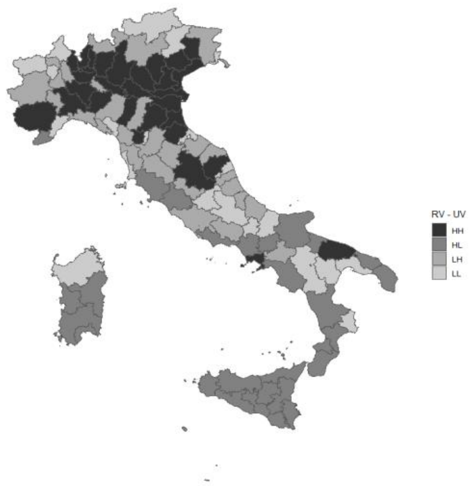

Figure 3: UV and RV map.

Table 1 in the Appendix shows the indexes for each province. Figure 3 shows that Italian provinces are rather clustered; a large part of northern provinces show a high level both for RV and UV. In the southern ones, there is a prevalence of high RV and low UV. According to the EEG literature, we should conclude that northern provinces are characterized by a widespread economic tissue, with a large number of heterogenous industries (UV is high). At the same time, the high value of RV means that industries are related and that Jacob externalities are at work. While the latter point should sustain growth and productivity, the former operates in the opposite direction, as the wide presence of heterogenous industries form a sort of barrier to knowledge diffusion, as cognitive proximity is very low. So the forces operate in an opposite way; which is winning? Southern provinces are more characterized by high RV and low UV.

This seems the best situation, few heterogeneous sectors (low UV) jointly to a large number of related industries sustain innovation and productivity. Hence southern provinces should perform better than the northern ones: a conclusion that is not true for Italy. What is wrong?

5. Bridging the Two Approaches

The EEG approach borrows from life and natural sciences; nevertheless, several remarks need. First of all, each territory is a singleton. There is not a comparison between the local and the "national" level. This is not true in the meta-community approach, where total entropy, or diversity, is decomposed in local (alpha) entropy and the divergence between local and meta-community (beta). Moreover, the economic interpretation of related and unrelated decomposition is not clear. Variety corresponds to the alpha-entropy of a single community; the UV is simply the entropy calculated at the higher slice of the tree and measures the concentration/dispersion of sectors in a single community, without ambiguity. What the EEG calls related variety, so important for Jacob's externalities, actually measures the Jensen-Shannon divergence between the distribution at the lowest slice (five-digit) and the one at the upper level (the two-digit). It is a relative entropy; when it is low, this means that the two distributions are similar and vice versa. In the EEG seems that the interpretation is different; RV is used to assess whether the five-digit level distribution is concentrated or not; when RV is low economies are more specialized and vice versa. In other words, RV and UV are used as two independent measures of entropy. Actually, this is not true: RV is a measure of relative entropy and for such a reason must be compared with the variety and the UV one, being their residual component; a high (low) value cannot be interpreted as a sign of dispersion (concentration) of the five-digit structure respect the two-digit one if . The correct approach should be to decompose the total entropy in the alpha and beta components. In case, the alpha entropy could be further decomposed in the UV and RV components.

Moreover, in the EEG approach, there is not a measure of interaction between the local community and the national one (the beta entropy). The latter delivers an important information, as it underlines sectoral specificity in the local communities compared to the national productive tissue. Moreover, spatial econometrics, in the last twenty years, has stressed the importance of spatial externalities among geographical entities. The diffusion of knowledge runs both in time and space. Under this point of view, focusing on a single geographical entity, completely detached by its context, could misestimate the true channels of knowledge diffusion.
By trying to unify the two approaches, the gamma entropy of the meta-community can be calculated in the following way $(q = 1)$:

$$H_\gamma = H_\alpha + H_\beta = RV + UV + H_\beta$$

Accordingly, using RV and UV does not provide a final picture of the total entropy of a local community. The beta-entropy delivers a piece of important information: how the local level diverges from the national one; the higher it is, the more dispersed the local community is w.r.t. to the national one.

Figure 4 shows a similar map of Figure 3**;** it plots alpha and beta higher or lower than the national level:

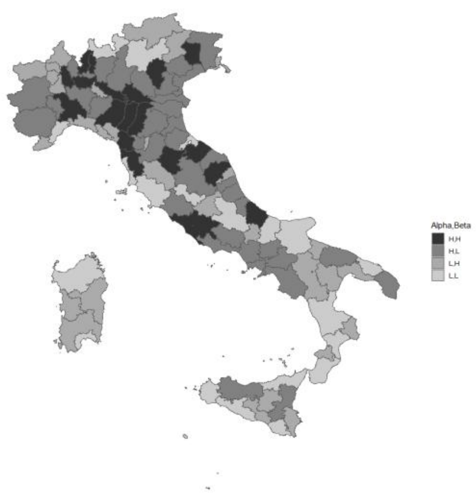

Figure 4:Alpha and Beta map.

The two maps deliver two different points of view: in Figure 4 the alpha-entropy corresponds to the Var of the EEG approach. The comparison between the maps, as expected, provides different conclusions, so the question is: what is the most suitable? As an example, let us focus on the province of Rome.Figure 3 and Table 1 shows that Rome is characterized by low RV (1.69) and high UV (3.60). How do we interpret this figure? Undoubtedly, a high entropy at the two-digit level (UV) means that this province has a widespread productive tissue. RV interpretation is more ambiguous; the low number induces to conclude that, inside each two-digit sector, industries are more concentrated. Actually, if we look correctly at RV, we should conclude that the divergence between the five-digit distribution (variety) and the one at level two, UV, is low; hence, even at the five-digit level, the number of industries is widespread, by mimicking what happens at the two-digit level. In the province of Rome, the highest two-digit employment share (10%) is in the A47 sector (Retail Wholesale); it is made of 90 five-digit industries (the highest populated two-digit sector). The A46 sector (Wholesale Trade) employs 5% of the total workers of the Rome province and it is composed of the highest number of five-digit sectors (80). The low value of RV should bring to conclude that, at a five-digit level, the Rome province is concentrated? We believe that this conclusion would be wrong. The right conclusion should be that there is a low divergence between the distribution of employers at the five-digit level (variety) and the one at two (UV). If we look at Figure 4 the same province is characterized both by high alpha and beta entropy: i.e. large dispersion of total entropy at five-digits (VAR) and large local dispersion w.r.t. the national average. This seems more coherent with the economy of this province; a large economy, with respectto the average, with a vital productive tissue.

As a counterexample, let us focus on a province that we know to be highly specialized, like the one of Prato. This is a small Tuscan province which is a well-known textile district; with its 7149 firms in the textile sector, Prato is the largest textile district in Europe. We should expect that this specialization produces a very low value of UV, as the production is practically concentrated in two sectors only (A14 and A13), and data confirms, as UV is low. Figure 3, and Table 1, show that Prato is characterized by alow value both for RV (the lowest value in Italy) and UV. The low value of UV is coherent with the data, as Prato is highly specialized, but not RV. In fact, if we look at the employment share of five-digit groups belonging to A14 and A13, they cover all the sub-groups of Ateco classification. In other words, the five-digit structure of these two-digit sectors is dense and largely widespread; if we would use the EEG interpretation, we should expect a high RV instead of a low one.More coherently, in Figure 4, Prato has low alpha and high beta. This confirms that is a specialized province (low variety) but, in addition, that its productive structure is well different from the national average, confirming its leading role in this type of production.

The last intriguing case we wish to analyze is high RV and UV. About UV, this undoubtedly means that two-digit sectors are widespread. What about RV? According to our interpretation, this should warn that the 5-digit structure embodied in the two-digit one is rather different from the one at two-digit. Let us focus on an important player for Italy, like the province of Milan. In Figure 5, we show the polynomial interpolation of the distribution of the employment share both two-digit (top) and five-digit (low); in the latter, the x-axis shows the corresponding Ateco 2 sectors grouping the Ateco 5.

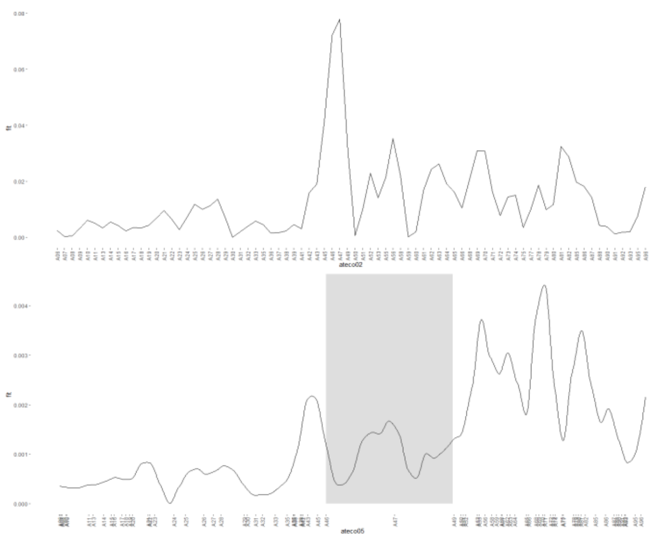

Figure 5:Distribution estimates in the province of Milan.

As Figure 5 shows, Ateco 2 has a peak in A46 and A47 but the corresponding five-digit share (grey area) does not show a particular concentration in these sectors; moreover, the five-digit distribution seems right-skewed, differently from the two-digit one. In other words, the structures are rather different; this explains the RV high value.

6. Conclusions

In this paper, we have reviewed the main characteristics of the entropy measures, both from an analytical and interpretative point of view. In particular, we have focused on the EEG literature and the way it exploits these measures for assessing the spreading of knowledge inside the industrial sectoral composition of a geographical entity. The entropy plays a leading role in several scientific disciplines, as biology, ecology, genetics, where it is used to analyze the evolution of the species in habitat and its classification into a tree structure (phylogenetic tree). The EEG mimics such an approach for investigating the industrial sectoral composition; in particular, the decomposition of the entropy (variety) at the lowest level of the SIC tree (usually the five-digit one) in related and unrelated variety provides a possible interpretation of how knowledge spills in the economy and its effects on growth and productivity. While it remains an interesting and useful approach, some shadows remain on the way the related variety is interpreted and used. By exploiting methodologies commonly used in natural sciences, we argue that total entropy should be decomposed into alpha and beta components, that are easier to interpret and more coherent with the mathematical foundations. Moreover, the beta entropy provides a measure of divergence of the local geographical entity w.r.t. the entire economy, or ecosystem. This is particularly useful in the context of a spatial transmission of knowledge. Local economies are not islands: they share knowledge, skills, markets with the rest of the economy. In the last twenty years, spatial econometrics has enlarged our horizon, allowing us to account for spatial spillovers among local economies, both in a direct (how a territory affects the neighbors) and indirect (how a territory is affected) way. Regional Economics and Economic Geography have gained another fundamental empirical tool for investigating the functioning of local economies. The decomposition of total entropy in gamma and beta components could be beneficial at integrating such an approach.

References

1) Aarstad, J., Kvitastein, O. A., & Jakobsen, S. E. (2016). Related and unrelated variety as regional drivers of enterprise productivity and innovation: A multilevel study. *Research Policy, 45*(4), 844-856.
2) Berger, W. H., & Parker, F. L. (1970). Diversity of planktonic foraminifera in deep-sea sediments. *Science, 168*(3937), 1345-1347.
3) Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure* (Vol. 7). New York: Free Press.
4) Boschma, R. (2005). Proximity and innovation: a critical assessment. *Regional studies, 39*(1), 61-74.
5) Boschma, R., & Iammarino, S. (2009). Related variety, trade linkages, and regional growth in Italy. *Economic Geography, 85*(3), 289-311.
6) Ceriani, L., & Verme, P. (2012). The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *The Journal of Economic Inequality, 10*(3), 421-443.
7) Chao, A., Chiu, C. H., & Jost, L. (2014a). Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annual review of ecology, evolution, and systematics, 45*, 297-324.
8) Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014b). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological monographs, 84*(1), 45-67.
9) Chavarro, D., Tang, P., & Rafols, I. (2014). Interdisciplinarity and research on local issues: evidence from a developing country. *Research Evaluation, 23*(3), 195-209.
10) Cowell, F. A. (1985). Measures of distributional change: An axiomatic approach. *The Review of Economic Studies, 52*(1), 135-151.
11) Cowell, F. A., & Kuga, K. (1981). Additivity and the entropy concept: an axiomatic approach to inequality measurement. *Journal of Economic Theory, 25*(1), 131-143.
12) Daly, A. J., Baetens, J. M., & De Baets, B. (2018). Ecological diversity: measuring the unmeasurable. *Mathematics, 6*(7), 119.
13) Daróczy, Z. (1970). Generalized information functions. *Information and control, 16*(1), 36-51.

14) de la Vega, C. L., Urrutia, A., & Volij, O. (2013). An axiomatic characterization of the Theil inequality ordering. *Economic Theory, 54*(3), 757-776.
15) Eagle, N., Macy, M., & Claxton, R. (2010). Network diversity and economic development. *Science, 328*(5981), 1029-1031.
16) Farchy, J., & Ranaivoson, H. (2011). Measuring the Diversity of Cultural Expressions: Applying the Stirling Model of Diversity in Culture. Technical Report 6, Unesco Institute for Statistics.
17) Frenken, K., Van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional studies, 41*(5), 685-697.
18) Guevara, M. R., Hartmann, D., & Mendoza, M. (2016). diverse: an R Package to Analyze Diversity in Complex Systems. *R J., 8*(2), 60.
19) Havrda, J., & Charvát, F. (1967). Quantification method of classification processes. Concept of structural α -entropy. *Kybernetika, 3*(1), 30-35.
20) Hidalgo, C. (2015). *Why information grows: The evolution of order, from atoms to economies.*Basic Books.
21) Hidalgo, C. A., Klinger, B., Barabási, A. L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science, 317*(5837), 482-487.
22) Hill, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology, 54*(2), 427-432.
23) Jost, L. (2006). Entropy and diversity. *Oikos, 113*(2), 363-375.
24) Jost, L. (2010). The relation between evenness and diversity. *Diversity, 2*(2), 207-232.
25) Marcon, E., & Hérault, B. (2015). Decomposing phylodiversity. *Methods in Ecology and Evolution, 6*(3), 333-339.
26) Marcon, E., Hérault, B., Baraloto, C., & Lang, G. (2012). The decomposition of Shannon's entropy and a confidence interval for beta diversity. *Oikos, 121*(4), 516-522.
27) Marcon, E., Scotti, I., Hérault, B., Rossi, V., & Lang, G. (2014). Generalization of the partitioning of Shannon diversity. *PloS ONE, 9*(3), e90289.
28) Patil, G. P., & Taillie, C. (1982). Diversity as a concept and its measurement. *Journal of the American statistical Association, 77*(379), 548-561.
29) Pavoine, S., Love, M. S., & Bonsall, M. B. (2009). Hierarchical partitioning of evolutionary and ecological patterns in the organization of phylogenetically-structured species assemblages: application to rockfish (genus: Sebastes) in the Southern California Bight. *Ecology letters, 12*(9), 898-908.
30) Pielou, E. C. (1970). *Introduction to Mathematical Ecology*. JohnWiley & Sons Inc, New York.
31) Rafols, I. (2014). Knowledge integration and diffusion: Measures and mapping of diversity and coherence. In *Measuring scholarly impact* (pp. 169-190). Springer, Cham.
32) Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for information Science and Technology, 61*(9), 1871-1887.
33) Rao, C. R. (1982). Diversity and dissimilarity coefficients: a unified approach. *Theoretical population biology, 21*(1), 24-43.
34) Rao, C., & Nayak, T. (1985). Cross entropy, dissimilarity measures, and characterizations of quadratic entropy. *IEEE Transactions on Information Theory, 31*(5), 589-593.
35) Rényi, A. (1961). *On measures of entropy and information. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Contributions to the Theory of Statistics*. The Regents of the University of California.
36) Rhoades, S. A. (1993). The herfindahl-hirschman index. *Fed. Res. Bull.*, 79, 188.
37) Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal, 27*(3), 379-423.
38) Simpson, E. H. (1949). Measurement of diversity. *Nature, 163*(4148), 688-688.
39) Stirling, A. (1998). On the economics and analysis of diversity. *Science Policy Research Unit (SPRU), Electronic Working Papers Series, Paper*, 28, 1-156.
40) Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface, 4*(15), 707-719.
41) Theil, H. (1967). *Economics and Information Theory*. North-Holland, Amsterdam.
42) Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of statistical physics, 52*(1), 479-487.
43) Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. *PloS ONE, 10*(5), e0127298.

Appendix

Table 1: Alpha and Beta entropies

Provinces	Alpha	Beta	Gamma	RV	UV	VAR
Agrigento	5.18	0.35	5.54	1.95	3.23	5.18
Alessandria	5.36	0.39	5.76	1.79	3.58	5.36
Ancona	5.38	0.29	5.67	1.67	3.71	5.38
Arezzo	5.26	0.48	5.74	1.64	3.61	5.26
Ascoli Piceno	5.32	0.23	5.55	1.68	3.64	5.32
Asti	5.28	0.38	5.66	1.72	3.56	5.28
Avellino	5.29	0.33	5.62	1.72	3.57	5.29
Bari	5.41	0.18	5.59	1.87	3.54	5.41
Barletta	5.19	0.40	5.60	1.85	3.34	5.19
Belluno	4.74	1.00	5.74	1.34	3.40	4.74
Benevento	5.28	0.26	5.54	1.78	3.50	5.28
Bergamo	5.54	0.29	5.83	1.87	3.67	5.54
Biella	4.92	0.82	5.74	1.58	3.34	4.92
Bologna	5.50	0.30	5.80	1.80	3.70	5.50
Bolzano	5.09	0.48	5.57	1.68	3.41	5.09
Brescia	5.50	0.26	5.76	1.89	3.61	5.50
Brindisi	5.25	0.30	5.54	1.84	3.41	5.25
Cagliari	5.25	0.29	5.55	1.76	3.49	5.25
Caltanissetta	5.10	0.39	5.49	1.75	3.35	5.10
Campobasso	5.14	0.34	5.48	1.67	3.47	5.14
Caserta	5.34	0.25	5.58	1.86	3.47	5.34
Catania	5.34	0.27	5.61	1.89	3.45	5.34
Catanzaro	5.20	0.29	5.49	1.83	3.37	5.20
Chieti	5.27	0.41	5.68	1.72	3.55	5.27
Como	5.41	0.39	5.80	1.77	3.64	5.41
Cosenza	5.23	0.25	5.48	1.90	3.32	5.23
Cremona	5.43	0.41	5.84	1.85	3.58	5.43
Crotone	5.00	0.53	5.53	1.73	3.26	5.00
Cuneo	5.43	0.32	5.75	1.83	3.60	5.43
Enna	5.05	0.39	5.45	1.81	3.24	5.05
Fermo	4.66	1.03	5.68	1.51	3.14	4.66
Ferrara	5.30	0.30	5.59	1.76	3.54	5.30
Firenze	5.40	0.35	5.76	1.73	3.67	5.40
Foggia	5.23	0.28	5.51	1.86	3.37	5.23
Forlì-Cesena	5.41	0.29	5.71	1.79	3.62	5.41
Frosinone	5.26	0.31	5.57	1.66	3.60	5.26
Genova	5.13	0.60	5.73	1.60	3.53	5.13
Gorizia	5.13	0.55	5.68	1.63	3.50	5.13
Grosseto	5.09	0.31	5.40	1.75	3.34	5.09
Imperia	4.95	0.38	5.34	1.78	3.17	4.95
Isernia	5.04	0.43	5.47	1.56	3.49	5.04
L'Aquila	5.09	0.35	5.44	1.70	3.48	5.17
La Spezia	5.17	0.42	5.59	1.55	3.54	5.09
Latina	5.28	0.25	5.54	1.79	3.49	5.28
Lecce	5.34	0.22	5.56	1.92	3.42	5.34
Lecco	5.42	0.51	5.92	1.90	3.51	5.42
Livorno	5.19	0.36	5.55	1.72	3.47	5.19
Lodi	5.22	0.44	5.66	1.65	3.57	5.22
Lucca	5.31	0.38	5.69	1.70	3.61	5.31
Macerata	5.33	0.40	5.74	1.76	3.57	5.33
Mantova	5.40	0.41	5.81	1.76	3.64	5.40
Massa-Carrara	5.20	0.41	5.61	1.73	3.47	5.20
Matera	5.12	0.47	5.59	1.62	3.50	5.12

Messina	5.22	0.27	5.49	1.85	3.37	5.22
Milano	5.50	0.39	5.89	1.78	3.72	5.50
Modena	5.36	0.45	5.81	1.72	3.64	5.36
Monza Brianza	5.58	0.35	5.93	1.92	3.65	5.58
Napoli	5.44	0.22	5.66	1.89	3.55	5.44
Novara	5.35	0.42	5.77	1.75	3.60	5.35
Nuoro	5.06	0.39	5.44	1.76	3.30	5.06
Oristano	5.05	0.39	5.43	1.82	3.23	5.05
Padova	5.59	0.20	5.79	1.91	3.68	5.59
Palermo	5.30	0.32	5.62	1.83	3.46	5.30
Parma	5.33	0.39	5.73	1.74	3.60	5.33
Pavia	5.46	0.26	5.71	1.84	3.62	5.46
Perugia	5.46	0.20	5.66	1.83	3.63	5.46
Pesaro Urbino	5.31	0.40	5.71	1.72	3.58	5.31
Pescara	5.29	0.26	5.54	1.70	3.58	5.29
Piacenza	5.32	0.37	5.69	1.79	3.52	5.32
Pisa	5.32	0.45	5.77	1.68	3.64	5.32
Pistoia	5.35	0.30	5.65	1.78	3.58	5.35
Pordenone	5.34	0.49	5.83	1.79	3.55	5.34
Potenza	5.04	0.45	5.49	1.56	3.48	5.04
Prato	4.66	1.09	5.75	1.53	3.13	4.66
Ragusa	5.25	0.30	5.55	1.90	3.34	5.25
Ravenna	5.39	0.28	5.67	1.75	3.64	5.39
Reggio Calabria	5.22	0.31	5.53	1.97	3.25	5.22
Reggio Emilia	5.39	0.44	5.83	1.82	3.57	5.39
Rieti	4.99	0.38	5.38	1.55	3.44	4.99
Rimini	5.14	0.34	5.48	1.67	3.46	5.14
Roma	5.30	0.38	5.68	1.69	3.60	5.30
Rovigo	5.37	0.31	5.68	1.76	3.60	5.37
Salerno	5.32	0.21	5.54	1.86	3.46	5.32
Sassari	5.12	0.29	5.41	1.71	3.40	5.12
Savona	5.12	0.34	5.46	1.68	3.44	5.12
Siena	5.16	0.41	5.57	1.57	3.60	5.16
Siracusa	5.20	0.38	5.59	1.77	3.43	5.20
Sondrio	5.16	0.36	5.52	1.61	3.54	5.16
Sud Sardegna	5.08	0.40	5.49	1.78	3.30	5.08
Taranto	5.05	0.50	5.55	1.64	3.41	5.05
Teramo	5.32	0.31	5.62	1.73	3.59	5.32
Terni	5.23	0.36	5.59	1.64	3.59	5.23
Torino	5.43	0.29	5.71	1.74	3.68	5.43
Trapani	5.23	0.30	5.53	1.89	3.35	5.23
Trento	5.24	0.29	5.53	1.62	3.62	5.24
Treviso	5.54	0.34	5.88	1.85	3.69	5.54
Trieste	5.17	0.55	5.72	1.57	3.61	5.17
Udine	5.38	0.32	5.70	1.75	3.63	5.38
Valle d'Aosta	4.94	0.52	5.46	1.46	3.48	4.94
Varese	5.51	0.36	5.87	1.81	3.69	5.51
Venezia	5.37	0.25	5.62	1.76	3.62	5.37
Verbano-Cusio	5.19	0.38	5.57	1.72	3.47	5.19
Vercelli	5.17	0.56	5.73	1.59	3.58	5.17
Verona	5.51	0.21	5.72	1.86	3.66	5.51
Vibo Valentia	5.07	0.42	5.49	1.88	3.19	5.07
Vicenza	5.54	0.40	5.93	1.90	3.64	5.54
Viterbo	5.26	0.35	5.61	1.89	3.38	5.26