

Studies on Regional Wealth Inequalities: the case of Italy

Marcel Ausloos^{1,2,3} and Roy Cerqueti⁴

¹School of Management, University of Leicester,
University Road, Leicester, LE1 7RH, UK
e-mail address: ma683@le.ac.uk

²eHumanities group*,
Royal Netherlands Academy of Arts and Sciences,
Joan Muyskenweg 25, 1096 CJ Amsterdam, The Netherlands
³GRAPES[†], rue de la Belle Jardiniere 483,
B-4031, Angleur, Belgium

e-mail address: marcel.ausloos@ulg.ac.be

⁴University of Macerata, Department of Economics and Law,
via Crescimbeni 20, I-62100, Macerata, Italy
e-mail address: roy.cerqueti@unimc.it

Jan. 24, 2016

Abstract

The paper contains a short review of techniques examining regional wealth inequalities based on recently published research work [1, 2, 3, 4, 5, 6], but also presenting unpublished features. The data pertains to Italy (IT), over the period 2007-2011: the number of cities in regions, the number of inhabitants in cities and in regions, as well as the aggregated tax income of the cities and of regions. Frequency-size plots and cumulative distribution function plots, scatter plots and rank-size plots are displayed. The rank-size rule of a few cases is discussed. Yearly data of the aggregated tax income is transformed into a few indicators: the Gini, Theil, and Herfindahl-Hirschman indices. Numerical results confirm that IT is divided into very different regional realities. One region is selected for a short discussion: Molise. A note on the "first digit Benford law" for testing data validity is presented.

1 Introduction

In studying geo-complexity, many variables come to mind. Beside geo-political aspects, many considerations are social, economic and financial ones. One is aware that local aspects should be distinguished from global ones. Various levels can be imagined for discussion and modeling. Here below, the paper contains

*Associate Researcher

[†]Group of Researchers for Applications of Physics in Economy and Sociology

a short review of techniques and results examining regional wealth inequalities based on recently published research work [1, 2, 3, 4, 5, 6], but also presenting unpublished features. In brief, our aim is to find correlations between long lived macro-level and shorter lived micro-level features.

The data pertains to Italy (IT), over the quinquennium period 2007-2011: (i) the number of cities in regions, (ii) the number of inhabitants in cities and (iii) in regions, based on the 2011 Census, as well as (iv) the yearly aggregated tax income of the cities and (v) of regions, from 2007 till 2011, - but the ideas can be carried forward to other countries or systems. Rank-size plots and cumulative distribution function plots, are displayed in Sect.2. Frequency size and scatter plots are to be found in the listed references.

In Sect. 3, the yearly data of the aggregated tax income (ATI), i.e. the city contribution to the national GDP from citizens income tax, In Sect. 4 financial inequalities are examined: the ATI data is transformed into a few usual indicators of wealth distributions: the Theil index [7], a Herfindahl-Hirschman index [8] and the Gini coefficient [9].

One region has been selected for illustrating a smaller entity than the entire country, i.e. the Molise region. In order not to distract from the aim, a consideration on data validity, with emphasis on Molise and IT is given at the end through Benford Law first digit [10, 11]; see Sect. 5.

2 Administrative distribution of cities in regions

As far as 2011, Italy is composed of 20 regions, 110 provinces and 8092 municipalities, respectively. Since 2007, the number of IT cities has been yearly evolving to become respectively: 8101, 8094, 8094, 8092, 8092, - from 2007 till 2011.

In brief, several cities have thus merged into new ones, other were phagocytized. To sum up: 13 cities \rightarrow 4, in two steps, over the studied quinquennium. For completeness, let us mention them: Campolongo Tapogliano (UD), Ledro (TN), Comano Terme (TN), and Gravedona ed Uniti (CO). Moreover, a few (7) cities changed both province *and* region membership: such 7 cities moved from PU (the province of Pesaro and Urbino) in the Marche region, to RN (province of Rimini) in the Emilia Romagna region (Castel delci, Maiolo, Novafeltria, Pennabilli, San Leo, Sant' Agata Feltria, Talamello). For completeness, notice that 228 municipalities have changed from a province to another one, but nevertheless remained in the same region.

For completeness, let it be mentioned that in 2007, the number of provinces was increased from 103 (to 110) by 7 units (BT, CI, FM, MB, OG, OT, VS)¹ thereafter, leading to 110 provinces; 4 provinces have been instituted by a regional law of 12 July 2001 in Sardinia and became operative in 2005 (CI, MB, OG, OT), while BT, FM and VS have been created on June 11th, 2004 and became operative on June 2009.

The number of regions has remained constantly equal to 20 since 1963. It should be here mentioned that Molise (MOL) is the youngest Italian region, being established in 1963, when the region "Abruzzi e Molise" was split. Campobasso (CB) and Isernia (IS).

¹see ISO code: http://en.wikipedia.org/wiki/Provinces_of_Italy. .

For further historical and data details, we address the reader to: [http : //www.comuni – italiani.it/regioni.html](http://www.comuni-italiani.it/regioni.html). The above indicates that some care had to be taken to compare entities at various times in the quinquennium, i.e. when necessary "data manipulation through summation" occurred.

Let us illustrate the global view of IT cities through a histogram presenting (counting) the number of cities having a population of N inhabitants, as in Fig.1, according to the 2011 Census. In order to emphasize the 6 main cities, the y -axis is truncated at a value = 10 for readability. It can be shown that these 6 cities are outliers in subsequent statistical analysis of all pertinent data. In so doing, the notion of King (K) and that of Vice-Roy (VR) cities can be introduced [2, 12, 13]. Such outliers always occur, at low rank when considering rank-size plots, and much influence classical fits with simple empirical laws (see also a discussion below). For example, the rank-size law of the number ($N_{c,r}$) of cities in a given region for the IT regions ($N=20$) in various recent years is shown in Fig. 2. N.B. the data for 2007 and for 2010-2011 are displaced for readability. The fit of $N_{c,r}$, for the (undisplaced) data, with a 3-parameter function

$$g_3(r) = A (N r)^{-\gamma} (N - r + 1)^\beta \quad (2.1)$$

underlines the existence of 3 "regimes": at low, middle and high rank, Fig. 2. Observe that this function, Eq. (2.1) generalizes the mere power law $N_r = N_1 r^{-\alpha}$, and takes into account the finite size of the system and the strict vanishing of the distribution at some maximum rank [13, 14, 15, 16, 17, 18] .

3 Aggregated Tax Income

Considering the rank-size law for economic data, similar situations occur. On one hand, it was observed that the distribution of ATI between regions involved with time, more precisely than demographic values [1, 2, 3, 4],

On the other hand, looking at the data, after calculating either the number ($N_{i,r}$) of inhabitants in a region or the regional ATI (ATI_r), i.e. the ATI sum for the relevant cities in a region, for each year and also for the subsequent quinquennium average, the change in regional membership appeared to be very weakly relevant. Thus, presenting unpublished data, consider the case of the Molise (MOL) region. Molise has $N = 136$ cities (and is made of 2 provinces). A rank-size law, is hardly represented by a power law (not shown), nor by a simple 2-parameter free function taking into account the finite size of the system, as in

$$g_2(r) = \kappa_2 \left[\frac{N r}{N - r + 1} \right]^{-\chi}, \quad (3.1)$$

even removing outliers, like the King (Campobasso) and the 3 Vice-Roys (Termoli, Isernia, Venafro) cities in MOL.

These 2-parameter free and 3-parameter free functions (Eqs.(3.1)-2.1) respectively, for rank-size law research are compared in Fig. 3 and Fig. 4. The various fit coefficients are given in Table 1.

It seems, to us and hopefully to the readers, that one can conclude about the necessity of first examining histograms of distributions, during the examination of ranked data, in order to pin point relevant outliers, on one hand, and on the other hand, in view of finely reproducing the empirical data through adequate but simple functions.

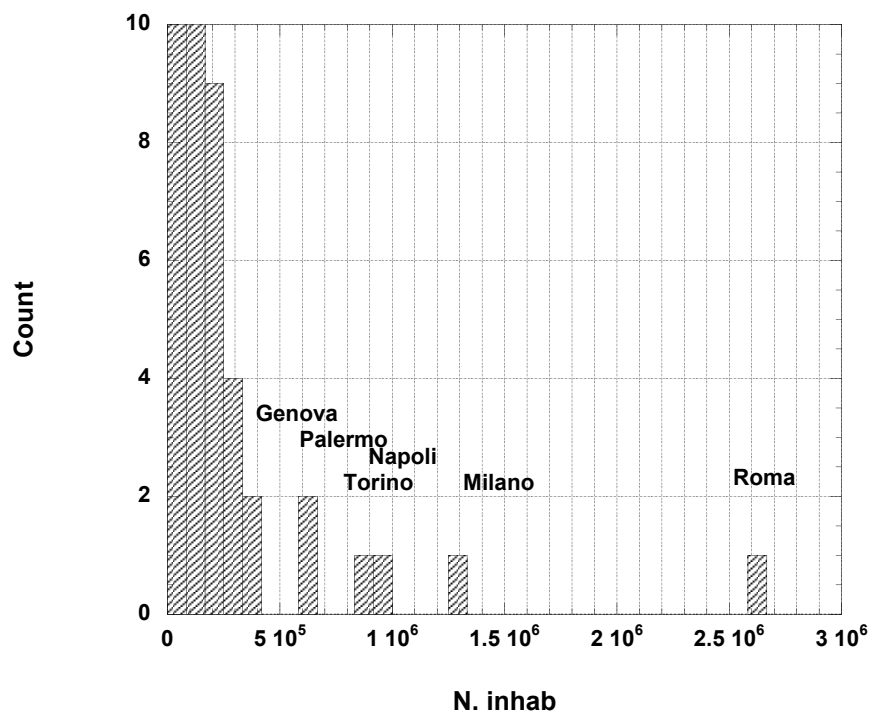


Figure 1: Histogram of the number of large cities in IT having a population of N inhabitants; the y -axis is truncated at a value = 10 for readability in order to emphasize the 6 main cities.

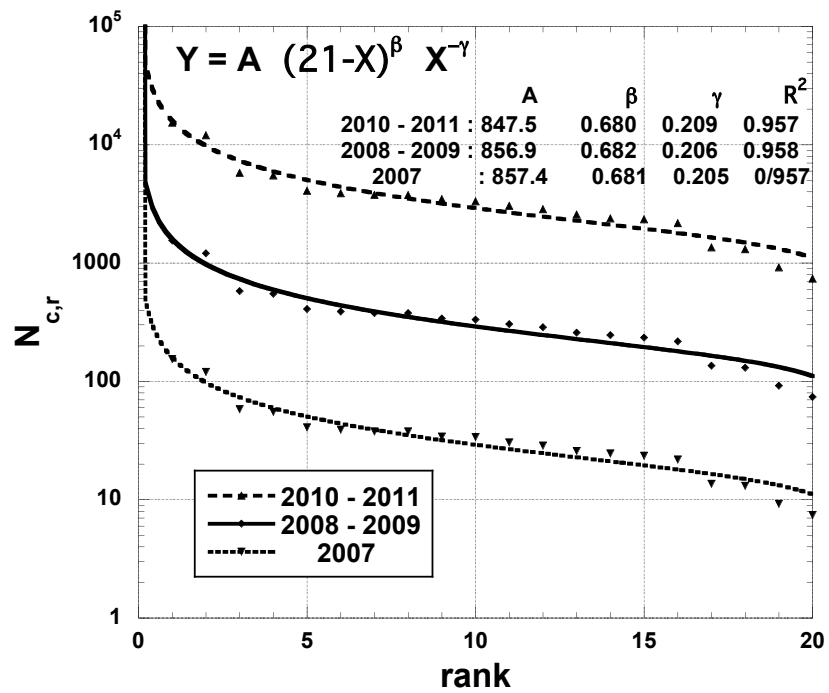


Figure 2: Display of the rank-size law of the number ($N_{c,r}$) of cities in a given region for the IT regions ($N=20$) in various recent years; data for 2007 and for 2010-2011 are displaced for readability; fit values for a 3-parameter function, as Eq. (2.1), are given for the undisplaced data

2 param			3 param			
	all	-KVR			all	-KVR
κ_2	3.525 ± 0.26	6.443 ± 0.113	2007	A	201.0 ± 251.6	47.090 ± 9.55
χ	1.049 ± 0.02	0.730 ± 0.007	2007	γ	1.079 ± 0.02	0.809 ± 0.01
			2007	β	0.226 ± 0.25	0.361 ± 0.04
R^2	0.979	0.990	2007	R^2	0.979	0.993
κ_2	3.672 ± 0.28	6.713 ± 0.122	2008	A	205.1 ± 258.4	45.467 ± 8.41
χ	1.046 ± 0.02	0.726 ± 0.007	2008	γ	1.076 ± 0.02	0.795 ± 0.01
			2008	β	0.226 ± 0.256	0.359 ± 0.03
R^2	0.978	0.989	2008	R^2	0.979	0.994
κ_2	3.539 ± 0.26	6.842 ± 0.117	2009	A	318.1 ± 351.9	38.2595 ± 7.46
χ	1.054 ± 0.02	0.713 ± 0.007	2009	γ	1.089 ± 0.02	0.783 ± 0.01
			2009	β	0.140 ± 0.225	0.390 ± 0.04
R^2	0.979	0.990	2009	R^2	0.980	0.993
κ_2	3.552 ± 0.27	6.929 ± 0.115	2010	A	321.8 ± 359.6	47.233 ± 8.93
χ	1.053 ± 0.02	0.709 ± 0.007	2010	γ	1.087 ± 0.02	0.815 ± 0.01
			2010	β	0.137 ± 0.227	0.406 ± 0.04
R^2	0.978	0.990	2010	R^2	0.980	0.993
κ_2	3.605 ± 0.27	7.031 ± 0.117	2011	A	432.8 ± 440.0	35.645 ± 7.28
χ	1.053 ± 0.02	0.709 ± 0.007	2011	γ	1.086 ± 0.02	0.780 ± 0.01
			2011	β	0.141 ± 0.23	0.406 ± 0.04
R^2	0.978	0.990	2011	R^2	0.980	0.993

Table 1: Examples of empirical rank-size law fit: the case of ATI for MOL region with $N = 136$ cities, with either 2 or 3 parameters, taking into account all data points but also when excluding the K and 3 VR cities mentioned in the text. Data are expressed in Euros.

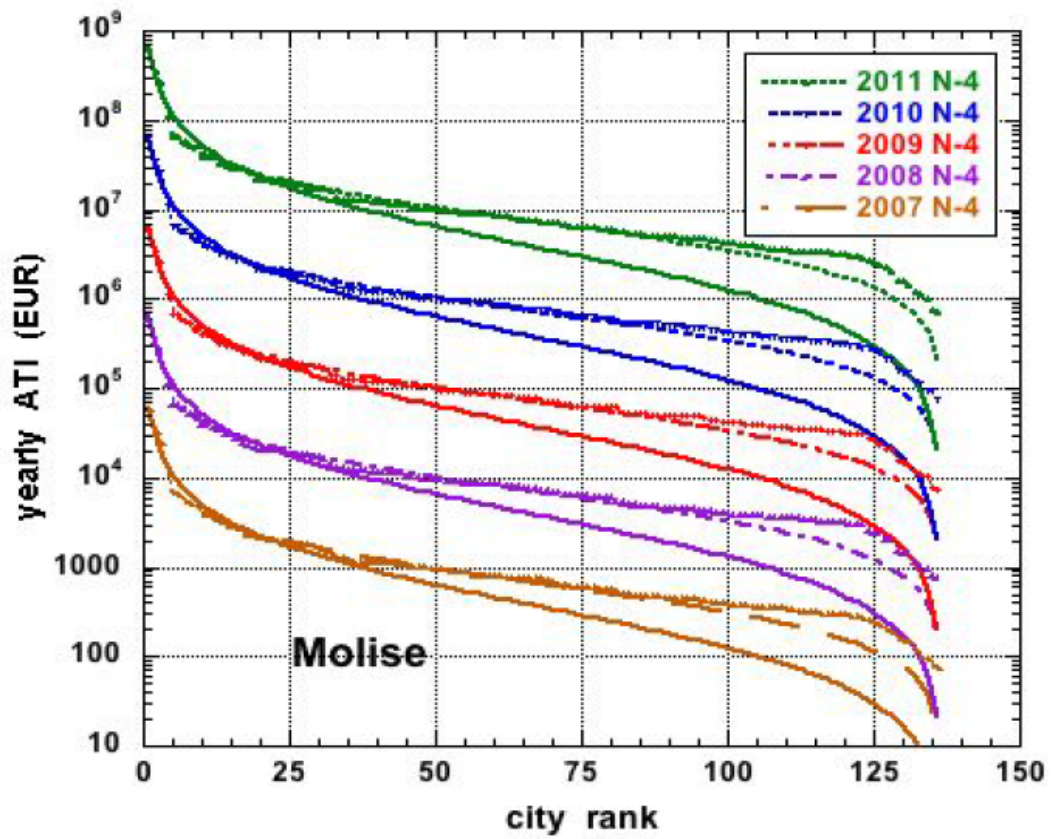


Figure 3: Displaying the rank-size law of the yearly ATI for Molise, with $N = 136$ cities, but after removing the King and 3 Vice-Roy cities, in various recent years: data for 2007-08 and for 2010-2011 are displaced for readability; fit values for a 2-parameter function, as Eq. (3.1), are given for the (undisplaced) data in Table 1.

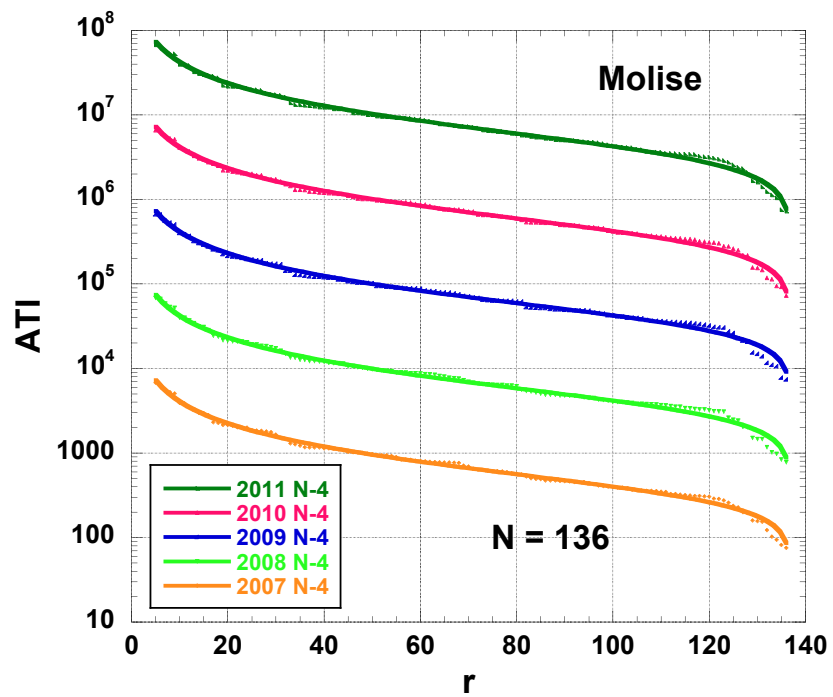


Figure 4: Displaying the rank-size law of the ATI for Molise, with $N = 136$ cities, but after removing the King and 3 Vice-Roy cities, in various recent years: data for 2007-08 and for 2010-2011 are displaced for readability; fit values for a 3-parameter function, as Eq. (2.1), are given for the (undisplaced) data in Table 1.

4 Financial inequalities

From the above it appears that a rank-size law can be useful for examining the evolution of entities in a given ensemble. Other techniques are based on the values leading to a ranking. These techniques are mainly based on a few indicators: the Theil index [7], a Herfindahl-Hirschman index [8] and the Gini coefficient [9]. They stem from wealth distribution investigations and indicate (financial) inequalities or disparities.

4.1 Theil index

The Theil index represents one of the most common statistical tools to measure inequality among data (Miskiewicz 2008, Clippe and Ausloos 2012). Basically it represents a number which synthesizes the degree of dispersion of an agent in a population with respect to a given variable (= measure). The most relevant field of application of the Theil index is represented by the measure of income diversity. The Theil index [7] is defined as:

$$Th = -\frac{1}{N} \sum_{i=1}^N \frac{y_i}{\sum_j y_j} \cdot \ln \left(\frac{y_i}{\sum_j y_j} \right) \quad (4.1)$$

where y_i is the ATI of the i -th city, in our case, and the sum $\sum_j y_j$ is the aggregation of ATI in regions or in the entire Italy; N is the relevant number of cities when aggregating, according to the j set of interest.

Moreover, it can be easily shown that the Theil index can be expressed in terms of a negative entropy

$$H = - \sum_i \frac{y_i}{\sum_j y_j} \cdot \ln \left(\frac{y_i}{\sum_j y_j} \right) \quad (4.2)$$

where $\frac{y_i}{\sum_j y_j}$ is the "market share" of the i -th city, thus indicating a deviation from the "ideal" maximum disorder, $\ln(N)$:

$$H = \ln(N) - Th \text{ or } Th = \ln(N) - H. \quad (4.3)$$

4.2 Herfindahl index

The Herfindahl index, also known as Herfindahl-Hirschman index (HHI), is a measure of concentration [8]. It is usually applied to describe company sizes (\sim *concentration*) with respect to the entire market: a HHI index below 0.01 indicates a highly competitive index. (From a portfolio point of view, a low HHI index implies a very diversified portfolio).

Adapted to the case of ATI of cities, HHI is an indicator of the amount of competition (for wealth, here) among municipalities in a region or in the entire country. The higher the value of HHI, the smaller the number of cities with a large value of ATI, the weaker the competition in concurring to the formation of Italian GDP. Formally:

$$HHI = \sum_{i \in L_{50}} \left(\frac{y_i}{\sum_j y_j} \right)^2, \quad (4.4)$$

where L_{50} is the set of the 50 largest cities in terms of ATI, and y_i is the ATI of the i -th city. The value 50 is conventional, whence HHI in Eq. (4.4) is the sum of the squares of the market shares of the 50 largest cities, when the market shares are expressed as fractions.

A normalized Herfindahl index is sometimes used and defined as:

$$H^* = \frac{(HHI - 1/N)}{1 - 1/N}. \quad (4.5)$$

with the appropriate N .

4.3 Gini coefficient

The Gini index (Gi) [9] can be viewed as a measure of the level of fairness of a resource distribution among a group of entities. Referring to the specific case treated here, it can be defined through the Lorenz curve, which plots the proportion f of the total Italian ATI that is cumulatively provided by the bottom $x\%$ of the cities. If Lorenz curve is a line at 45 degrees in an $f(x)$ plot, then there is perfect equality of ATI. The Gini coefficient is the ratio of the area that lies between the line of equality and the Lorenz curve over the total area under the line of equality. A Gini coefficient of zero, of course, expresses perfect equality, i.e. all ATI values are the same, while a Gini coefficient equal to one (or 100%) expresses maximal inequality among values, e.g. only one city contributes to the total Italian ATI.

A new type of display is hereby proposed. It seems of interest, for emphasizing the structure (like the maximum position and the corresponding percentage of relevant population) to display the data as the difference between the Lorenz curve (L_r) and the line of perfect equality in ATI:

$$\Delta L_j = \frac{1}{N_{c,r}} \left[\sum_{i=1}^j i y_{i,r} - j \right] \quad (4.6)$$

with $j \leq N_{c,r}$.

4.4 Two cases: IT and MOL

As an illustration of the values of such indices, let us give them, and those of the corresponding quantities of interest, in Table 2, when calculated for the 5 years of the quinquennium for IT and MOL.

It is observed that the yearly variations are rather small. The Theil index is about 1.275. The Gini coefficient is $\simeq 0.75$. The HHI is approximately given by $7 \cdot 10^{-3}$ for IT and $7 \cdot 10^{-2}$ for MOL. Moreover, in each case, the indices of MOL have a value lower than the corresponding one for IT. The scale effect is the greatest in HHI. Indeed, MOL is known as one of the poorest regions in Italy. Molise is mostly mountainous; the economy relies heavily on agriculture and livestock raising, although food and garment industries are undergoing some development. The 4 richest cities are markedly dominating in terms of wealth, thereby leading to heterogeneity, reflected in the indices values.

The yearly variation of the Gini coefficient for IT has been presented in [1]. The proposed way, according to Eq.(4.6), is shown on Fig. 5 for IT for the 5 years of interest. It can be observed that the maximum of such a curve occurs

MOL	2007	2008	2009	2010	2011	$\langle 5yav \rangle$
Entropy	3.6245	3.6314	3.6371	3.6407	3.6396	3.6350
Max. Entropy	4.9127	4.9127	4.9127	4.9127	4.9127	4.9127
Theil index	1.2882	1.2813	1.2756	1.2719	1.2730	1.2777
$10^2 HHI$	7.6722	7.6097	7.6336	7.6014	7.5998	7.6225
$10^2 H^*$	6.9883	6.9253	6.9494	6.9170	6.9153	6.9382
Gini Coeff.	0.7007	0.6989	0.6959	0.6957	0.6967	0.6975
whole IT	2007	2008	2009	2010	2011	$\langle 5yav \rangle$
Entropy (H)	7.2476	7.2603	7.2659	7.2669	7.2826	7.2650
Max. Entropy	8.9986	8.9986	8.9986	8.9986	8.9986	8.9986
Theil index	~ 1.751	1.7383	1.7327	1.7317	1.7160	1.7336
$10^3 HHI$	7.332	7.236	7.205	7.230	7.115	7.222
$10^3 H^*$	7.209	7.113	7.083	7.107	6.992	7.099
Gini Coeff.	0.7591	0.7576	0.7566	0.7565	0.7547	0.75685

Table 2: Statistical characteristics of the ATI data distribution as a function of time for the Molise region and for the whole Italy: $N= 136$ and 8092 , respectively. Entropy is H (see Eq. (4.3)); Max. Entropy $\equiv \ln(N)$. Theil index is taken from Eq. (4.1); the Herfindahl index is HHI (see Eq. (4.4)); the normalized Herfindahl index is H^* (see Eq. (4.5))

at 0.81 of the maximum ATI. The corresponding case for Molise is shown in Fig. 6, but only for the 5 year average data, since it can be imagined that, like for IT, the Gini coefficient of MOL hardly changed during the quinquennium; see Table 2 for such a justification. However, the maximum of the ΔL_j curve occurs at 0.83 of the maximum ATI, - which is after finer examination slightly different from IS and CB, and MOL. It is likely obvious, *but a posteriori*, that the overall Gini coefficient and ΔL_j curve are much influenced by the larger partner in the examined set, as shown for the case of CB in MOL in Fig. 6.

5 Benford Law

Finally, one statistical tool which can serve as a quality/reliability rapid data check is the so-called Benford's law [10, 11]. First digit Benford law for the cumulated distribution of the Molise 186 cities AIT each year of the 2007-2011 quinquennium is shown in Fig.7: marked deviations are seen, in contrast to other IT regions, as discussed in [6], even though in these some ATI manipulation could have been expected.

6 Conclusion

In the main text, we have applied techniques to investigate regional wealth and administrative (demographically based) disparities. through recent ATI data in Italy over the time interval 2007-2011. The rank-size rule is discussed. Yearly data of the aggregated tax income when transformed into the Gini, Theil, and Herfindahl-Hirschman indices has been analyzed. One IT region has been selected for illustration and discussion: Molise. A note on the "first digit Benford

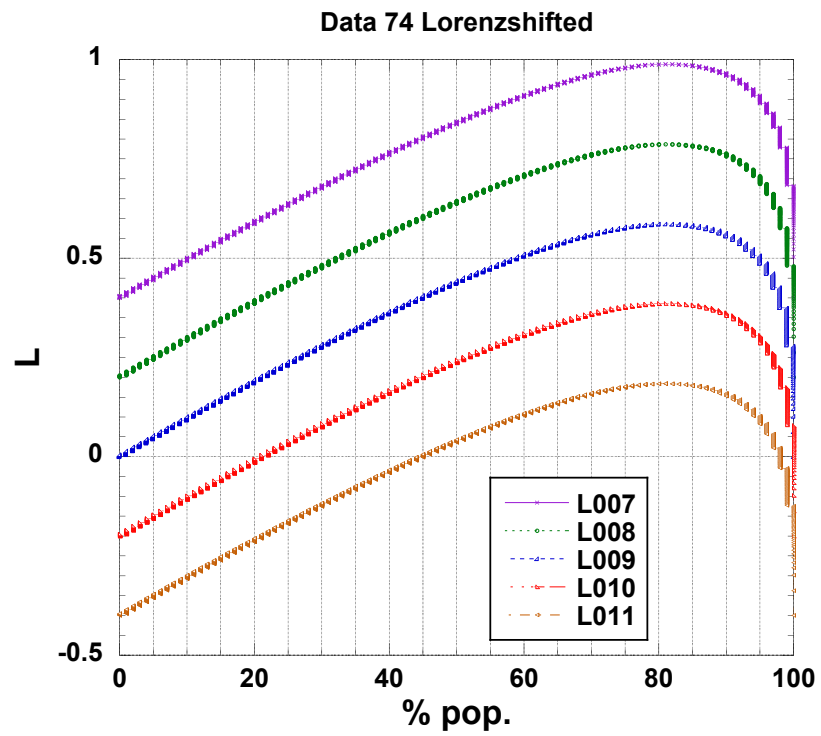


Figure 5: Another way of representing the Lorenz curve, through Eq.(4.6), leading to the Gini coefficient: the figure presents the case of the cumulated distribution of the AIT in IT 8092 cities during the 5 years of the 2007-2011 quinquennium; the data (except L009) is displaced for readability

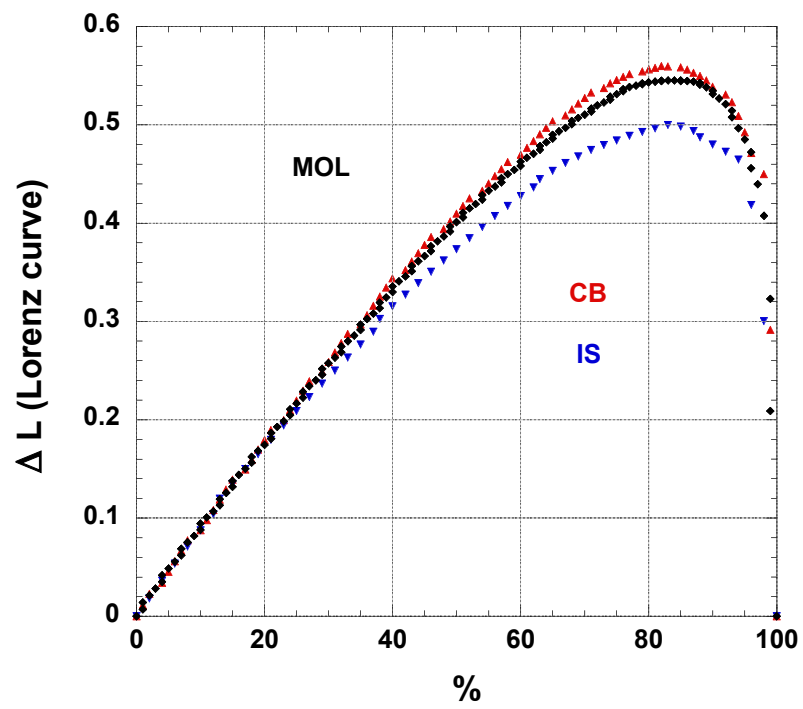


Figure 6: So called ΔL Lorenz curve, Eq.(4.6), for the cumulated distribution of the AIT in MOL (136 cities), but also for IS (52 cities) and for CB (84 cities) cities during the 5 years of the 2007-2011 quinquennium

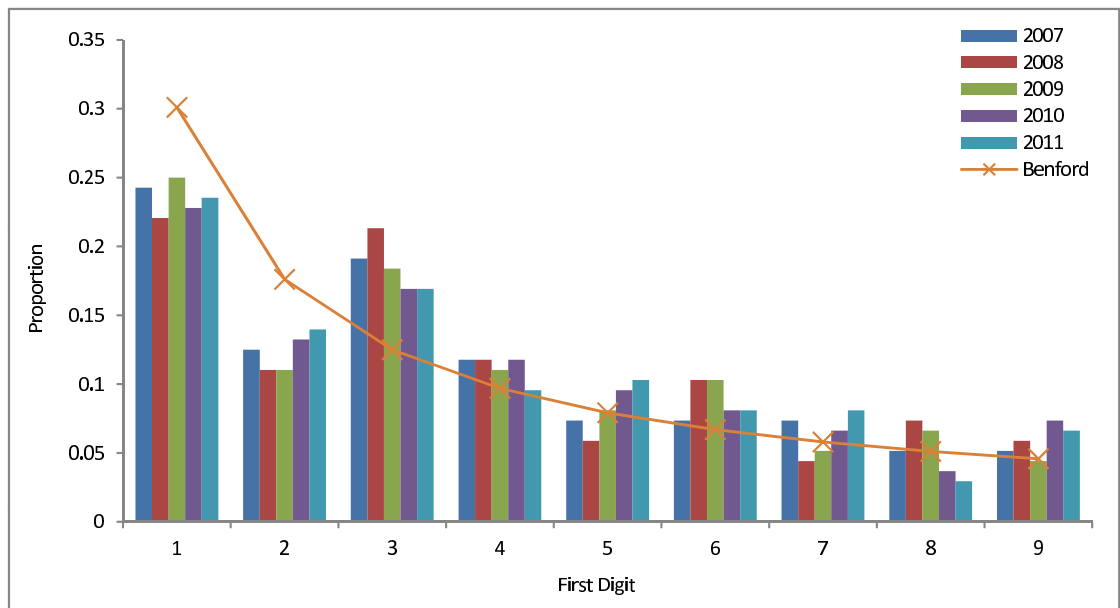


Figure 7: First digit Benford law for the cumulated distribution of the Molise 186 cities AIT each year of the 2007-2011 quinquennium

law” for testing data validity has been presented, indicating an unexpected behavior for the Molise region

Of course, there are other data and means to discuss the wealth of regions, like through measuring territorial dispersion rates of small and medium-sized enterprise creation patterns [21] or as such SME creation related to innovation outputs [22].

Acknowledgments

MA thanks the FENS 2015 organizers, in particular R. Kutner and D. Grech, for their invitation (and as usual very warm welcome) to the Rzeszow meeting. Thanks to K. Kulakowski for providing extra logistic means. The Benford Law plot (Fig.7) is courtesy of T.A. Mir.

References

- [1] R. Cerqueti, M. Ausloos, Statistical Assessment of Regional Wealth Inequalities: the Italian Case, *Quality and Quantity*, 49(6), 2307-2323 (2015). doi: 10.1007/s11135-014-0111-y
- [2] R. Cerqueti and M. Ausloos, Evidence of economic regularities and disparities of Italian regions from aggregated tax income size data, *Physica A* 421 (2015) 187-207. doi:10.1016/j.physa.2014.11.027
- [3] R. Cerqueti and M. Ausloos, Cross Ranking of Cities and Regions: Population vs. Income, *J. Stat. Mech.* 7 (2015) P07002

- [4] R. Cerqueti and M. Ausloos, Socio-economical Analysis of Italy: the case of hagiotoponym cities, *The Social Science Journal* (2015). doi:10.1016/j.soscij.2015.03.004
- [5] M. Ausloos and R. Cerqueti, Religion-based Urbanization Process in Italy: Statistical Evidence from Demographic and Economic Data, *Quality and Quantity* in press (2015). doi: 10.1007/s11135-015-0220-2
- [6] T.A. Mir, M. Ausloos, R. Cerqueti, Benford's law predicted digit distribution of aggregated income taxes: the surprising conformity of Italian cities and regions. *Eur. Phys. J. B* 87 (2014) 26
- [7] H. Theil, *Economics and Information Theory*, Chicago: Rand McNally and Company (1967).
- [8] A.O. Hirschman, The paternity of an index, *The American Economic Review* 54(5) (1964) 761-762.
- [9] C. Gini, ndice di Concentrazione e di Dipendenza. *Biblioteca dell'Economista*, serie V, vol. XX, Utet, Torino. (in Italian) (1910); English translation in *Rivista di Politica Economica* 87 (1997), 769-789.
- [10] S. Newcomb, Note on the frequency of use of different digits in natural numbers, *Am. J. Math.* 4 (1881) 39-40.
- [11] F. Benford, The law of anomalous numbers, *Proc. Am. Phil. Soc.* 78 (1938) 551-572.
- [12] M. Jefferson, The law of primate city, *Geographical Review*, 29(2),226-232 (1939).
- [13] M. Ausloos, A scientometrics law about co-authors and their ranking: the co-author core, *Scientometrics* 95 (2013) 895-909.
- [14] I. Popescu, M. Ganciu, M.C. Penache, D. Penache, On the Lavalette ranking law, *Romanian Reports in Physics* 49, 3-27 (1997).
- [15] I. Popescu, On a Zipf's Law Extension to Impact Factors, *Glottometrics* 6, 83-9 (2003).
- [16] R. Mansilla, E. Köppen, G. Cocho, P. Miramontes, 2007. On the behavior of journal impact factor rank-order distribution, *Journal of Informetrics* 1, 155-160 (2003).
- [17] I.A. Voloshynovska, Characteristic Features of Rank-Probability Word Distribution in Scientific and Belletristic Literature, *Journal of Quantitative Linguistics* 18, 274-289 (2011).
- [18] M. Ausloos, Toward fits to scaling-like data, but with inflection points & generalized Lavalette function, *Journal of Applied Quantitative Methods* 9 (2014) 1-21.
- [19] P. Wessa, Kendall τ Rank Correlation (v1.0.11) in Free Statistics Software (v1.1.23-r7), Office for Research Development and Education (2012) [http : //www.wessa.net/rwasp_kendall.wasp/](http://www.wessa.net/rwasp_kendall.wasp/).

- [20] P. Wessa, Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7, (2014) [http : //www.wessa.net/](http://www.wessa.net/).
- [21] M. Nistotskaya, N. Charron, and V. Lapuente, The wealth of regions: quality of government and SMEs in 172 European regions, *Environment and Planning C: Government and Policy* 32 (2014).
- [22] M. Ausloos, F. Bartolacci, N. G. Castellano, and R. Cerqueti, Effects of Innovation on Performance from STAR Market Ranking of Companies, working paper