VOLUME LXXV – N. 4 OTTOBRE – DICEMBRE 2021

RIVISTA ITALIANA DI ECONOMIA DEMOGRAFIA E STATISTICA

Sede Legale: C/O Studio Associato Cadoni, Via Ravenna n. 34 – 00161 ROMA.
info@sieds.it, rivista@sieds.it

SIEDS
SOCIETÀ ITALIANA
DI ECONOMIA DEMOGRAFIA E STATISTICA

SEDE LEGALE:
C/O Studio Associato Cadoni, Via Ravenna n. 34 – 00161 ROMA

info@sieds.it

rivista@sieds.it

VOLUME FUORI COMMERCIO – DISTRIBUITO GRATUITAMENTE AI SOCI

INDICE

AN ATTEMPT TO QUANTIFY THE TECHNOLOGICAL CHANGE IN ITALY THROUGH A MULTISECTORAL FRAMEWORK: A COMPARATIVE ANALYSIS[1]

Clio Ciaschini, Margherita Carlucci, Francesco M. Chelli,
Luca Salvati, Giuseppe Ricciardo Lamonica

1. Introduction

In recent years, European countries are experiencing the so called "Fourth Industrial Revolution" that is meant to have a pervasive impact, in its magnitude and ramification, on all the aspects of the society (Schwab, 2017). Its effects, indeed, would spread in a pervasive way in both private and public sectors, including academia and civil society. This Revolution is intended to be the final point of a slow process that has begun in 1760 with the First Industrial Revolution that introduces the mechanization of production by means of water and steam power. The further development goes in the direction of the introduction of electricity in the production and constitutes the Second Industrial Revolution. The Third Industrial Revolution introduces the automated production by means of information and communication technology. (Mattioli, Lamonica, 2013). This automation tendency would be completed in the Fourth Industrial Revolution that is expected to merge the physical, digital, and biological areas. (Xu, 2018). Even if this latest Revolution is intended to develop the process started during the Third one, it is aimed to create a discontinuity with the previous one in relation to the speed, aim and impact on the whole society. At first, the speed of the recent innovations has an exponential rate, if compared to the previous Revolutions in every industry and country. The strength of the transformations involves the whole productive system and governance. This new environment increases and widens the possibilities of people to access to knowledge by means of mobile devices with extraordinary processing power. This disruptive force emerged by new technology, such as artificial intelligence, robotics, the Internet of Things, and biotechnology. (Wortmann and Flüchter, 2015). Recently, large progress has developed in Artificial Intelligence encouraged by computing power and the huge availability of data, from software to find new drugs to algorithms to calculate cultural interests and monitor life habits of Internet users. The

[1] C.C. conceived the idea. C.C. and G.R.L. developed the theory, performed the computations, and discuss the results. M.C. and F.M.C. supervised the work. L.S. helped in the final revision of the manuscript.
.

Fourth Industrial Revolution could raise income levels and the quality of life of people around the world. Technological progress will also influence the supply side with long term gains in efficiency and productivity. The decrease of costs of transportation and communication and the implementation and development of global supply chains stimulate the emergence of new markets and lead the economic growth. The negative aspect is to be found in the labour market. Machines will substitute the so called "blue collars" workers and there will be a gap between returns to capital and labour. On the other side, new skills will emerge that can include the so called "superstar workers", characterised by higher skilled tasks not subjected to the substitutability between human and machine. (Brynjolfsson and McAfee, 2014).

Our aim is to provide a quantitative and comparative picture of the transition from the Third to the Fourth Industrial Revolution. Such transition is important since, even if they are two distinct phases, the second one is grounded on the first and each country attains the second at different times, according to its performance. (Popkova et al., 2019). For a comparative analysis with our country, we have chosen The Netherlands: this country occupies the 4th position in the Global Competitive Index (hereafter GCI) ranking of 2019, behind Singapore the United States and Hong Kong, overtaking Germany and Switzerland. (Schwab, 2019). Its innovation capability is at 10th place in the world and its ICT adoption is at 24th place. Opposite, Italy is graded at the 30th place (53rd in ICT adoption and 22nd place in Innovation Capability). This work starts from the observation that, starting from the early 2000, the different countries invest in ICT in different time periods and at different time lags, so the technological outcomes do not find immediate and evident feedbacks in the macroeconomic data. To reach this aim, focusing on the period of the Third Industrial Revolution, we would quantify, if possible, and at what extent, the advancement of the technological progress in Italy, based on observed data, in relation to The Netherlands.

In this work, the linkage analysis has been performed. Hauknes and Knell, (2009), define intersectoral linkages as "techno-economic connections between industries, embodied in the exchange of tangibles and intangibles". The linkage analysis, pioneered by Leontief in 1936, (Leontief, 1941), began to spread starting from 1950s, in consequence of the acceleration of industrialisation in developing countries. Industrialisation, indeed, induced scholars to focus on how to quantify the relative importance of the various industries, to the aim of identifying key industries for economic development and high-speed economic growth, (Cuello et al., 1992). In particular, the linkage analysis aims at assessing the relationship between and within industries and evaluating the role of the industries and the performance of the economic system and optimize the industrial structure of the national economy. Our work is based on two approaches belonging to the "Classical Multiplier Method", i.e., the Rasmussen and Laumas methods, and compares two Leontief- based

linkages showing the differences between a technology-based approach (Rasmussen linkages) and the approach that weights Rasmussen's Forward Linkages for value added and Rasmussen's Backward Linkages for final demand. (Rasmussen, 1956; Laumas, 1976). These traditional outcomes of Input- Output analysis, today acquire a renewed relevance in the study of value chains in terms of upstream and downstream position. (Guerrieri and Meliciani, 2005).

The paper is divided in four Sections. Section 2 describes data and methodology adopted in the work; Section 3 shows the results of linkages analysis. In Section 4, the conclusion.

2. Data description and methodology

Data used in our analysis belong to the World Input-Output Database, (WIOD) that provides an historical series of Input-Output tables of all the European countries. (Timmer et al, 2015). The choice of this database is linked to the fact that it provides a homogeneous statistical basis through which the outcomes can be conveniently compared. From this database we retrieved national Input-Output tables of Italy and the Netherlands. To detect the impact of ICT on macroeconomic data, we have analyzed the four ICT sectors in the 56 sectors classification of WIOD, shown in Table 1.

Table 1 – *ICT Sectors included in Section J of the International Standard Industrial Classification of All Economic Activities Rev. 4 (ISIC 4.0)*

ISIC Code	**Section J — INFORMATION AND COMMUNICATION**
Division 58	Publishing activities
Division 59	Motion picture, video and television programme production, sound recording and music publishing activities
Division 61	Telecommunications
Division 62	Computer programming, consultancy and related activities

In addition, with reference to the recommendations of the International Standard Industrial Classification of All Economic Activities Rev. 4 (ISIC 4.0), and following De Siena (2019), we have further extracted four ICT components from four macro sectors, two sectors related to manufacture, i.e. Manufacture of computer, electronic and optical products and Repair and installation of machinery and equipment, and two sectors related to trade, i.e. Wholesale trade, except of motor vehicles and motorcycles and Retail trade, except of motor vehicles and motorcycles. From each of these four sectors we extract a portion that we attribute to ICT. This component

has been obtained, aggregating the four sectors of Section J- ICT and calculating the weights of ICT included in each of these four sectors. By means of these weights, we decompose each of these four sectors, in the part of the sector linked to ICT, and the sector not related to ICT. The newly defined *ICT- related* sectors are shown in Table 2.

Table 2 – *"ICT- related" sectors, our elaborations on International Standard Industrial Classification of All Economic Activities Rev. 4 (ISIC 4.0)*

ICT related sectors	
Division 26- ICT	Manufacture of computer, electronic and optical products
Division 33-ICT	Repair and installation of machinery and equipment,
Division 46-ICT	Wholesale trade, except of motor vehicles and motorcycles
Division 47-ICT	Retail trade, except of motor vehicles and motorcycles

This new classification leads to a 60 sector Input- Output matrix both for Italy and the Netherlands.

The methodology consists in the linkage analysis. Intersectoral linkages are defined as techno economic connections between industries, determined by the exchange of goods and services. Linkage analysis allows to define and quantify the role of each sector in relation to the provider sectors, upstream position, and customer sectors, downstream position (Hauknes, 2009; Reis and Rua, 2009).

The basic model for the linkage analysis is the Leontief model (Leontief, 1941) that reads as:

$$x = Ax + f \tag{1}$$

where x defines the vector of total output by industry and equals the intermediate consumption vector, Ax, plus the final demand vector by industry, f.
Starting from equation (1), we can write:

$$x = Rf \tag{2}$$

where: $R = (I-A)^{-1}$ is the Leontief Inverse, i.e., the matrix of direct and indirect intermediate requirements per unit of output. This matrix provides a fundamental tool for the quantification of the relevance of every industry in stimulating upstream the other industries of the economy and contributing downstream to the provision and realization of the sectoral outputs.

Our work is based on the Rasmussen Linkages, (Rasmussen, 1956). This approach is directly based on the Leontief inverse, and for that reason is referred to as technological linkage.

Rasmussen approach provides two indexes, the Backward Linkage that indicates the average direct and indirect impact on the whole productive system, in terms of average purchases of intermediate goods, due to a unitary increase in final demand of the good produced in a given sector. Mathematically:

$$BL_j = \frac{\frac{1}{n}\sum_{i=1}^{n} R_{ij}}{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} R_{ij}} \tag{3}$$

The Backward Linkage of sector *j* is obtained computing the average of the column of the Leontief inverse and dividing matrix *R* by the average of the whole matrix, average coefficient of the sector divided by average coefficient of the whole economy).

The Forward Linkage defines the average direct and indirect impact on the whole productive system, in terms of sales of intermediate good by the sector to the other sectors in the economy, due to a unitary increase in the final demands for the goods produced by all the other sectors. In formulas:

$$FL_i = \frac{\frac{1}{n}\sum_{j=1}^{n} R_{ij}}{\frac{1}{n^2}\sum_{j=1}^{n}\sum_{j=1}^{n} R_{ij}} \tag{4}$$

The Forward Linkage of sector *i* results from the average of the row of the Leontief inverse and dividing it by the average of the coefficients of the whole matrix.

A further analysis will be performed with reference to the linkage weighted by the shares of value added and final demand, (Laumas, 1976). This method highlights the relevance of each sector in facing two basic instances of economic content: the final demand provision and the income generation (value added). Rasmussen linkages are, indeed modified through the correction of each index with the weight of the sector in satisfying its final demand and value added. In particular, the Backward Linkage weighted by Value Added reads as:

$$BL_j^{va} = \frac{\frac{1}{n}\sum_{i=1}^{n} w_j R_{ij}}{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_j R_{ij}} \tag{5}$$

where: $w_i = \frac{va_i}{\sum_{i=1}^{n} va_i}$ is the weight represented by the share of value added of sector *i* over the total value added. The so obtained weighted backward linkage is determined scaling the Leontief inverse by column, with the percentage weight of

the sectoral value added with respect to the total value added and dividing the average of the column by the average of the matrix.

The Forward Linkage weighted by the final demand is determined analytically as:

$$FL_i^f = \frac{\frac{1}{n}\sum_{j=1}^{n} w_i R_{ij}}{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i R_{ij}}$$

where: $w_j = \frac{f_j}{\sum_{i=1}^{n} f_j}$ defines the weight, i.e., the share of final demand of sector *j* over the total final demand. This index is obtained scaling the Leontief by row using the percentage of the final demand of the given sector over the total final demand and dividing the average of the row by the average of the matrix.

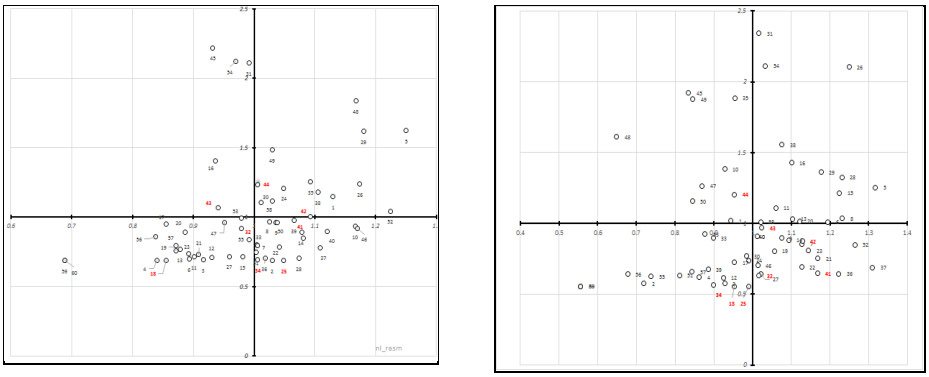
3. Results of Linkage Analysis

Figure 1 shows the outcomes of Rasmussen analysis with reference to the two countries under scrutiny, i.e. The Netherlands and Italy. In the graph, the axes are centred on the value 1. Sectors in the first quadrant have both backward and forward linkages higher than 1, and are defined, following Rasmussen, *Key Sectors*.

This group of sectors sells and purchases goods and services in an amount higher than the average of all the other sectors. The second quadrant refers to the *Prime Vendors*, i.e., those sectors that sell by an amount higher than the average and purchases for an amount lower than the average of the other sectors. In the third quadrant we can find the *Low Impact Sectors* that sell and purchase goods and services for an amount lower than the average of the other sectors. The fourth quadrant comprises the *Prime Users*, that sell for an amount lower than the average and buy for an amount higher than the other sectors.

In the two Figures, *Key Sectors* are highlighted in red. These Sectors constitute 25% of the total for both Italy and the Netherlands. The difference between the outcomes of the two countries consists in the composition of the sectors. *Key Sectors* for Italy are, in prevalence, traditional sectors, linked to the so called *Made in Italy*, such as: *S6-Manufacture of textiles, wearing apparel and leather products, S8-Manufacture of paper and paper products, S11-Manufacture of chemicals and chemical products, S13-Manufacture of rubber and plastic products, S15-Manufacture of basic metals* and *S16- Manufacture of fabricated metal products, except machinery and equipment.* Differently from Italy, The Netherlands seems to show a higher propensity to services sectors.

Figure 1 – *Outcomes of Rasmussen Linkages for The Netherlands (a) and Italy (b).*

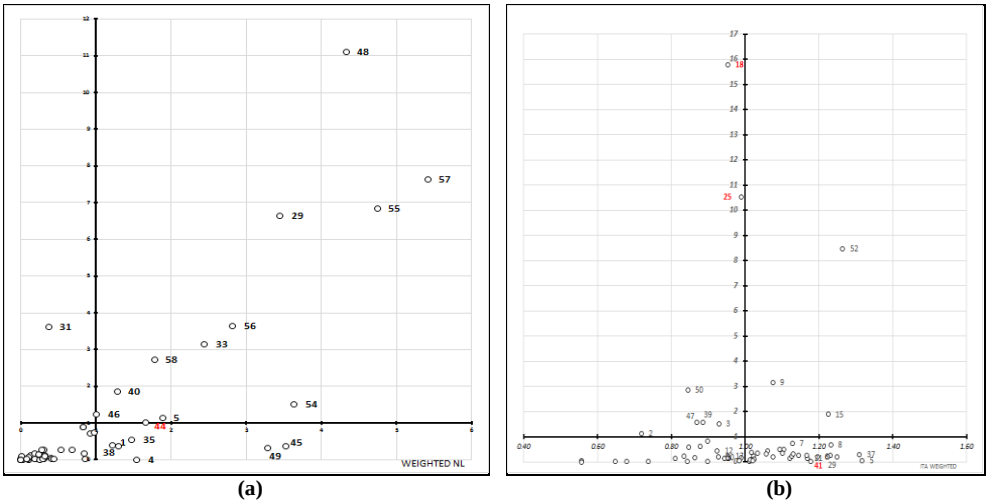


(a) **(b)**

In The Netherlands, *Key Sectors* group comprises sectors as *S48-Real estate activities and S49-Legal and accounting activities; activities of head offices; management consultancy activities.* In addition, this group includes also *S42-Motion picture, video and television programme production, sound recording and music publishing activities; programming and broadcasting activities* and *S44-Computer programming, consultancy, and related activities; information service activities*, these sectors, as described in Section 2, belong to the ICT sectors of Section J of ISIC Rev.4. Figures 2(a)- (b) shows the outcome of the Weighted Rasmussen Linkages. Using these indexes, the situation changes radically. Figure 2(a) shows an outcome that seems to underline, even more than that of Rasmussen approach, a propensity of The Netherlands to the shift to sectors linked to services, in terms of fulfilment of final demand and creation of value added. Indeed, within the twelve *Key Sectors* (almost the 20%), retrieved for this country, we can find, *S55- Public administration and defence; compulsory social security*, *S56- Education* and *S57- Human health and social work activities.*

In Italy, as highlighted by Figure 2(b), with this approach, there are only three *Key Sectors*, (5% of the total), i.e., *S9- Printing and reproduction of recorded media*, *S15- Manufacture of basic metals* and *S52- Advertising and market research.* It is worth noting, that, almost all the sectors related to "Made in Italy", even losing their status of *Key Sectors*, keep a relevant position as *Prime Users*. With reference to ICT, the Figures show that, in Italy, there are no ICT-sectors within *Key Sectors* group, while *S44-Computer programming, consultancy, and related activities; information service activities*, is, also in the weighted approach, included in *Key Sectors* for The Netherlands.

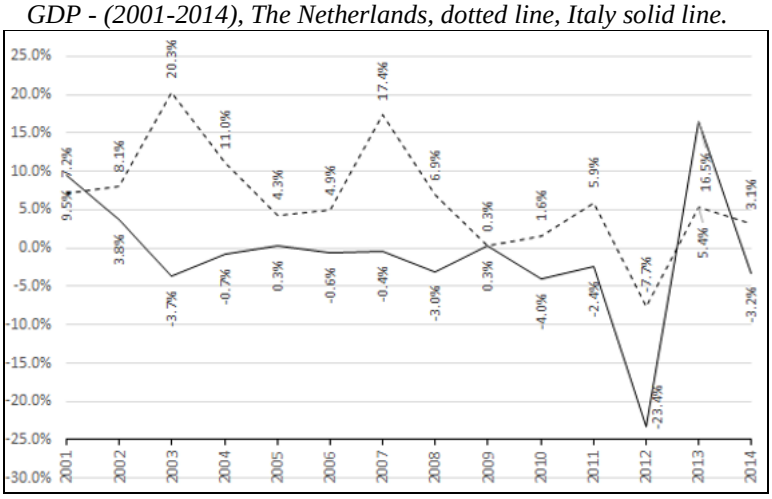
Figure 2 – *Outcomes of Weighted Rasmussen Linkages, The Netherlands (a) and Italy (b).*



(a) **(b)**

The last analysis has been performed with reference to the historical series of the percentage change of the output in ICT with respect to GDP.

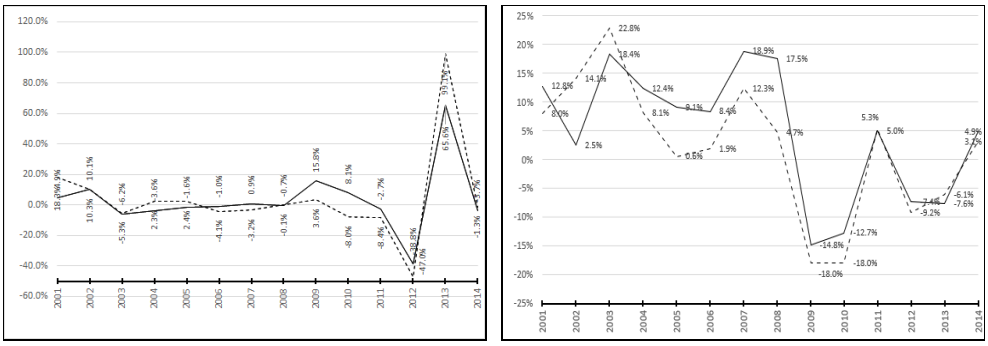
Figure 3 – *Percentage change of Information and Communication Technology output w.r.t. GDP - (2001-2014), The Netherlands, dotted line, Italy solid line.*



As highlighted by Figure 3, in The Netherlands the growth rate is positive in most of the time interval, while Italy always performed a negative, or nearly zero growth rate. Both of the countries exhibit a negative growth rate around year 2012, probably

in consequence of the crisis of subprime mortgages of 2007-2008 and then there is a recovery. The already discussed tendency is also confirmed by the historical series of the percentage changes of private investments and public expenditure in ICT with respect to GDP. Figures 4(a) and 4(b), indeed, show a nearly zero growth rate for Italy and a positive rate for The Netherlands, except for the years after 2007-2008, with negative growth rates in consequence of the economic crisis. In the latest years, Figures show a slight recovery in both countries.

Figure 4 – *Perc. change of investments (solid line) and government. final expenditure (dotted line) in ICT w.r.t. GDP, (a) Italy and (b) The Netherlands, (2001-2014).*



(a) **(b)**

4. Conclusion

In recent years, studies on technological progress highlight the emergence of the era of Fourth Industrial Revolution focused on digitalisation of productive processes. This new phase of Industrial Revolution is intended to lead the transition from Human to Machine to Machine to Machine transformation.

The aim of the work is to concentrate on the previous phase of the Industrial Revolution, i.e., the Third Industrial Revolution considered as a preparatory phase of provision of new tools for the information and communication, at the basis of the expected future technological advance.

The aim of this work is to evaluate quantitatively, as possible, the place of the Italian economy in implementing the Information and Communication Technology, comparing the Italian outcomes with those of The Netherlands, a European country that occupies the 4th place, behind Singapore, the United States and Hong Kong, overtaking Germany and Switzerland. Its innovation capability is at 10th place in the world and its ICT adoption is at 24th place.

The analysis adopts the multisectoral Input-Output viewpoint, starting from the last available Input- Output Flow Tables retrieved from the WIOD database. This database provides a homogeneous statistical basis from which the sectoral macroeconomic results of the two countries can be conveniently compared.

The linkage analysis is referred to Rasmussen linkages, and then final demand and value-added weighted Rasmussen indexes. Emerging results from linkage analysis with Rasmussen show 25% of sectors that exhibit values of linkages higher than the average with respect to the other sectors in stimulating the whole productive system upstream and downstream for both the countries. Nevertheless, it is possible to highlight several peculiarities in the performance of Italian with respect to Dutch economy. In Italy, the majority of Key Sectors seem to pertain to the so called "Made in Italy", traditional sectors. These sectors are included within the category of Low Impact sectors in The Netherlands, where most sectors are linked to services.

Further differences between the two economies emerge when deepening the analysis by means of the weighted Rasmussen Linkages. In this case, Italy has only three Key Sectors (5% of all sectors), while The Netherlands seems to have twelve Key Sectors (20% of all sectors) dealing with services. With reference to the aim of detecting the emergence of ICT in empirical data, there are no ICT-sectors within Key Sectors in Italy while there are two ICT-related Key sectors for The Netherlands with Rasmussen and one ICT-related Key sectors following the weighted Rasmussen approach. The historical series analysis confirms a longer period interest for The Netherlands, in promoting investments and public expenditure in ICT sectors with respect to Italy, even if in recent years Italy exhibits a slight recovery maybe also due to the approaching innovative policies linked to the National Plan Industry 4.0. Further developments go in the directions of extending this comparative analysis to the other European countries.

References

BRYNJOLFSSON E., MCAFEE A. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies.* WW Norton & Company.

CUELLO F. A., MANSOURI F., HEWING, G. J. D. 1992. The identification of structure at the sectoral level: a reformulation of the hirschman-rasmussen key sector indices, *Economic Systems Research*, Vol.4, No.4, pp. 285–296.

DE SIENA L. 2019. Agglomeration Economies in the Italian ICT Sector: The Structural Features. *Rivista Geografica Italiana.*

GUERRIERI P., MELICIANI V. 2005. Technology and international competitiveness: the interdependence between manufacturing and producer services, *Structural Change and Economic Dynamics*, Vol.16, No.4, pp. 489–502.

HAUKNES J., KNELL M. 2009. Embodied knowledge and sectoral linkages: An input–output approach to the interaction of high-and low-tech industries. *Research Policy*, Vol.38, No.3, pp. 459-469.

ISIC Rev. 4 (ISIC 4.0) retrieved at: https://unstats.un.org/unsd/publication/seriesm/seriesm_4rev4e.pdf (accessed 11/06/2021).

LAUMAS P.S. 1976. An international comparison of the structure of production. *Economia Internazionale*, Vol.29, No.1-2, pp. 2–13.

LEONTIEF W. 1941. *The Structure of American Economy 1919-1929*. 1st edition, Cambridge, Mass., Harvard University Press.

MATTIOLI E., LAMONICA G.R. 2013. The ICT role in the Word economy: an input-output analysis. *Journal of World Economic Research*, Vol. 2, No.2, pp. 20–25.

POPKOVA E.G., RAGULINA Y.V., BOGOVIZ A.V. 2019. Fundamental differences of transition to industry 4.0 from previous industrial revolutions. In *Industry 4.0: Industrial Revolution of the 21st Century* (pp. 21-29). Springer, Cham.

RASMUSSEN P.N. 1956. *Studies in Inter-sectoral Relations*, E. Harck, Copenhagen, Denmark.

REIS H., RUA A. 2009. An input–output analysis: Linkages versus leakages. *International Economic Journal*, Vol. 23, No.4, pp.527-544.

SCHWAB K. 2019. The Global Competitiveness Report, WEF (2019) http://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2019.pdf (accessed 11/06/2021).

SCHWAB K. 2017. *The fourth industrial revolution*. Currency.

TIMMER M.P., DIETZENBACHER E., LOS B., STEHER. R, DE VRIES G.J. 2015. An Illustrated User Guide to the World Input Output Database: The Case of Global Automotive Production. *Review of International Economics*, Vol.23, No.3, pp. 575 -605. doi: 10.1111/roie.12178.

WORTMANN F., FLÜCHTER K. 2015. Internet of things. *Business and Information Systems Engineering*, Vol.57, No.3, pp. 221-224.

XU M., DAVID J. M., KIM S.H. 2018. The fourth industrial revolution: Opportunities and challenges. *International Journal of Financial Research*, Vol.9, No.2, pp. 90-95.

SUMMARY

An attempt to quantify the technological change in Italy through a multisectoral framework: a comparative analysis

The fourth Industrial Revolution, centred on the digitalisation of the productive processes, is intended to develop the process started during the Third one, but also to create a discontinuity with the previous one in relation to the speed, aim and impact on the whole society.

The aim of this work is to evaluate quantitatively, as possible, the place of the Italian economy in implementing the Information and Communication Technology, comparing the Italian outcomes with those of The Netherlands, that occupies the 4th place in the Global Competitive Index ranking.

Starting from the last available Input- Output Flow Tables retrieved from the WIOD database, an interindustry linkage analysis is performed, whose results could confirm and detail the different position of the two economies in the world ranking. The linkage analysis confirms the higher propensity of The Netherlands of investing in ICT and the major concern of Italy in the valorisation of the "Made in Italy".

Clio CIASCHINI, Polytechnic University of Marche, c.ciaschini@staff.univpm.it
Margherita CARLUCCI, "La Sapienza" University of Rome, margherita.carlucci@uniroma1.it
Francesco M. CHELLI, ISTAT, Polytechnic University of Marche, f.chelli@staff.univpm.it
Giuseppe RICCIARDO LAMONICA, Polytechnic University of Marche, g.ricciardo@staff.univpm.it
Luca SALVATI, University of Macerata, bayes00@yahoo.it.

GENDER EQUALITY IN THE ITALIAN REGIONS

Nicola Maria D'angelo, Tommaso Paoletti

1. Introduction

Gender inequality is a recent, important and debated phenomenon. In the last decades it has become a political and academic matter of interest, generating lively debates and plenty of indicators in order to find an objectively and accurate way to measure it (Bozzano, 2012). Indeed, the degree of gender disparity in both developing and developed countries, becoming one of the main requested conditions for an inclusive and sustainable society. Due to feminist movement and equal social policies, the inclusion gap between male and female has been reduced significantly compared to the past decades. Despite these gains, many challenges remain: women are still victims of gender-based violence, discrimination, social exclusion and less represented than men at all levels of political and economic leadership. Due to its discriminating nature, gender inequality is arguably an ethical issue, but also "an important economic, business and societal issue with a significant impact on the growth of nations" (Hausmann, Tyson, and Zahidi, 2007). Gender discrimination, indeed, is one of the causes of an economic slowdown development, social improvement and, more generally, of sustainable and fair nation (Kabeer and Natali, 2013; Moorhouse, 2017; Profeta, 2017; Di Bella, 2021).

A clear and precise definition of gender equality has been given in the United Nations Report of the Economic and Social Council (United Nation, 1997), where the gender equality is described as "The condition in which people receive equal treatment, with equal ease of access to resources and opportunities, regardless of gender… ". Moreover, other authoritative sources that explicitly mention the gender equality as a fundamental right to be guaranteed to all citizens are the Italian Constitution (Art. 3, 1948) and Universal Declaration of Human Rights (Art. 22, 1948). To validate the urgency of the abovementioned problem, in 2015 the Unites Nations countries fixed 17 Sustainable Development Goals (SDGs), including the Gender Equality, 'to achieve a better and more sustainable future for all' by 2030 (Di Bella, 2021).

Bases on these institutional documents, we selected the domains that better allow to describe and analyze this multidimensional and complex phenomenon: Education, Work, Power and Safety (Table 1).

Table 1 – *Sources, target and domains explicitly refer to Gender Equality.*

Source	Target	Domain
ITALIAN COSTITUTION	**Art. 34** – "Schools are open to everyone. Primary education, given for at least eight years, is compulsory and free of tuition. Capable and deserving pupils, even if lacking financial resources, have the right to attain the highest levels of education …"	**Education**
	Art. 37 – "Working women have the same rights and are entitled to equal pay as men for equal work …"	**Work**
	Art. 51 – "All citizens, regardless the gender, are eligible for public offices and for elective positions under equal conditions, according to the rules established by law …"	**Power**
UNIVERSAL DECLARATION OF HUMAN RIGHTS	**Art. 22** – "Everyone, as a member of society, has the right to social security and is entitled to realization, through national effort and international co-operation and in accordance with the organization and resources of each State, of the economic, social and cultural rights indispensable for his dignity and the free development of his personality."	**Safety**
	Art. 26 – "Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit."	**Education**
SDG	"End all forms of discrimination against all women and girls everywhere."	**Safety**
	"Ensure women's full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life."	**Power**

From these definitions, we have analyzed the gender phenomenon through a synthetic measure of disparity between male and female based on the methodology of composite indices, which allows us to find a suitable combination of the individual indicators, aimed at yielding a measurement of active inclusion either of male and female in Italian regions.

2. Domains and Indicators

Measuring gender inequality is, therefore, a very challenging task, because of its complex and multidimensional nature.

Analyzing gender accessibility to resources and opportunities, it is important to be aware of the contextual factors. From a statistical point of view, concepts like accessibility, active inclusion and participation are theoretical concepts and so not directly measurable. Consequently, to monitor the gender social inclusion it is necessary to implement statistical indicators to approximate the different dimensions of these concepts.

In accordance with the definition given and the reference sources, we focused the analysis on 4 domains (Table 2):

- **Education**, analyzing the NEETs and the dropout rate of students between 18-24 years[1], the graduation and master's degrees pass rate[2];
- **Work**, represented by the percentage gender gap salary[1], the average of weekly working hours (20-64 years), the employment rate and the percentages of part time contracts[2];
- **Power**, composed of gender percentages at Municipal Councils and Regional Councils, as well as the rate of entrepreneurship[3];
- **Security**, analyzing the quota of voluntary homicides in the family, the percentage of suicides of people aged 15 and over, the percentage of abuses from partners or ex-partners and the percentage of victims calling to 1522 (the anti-violence and stalking number)[2,4]

All indicators have been considered in the period within 2015 and 2017, except for the quota of victims of voluntary homicides in the family, which is taken from the average of this period.

In the table below (Table 2) we show the structure of single domains and indicators with their data source, highlighting their polarity:

- Positive polarity (+), i.e., an increment of the graduation rate will increase the Education rate of the same region;
- Negative polarity (-), i.e., if the NEET rate increase the Education rate will decrease.

[1] https://ec.europa.eu/eurostat/data/database/
[2] https://dati.istat.it/index.aspx?lang=it#
[3] http://amisuradicomune.istat.it
[4] https://www.dati.gov.it/

Table 2 –*Domains and Indicators with data sources. Polarity (+, -) is shown in brackets.*

DOMAINS	INDICATORS	SOURCES
EDUCATION (+)	NEET RATE (-)	Eurostat
	SCHOOL DROPOUT RATE (-)	
	GRADUATION PASS RATE (+)	ISTAT
	MASTER'S DEGREE PASS RATE (+)	
WORK (+)	GENDER GAP PAY (+)	Eurostat
	WEEKLY AVARAGE WORKING HOUR (+)	
	EMPLOYMENT RATE (+)	ISTAT
	PART TIME CONTRACTS (-)	
POWER (+)	MUNICIPAL COUNCILS (+)	ISTAT
	REGIONAL COUNCILORS (+)	
	ENTREPRENEURSHIP (+)	
SAFETY (+)	ABUSE CONSUMED BY PARENTS AND PARTNERS (-)	ISTAT and Ministry of the Interior
	VICTIMS CALLING 1522 (-)	
	QUOTA OF VICTIMS OF VOLUNTARY MURDER COMMITTED BY PARTNERS (-)	
	SUICIDES RATE (-)	

3. Methodology of analysis

The difficulty in finding disaggregated data by gender and by regions, together with the complexity of this phenomenon has led to identify a selection of specific domains that can describe the problem on the basis of the aforementioned theoretical sources.

Although awarded of the risk of simplifying such a complex and delicate issue through a synthetic measure, we considered crucial to have a statistical tool which allows to objectively communicate and compare this fragmented and multidimensional issue.

3.1 Original matrix

Firstly, we constructed the original matrix (40x15) for each year considered (2015-2017), where rows represent Italian regions for both gender (the first 20 rows refer to male and the other 20 to female) and columns represent the elementary indicators.

In the selection of elementary indicators, we also included correlated dimensions. However, since gender inequality is a normative phenomenon, we decided to keep these correlations as they are relevant to the description of this issue (even if they are not statistically informative).

3.2 Normalization of elementary indicators

Since elementary indicators present values with different units of measurement, standardization was necessary to make them comparable. Therefore, the normalization makes sure that an increase in the indicators corresponds to an increase in the composite index, regardless of their polarity. Specifically, we applied the Adjusted Mazziotta-Pareto Index (AMPI) (Mazziotta and Pareto, 2016; Mazziotta and Preto, 2017) approach, which makes it possible to compare distributions with originally different variability over the time.

After the elementary indicators normalization, we compute a composite indicator for each single domain, and then we aggregate them in order to obtain the final composite index (Regional Gender Equality Index - RGEI).

Given the matrix $\mathbf{X}_d = \{x_{ijd}\}$, whose generic element x_{ijd} represents the value of the indicator j for the i-th region in the d-th domain, let *Inf* x_j and *Sup* x_j be respectively the overall minimum and maximum values of the indicator j across all the regions and all time periods considered. Denoting with *Ref* x_j the average of the indicator j all over the Italian regions in 2015 as the reference, the minimum and the maximum value of the possible range for each indicator (goalposts) are defined as:

$$\begin{cases} Min_{x_j} = Ref_{x_j} - \Delta \\ Max_{x_j} = Ref_{x_j} + \Delta \end{cases} \quad (1)$$

where $\Delta = (Sup_{xi} - Inf_{xi})/2$. Given the matrix $X_d = \{x_{ijd}\}$, a normalized matrix $R_d = \{r_{ijd}\}$ is then computed, where the generic element r_{ij} is obtained as follows (in the case of positive polarity of the indicator j):

$$r_{ij} = \frac{x_{ij} - Min_{x_j}}{Max_{x_j} - Min_{x_j}} \; 60 + 70 \quad (2)$$

where $Min\ x_j$ and $Max\ x_j$ are the goalposts for the indicator j. Indeed, if the indicator j has negative polarity, it is necessary to use the complement respect to 200 for computing the normalized value.

The multiplication factor 60 and the translation of 70 units allow to obtain vectors with values almost certainly between 70-130, where 100 corresponds to the reference value. One of the advantages of the AMPI method consists on an easier interpretation of the levels of the phenomenon. It is possible to notice immediately the values above the reference value (values greater than 100), the values with a level below it (values less than 100) and make comparison among them over the time.

3.3 Weighting and aggregation

Since we considered all the indicators equally important, we attributed the same weight to all of them. In order to obtain the composite index for the domain d-th, the elementary indicators for each region have been summarized through an arithmetic average, and for avoiding compensative side effects a penalty correlated to the variance of the indicators values has been applied to it:

$$AMPI\ d_i^{+/-} = Mr_i \pm Sr_i\ cv_i \tag{3}$$

where Mr_i and Sr_i, are respectively, mean and standard deviation of the normalized values of the region i-th, and $cvi = Sr_i/Mr_i$ is the coefficient of variation for the region i-th of the domain d-th. This aggregation method was applied to the four domains of each region, obtaining the gender equality index (RGEI) for each Italian region.

$$GEI = Md_i \pm Sd_i\ cvd_i \tag{4}$$

where Md_i is the arithmetic mean of the domains value of the region i-th, Sd_i their strander deviation and cvd_i represents the coefficient of variation. Since this is a positive polarity phenomenon, the penalty due to variability has been subtracted.

4. Results

The observed values of the RGEI show a marked disadvantage of female respect to male for each considered year (Table 3), confirming that there is a disparity between the two genders. Regarding the analysed domains, we can observe how males in almost all Italian regions have more opportunities and an easier access to social resources, with the exception of the Education domain where women show the highest values in every Italian region.

As far as the Education domain is concerned, we observe that Abruzzo and Molise females have the highest values over the entire period, while the lowest ones are found in the male population of Calabria, Sardegna and Sicilia.

Comparison over the years of the Work domain shows a marked difference between males index in the North regions, where Veneto, Trentino Alto Adige and Lombardia showed the best overall values, and females in the South, where Sicilia, Calabria and Campania displayed the lowest ones.

In the Power domain, given the complementarity of the gender data in the municipal councils and regional councillors, we observe how from 2015 to 2017 the extreme values between genders are shown in Calabria, Campania and Lazio with a marked penalty for females. Indeed, during this period, only two women held the role of President of the Region, specifically in Umbria and Friuli Venezia Giulia (Cottone, 2020).

During this period in the Security domain Veneto and Abruzzo displayed the highest level of safety for male, whilst Calabria and Friuli Venezia females showed the lowest one.

Considering the RGEI for the entire Italian peninsula, a slight but constant growth of the average national index has been recorded. Following ISTAT and Eurostat guidelines for the Italian territorial distinction, we analysed and compared the values of five macro-areas: North-west, North-east, Centre, South and Islands.

Table 3 – *Regional Gender Equality Index (RGEI) in the Italian regions for males and females.*

	RGEI	2015		2016		2017	
		M	**F**	**M**	**F**	**M**	**F**
North-West	PIEMONTE	108.52	99.28	108.70	99.69	108.75	100.03
	VALLE D'AOSTA	107.78	101.37	107.99	100.97	108.02	100.71
	LIGURIA	108.13	99.51	108.85	98.94	109.02	99.29
	LOMBARDIA	108.62	99.67	108.88	99.72	108.85	99.59
North-Est	TRENTINO A.A.	108.79	98.78	108.63	98.91	108.49	98.63
	VENETO	109.72	100.22	110.08	100.08	110.06	100.28
	FRIULI V.G.	109.37	99.25	109.46	99.95	109.00	100.07
	EMILIA ROMAGNA	108.73	100.56	109.13	100.69	109.00	100.78
Centre	TOSCANA	107.81	99.77	107.95	100.22	107.96	100.38
	UMBRIA	109.24	100.00	109.29	100.39	109.08	101.05
	MARCHE	108.94	101.12	109.07	101.43	109.19	101.36
	LAZIO	108.81	99.62	108.44	100.30	108.79	100.56
South	ABRUZZO	108.38	100.09	109.29	100.27	109.22	100.88
	MOLISE	108.55	100.12	107.34	99.59	108.31	100.96
	CAMPANIA	106.47	97.44	106.26	98.01	106.55	98.35
	PUGLIA	107.16	98.24	106.88	98.68	106.94	98.65
	BASILICATA	108.73	100.52	108.01	100.84	107.95	100.81
	CALABRIA	106.26	97.50	106.20	97.56	106.43	97.95
Islands	SICILIA	104.81	97.48	105.14	97.82	104.96	98.43
	SARDEGNA	105.36	97.68	105.87	98.05	105.76	98.18
	ITALY	100.00		100.17		100.26	

Table 3 clearly shows how in all Italian region males display the highest values compared to females of the same regions. More in detail we can observe that all males figures in each analysed year are higher than the national average RGEI with the maximum value reached by Veneto in 2016 (RGEI=110,08) and the minimum one by Sicilia in 2015 (RGEI=104,81). By contrast, the majority of females are below the goalspot ($RGEI_{(2015)}$= 100) with the maximum value represented by Marche in 2016 (REGEI= 101,43) and the minimum by Campania in 2015 (RGEI=97,44). The national average improved constantly gaining 0.17 in 2016 and 0.9 in 2017. This trend is also confirmed by the number of regions with a female RGEI over the goalspot moving from 8 in 2015, to 9 in 2016, reaching 11 in 2017.

For a straightforward and synthetic visualization of the phenomenon, we computed the mean values for macro areas (Table 4).

Table 4 - A*verage of Italian macro areas in 2015, 2016 and 2017*

RGEI	2015		2016		2017		Total	
	M	F	M	F	M	F	M	F
North-west	108.26	99.96	108.60	99.83	108.66	99.90	108.51	99.90
North-east	109.15	99.70	109.33	99.91	109.14	99.94	109.20	99.85
Centre	108.70	100.13	108.69	100.59	108.75	100.84	108.71	100.52
South	107.64	98.97	107.44	99.14	107.75	99.71	107.50	99.25
Islands	103.74	97.54	103.73	97.94	103.66	98.30	105.32	97.94

Accordingly to several studies on Italian society, Table 4 seems to represents an evident inequality level of inclusion between the North and the South of the peninsula (Di Bella, 2021).

Furthermore, the Islands present the lowest values for both genders with an average male RGEI of 105,32 and a female one of 97,94.

North-east instead shows the highest values for male (RGEI=109,19), with Veneto ranking as first all over the three years, while the Centre present the highest level of inclusion for females (RGEI= 100,58), with Marche placing second in 2015, and first in both 2016 and 2017.

5. Conclusions

During the realization of this research work we came across a systematic lack of data regarding gender issues, specifically gender disaggregated data or data composed through collection methods involving social and cultural factors. The RGEI may represent a solid and replicable measure of the phenomenon, with the aim to support the political and administrative decisions on both national and regional level of the country.From the analysis results we can clearly evince how gender gap is a common phenomenon all over Italy highlighting a significant difference between the northern and southern regions. Even if we noticed a slight improvement of the national average over the period, evidence of gender discriminations are still present in our daily society.

For example, although female numbers in the Instruction domain are higher than men in all the Italian regions, this performance is not reflected in the Work domain, where women are generally less paid and earn more part time contracts than men. This biased cultural mindset, based on the a very traditional model of household management, in which female role is relegated to the care of the family and the management of domestic activities, is a deterrence for the reintegration or the access on the labour market.The introduction of the parental leave in 2012 for fathers had a

mild, but significant contribution to reduce gender discriminations, however it impacted the labour system, and an extension of this paid leave may incentivize the reintegration of women on the labour market and the pursuit of their social, political and economic goals (Martino, 2018).

The Power domain also showed some marked inequalities of the presence of women in decision-making bodies. The application of 'gender equality by law' mechanism, as the quotas for women, brought an improvement in rebalancing the management bodies in the economic and political sectors.

The dimension of safety, sees acts of mistreatment and violence spread in a heterogeneous way throughout the peninsula. Even with the introduction of some national and international communication campaigns and the activation of the anti-stalking number (1522), the females are more likely to be victims of abuse than the male and feminicide seems still far to disappear from society.

We hope that a greater understanding of gender gaps, supported by analytical tools such as the one proposed in this article, may help to raise community awareness on an issue that, although it presents in the collective and political debate for several years now and some active policies were implemented in order to reduce the gap, still displays profound injustices and discriminations, in some cases even in its most violent forms.

References

BOZZANO M. 2012. Assessing Gender Inequality among Italian Regions: The Italian Gender Gap Index, *Quaderni di Dipartimento, No. 174*, Università degli Studi di Pavia, Dipartimento di Economia Politica e Metodi Quantitativi (EPMQ), Pavia.

COTTONE N. 2020. La difficile uguaglianza di genere, *Il Sole 24 ore*.

DI BELLA E., LEPORATTI L., GANDULLIA L., MAGGINO F. 2021. Proposing a regional gender equality index (R-GEI) with an application to Italy, *Regional Studies*, Vol. 55, No. 5, pp. 962-973.

HAUSMANN R. D., TYSON L., ZAHIDI S. 2007. The global gender gap index 2007, *The global gender gap report*, pp. 3-27.

ITALIAN CONSTITUTION 1948. Articoli 3-34-37-51.

KABEER N., NATALI L. 2013. Gender equality and economic growth: Is there a win-win?, *IDS Working Papers*, Vol. 417, pp. 1-58.

MARTINO E.M. 2018. Disuguaglianze di genere nel mercato del lavoro: l'impatto del Bonus Infanzia e del congedo di paternità. *Rivista delle Politiche Sociali / Italian Journal of Social Policy*, pp. 87-110.

MAZZIOTTA M., PARETO A. 2017. *Gli Indici Sintetici,* Torino, G.Giappichelli Editore.

MAZZIOTTA M., PARETO A. 2016. Methods for Constructing Composite Indices: One for All or All for One?, *Rivista Italiana Di Economia Demografia e Statistica*.

MOORHOUSE E. A. 2017. The many dimensions of gender equality and their impact on economic growth, *Forum for Social Economics*, Vol. 46, No. 4, pp. 350-370.

PROFETA P. 2017. Gender equality in decision-making positions: The efficiency gains, *Intereconomics*, Vol. 52, No. 1, pp. 34–37.

UNITED NATIONS 1997. *Report of the Economic and Social Council*, New York, United Nations.

UNITED NATIONS 2015. *Transforming our world: the 2030 Agenda for Sustainable Development, Resolution A/RES/70/1 of the General Assembly*, New York, United Nations.

UNITED NATIONS GENERAL ASSEMBLY 1948. Resolution 217A, *Universal Declaration of Human Rights,* Parigi, France, Art. 22-26.

SUMMARY

Gender equality in the Italian regions

This paper aims at analyzing the progressive evolution of the gender inclusion condition in the Italian regions over the period 2015-17, examining it across four domains: Education, Work, Power and Security with the goal of increasing knowledge about the topic of gender equality. The Adjusted Mazziotta-Pareto index (AMPI) has been applied to obtain a synthetic measure.

Nicola Maria D'ANGELO, dangelonicolamaria@yahoo.it
Tommaso PAOLETTI, tommasopaoletti@outlook.it

LONGEVITY-RISK-ADJUSTED GLOBAL AGE AS A MEASURE OF WELL-BEING

Mattia Mezzelani, Gloria Polinesi, Francesca Mariani, Maria Cristina Recchioni

1. Introduction

Survival analysis refers to a set of statistical techniques used to estimate the probability of the occurrence of a given event as a function of time (Collet, 2003). This collection of statistical procedures is also known as time-to-event analysis. It focuses on the study of the time elapsed from a specific moment up to the point when a specific event (death) will occur.
Models underlying survival analysis are usually expressed in terms of age-dependent mortality rate estimates.

Two types of approach emerge from the literature: a parametric approach that specifies a mathematical form for mortality rates (hazard rates) as a function of chronological age, and a non-parametric approach that sets the mortality rate of a given age as a weighted average of the 'raw' mortality rates.

According to the parametric approach, the seminal work of Gompertz (1825) implies that the total death rate is composed of an age-dependent term which increases exponentially with chronological age.

A first and non-exhaustive revised version of Makeham (1867) includes a non-age-dependent mortality rate in the Gompertz mortality law. In fact, the Gompertz–Makeham (GM) model has been expanded several times to allow for a specific shape of mortality. For example, Heligman and Pollard (1980) model mortality rate as an eight-parameter function of age, specifying the model in terms of mortality odds. Kostaki (1992) extends the Heligman-Pollard model to eliminate some biases in its estimation. Furthermore, Dellaportas *et al.* (2001) estimate the Heligman–Pollard model within a Bayesian framework.

Tabeau *et al.* (2002) review parametric functions used to model mortality and discuss the applicability of different parametric models to estimate the age and effects of mortality.

Recently, Milevsky (2020) developed a computational framework for inverting GM mortality hazard rates to present a new definition of age: the longevity-risk-adjusted global (L-RaG) age. The computation of longevity-risk-adjusted global age involves multifaceted aspects of aging by allowing the estimation of the number of years a person seems to be (biological age) from his/her chronological age. Indeed,

the computational process of L-Rag age begins by collecting data on mortality rates as a function of chronological age, as well as any other characteristics or elements associated with mortality.

On the other hand, non-parametric models used to estimate the survival function from lifetime data do not assume a theoretical distribution for *F*. This approach is inherent to studies of clinical trials where the following estimators assume particular importance: Kaplan-Maier estimator (Kaplan *et al.*, 1958), Nelson–Aalen estimator (Nelson, 1969), and longrank test (Peto and Peto, 1972).

We contribute to the existing literature by applying the above-mentioned work of Milevsky (2020) to obtain L-RaG age estimates among Italian regions for two different years (2011 and 2018), while proving that this quantitative indicator of biological age is related to a sentiment indicator of the perceived age (healthy life expectancy and one's life satisfaction) provided by the ISTAT database within the BES project[1]. The ISTAT indicators are obtained via surveys with the aim of detecting living conditions in Italian regions and reflecting human feeling. An interesting point is that an objective indicator (i.e., L-RaG age) can mimic a sentiment indicator, thus implying that human feeling relies on different lifestyle conditions. Analysis at regional level allows us to investigate the impact of the well-known difference between Northern and Southern Italy considering a biological indicator on different age classes.

In the following sections, we describe the GM mortality law and L-RaG age indicator, and we conclude with some results.

2. The GM mortality law

The GM mortality law is the linear relationship between the natural mortality rate and chronological age *x*, i.e., the number of years a person has been alive. Indeed, every adult life in the region *i*, for $\mathrm{i} = 1, \ldots, \mathrm{N}$ can be expressed by the following parametrization:

$$\mu_x[i] - \lambda[i] = \begin{cases} h[i]e^{g[i]x} & x < x^* \\ \lambda^* & x \geq x^* \end{cases}, \tag{1}$$

where $\mu_x[i] = 1/(1 - q_x[i])$ is the total hazard rate (THR) and $q_x[i]$ obtained from ISTAT mortality tables is the one-year decrement rate.

The region-specific parameters $\lambda[i],\ h[i], g[i]$ represent the accidental death rate, a.k.a. the Makeham constant, the initial natural mortality rate (INMR), and the

[1] Data collected from a sample of respondents that took a survey.

mortality growth rate (MGR), respectively. As well, x^*, i.e., the critical chronological age at which the Gompertzian regime ends, and $\lambda^* > \lambda[i]$, i.e., the plateau mortality rate faced when chronological ages are equal to x^*, are global parameters estimated at the global (Italian) level.

Parameters λ^* and x^* are also known as species-specific accidental mortality rate the species-specific lifespan (Richards, 2020).

Rearranging Eq. (1), the model can be expressed as:

$$ln(\mu_x[i] - \lambda[i]) = ln(h[i]) + g[i]x, \quad x < x^* \tag{2}$$

which is the standard linear representation of total hazard minus accidental death rates for all ages within the GM regime $[x, x^*]$.

The relationship can be explicitly written as:

$$\overbrace{ln\left(ln\left(\frac{1}{1-q_x[i]}\right) - \lambda[i]\right)}^{z} = \overbrace{ln(h[i]) + ln\left[(e^{g[i]x} - 1)/g[i]\right]}^{K_0} + \overbrace{g[i]}^{K_1} x \tag{3}$$

to estimate the region-specific GM parameters $\lambda[i], h[i], g[i]$, where the new constants K_0 and K_1 are defined for convenience, see Milevsky (2020). Global parameters x^* and λ^* are estimated via regression through the equation:

$$ln(h[i]) = L + (-x^*)g[i] + \varepsilon_j, \tag{4}$$

where $L = ln\lambda^*$.

The GM mortality law implies that (log) mortality rates increase linearly and then converge to a constant mortality plateau, which is also known as the compensation law. This leads to a linear negative relationship between the initial natural mortality (intercept) $ln(h[i])$ and MGR (slope) $g[i]$ as shown in Eq. (4).

The next Section focuses on the computation of the L-RaG age starting from the GM equation, using local and global mortality rates as input.

2.1 From GM mortality law to L-RaG age

According to the work of Milevsky (2020), *ξ(x,i)* denote the longevity-risk-adjusted global age for someone at the chronological age of *x* in region *i*, which may or may not correspond to his/her chronological age (x).

The L-RaG refers to the concept of biological age, i.e., the age (number of years) a person seems to be, whose computation involves mapping from mortality rates to a specific age by inverting the GM mortality law.

For this reason, the L-RaG age is forced to satisfy the compensation law described in Eq. (1), but considering the Italian average version of the GM parameters. Formally, we have:

$$ln[\Lambda_\xi[i] - \Lambda] = ln[H] + G\xi(x,i), \tag{5}$$

where $\Lambda_\xi[i]$ is the longevity-risk-adjusted global hazard rate and $\Lambda \geq 0, H > 0, G \geq 0$ represent the mean in Italy of $\lambda[i], h[i], g[i]$, respectively.

The longevity-risk-adjusted global hazard rate is set equal to the THR:

$$\Lambda_\xi[i] = \mu_x[i]. \tag{6}$$

Inverting the GM equation and solving by $\xi(x,i)$, the L-RaG age can be determined:

$$\xi(x,i) = \frac{ln[\lambda[i] - \Lambda + h[i]e^{g[i]x}] - ln[H]}{G}. \tag{7}$$

The validity of $\xi(x,i)$ is tied to $\lambda_i - \Lambda + h_i e^{g[i]x} > 0$. This fact is generally guaranteed because the GM model is applied to adult ages (35 and older).

Eq. (7) is used to estimate the L-RaG age in the Italian region. The data and results are discussed in Section 3.

3. Data and results

We consider data from ISTAT mortality tables for Italian regions considering for two years (2011 and 2018) males and females separately.

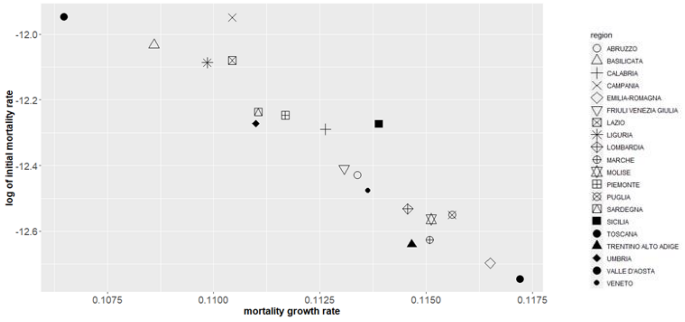
First of all, we extract the one-year decrement rate $q_x[i]$, then we estimate all the GM parameters as described in Section 2.

The compensation law is assured for all subgroups and years analyzed for values of x^* and λ^* equal to 95 and 0.06. The negative relationship between mortality growth rate and (log) initial mortality rate is shown in Figure 1.

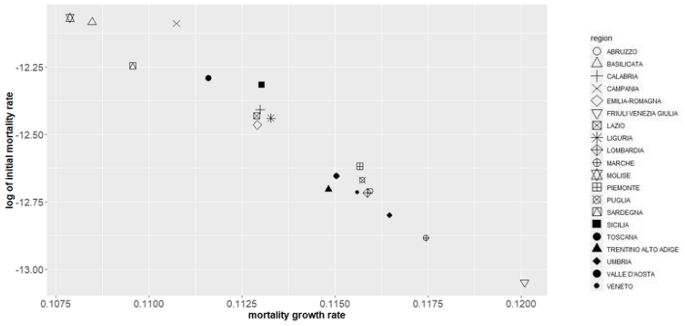
Figure 2 shows chronological ages versus the gap (difference) between chronological age and L-RaG age for females (a) and males (b) in each region considered.

Figura 1 – *Compensation law: females 2011-2018 (a, b) and males 2011-2018 (c, d).*

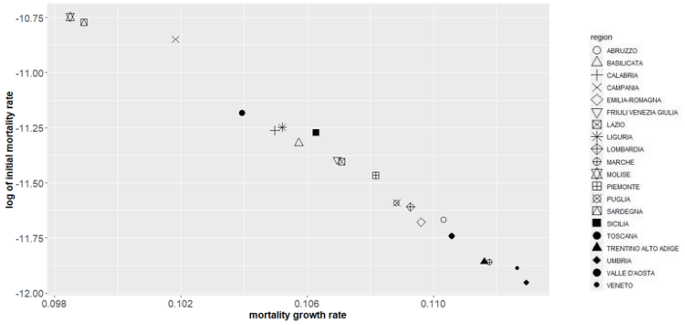
(a)



(b)



(c)



(d)

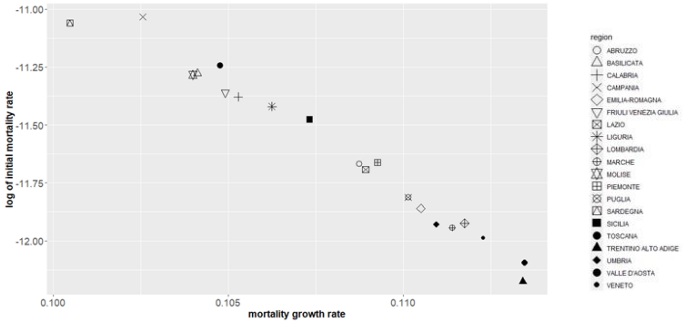
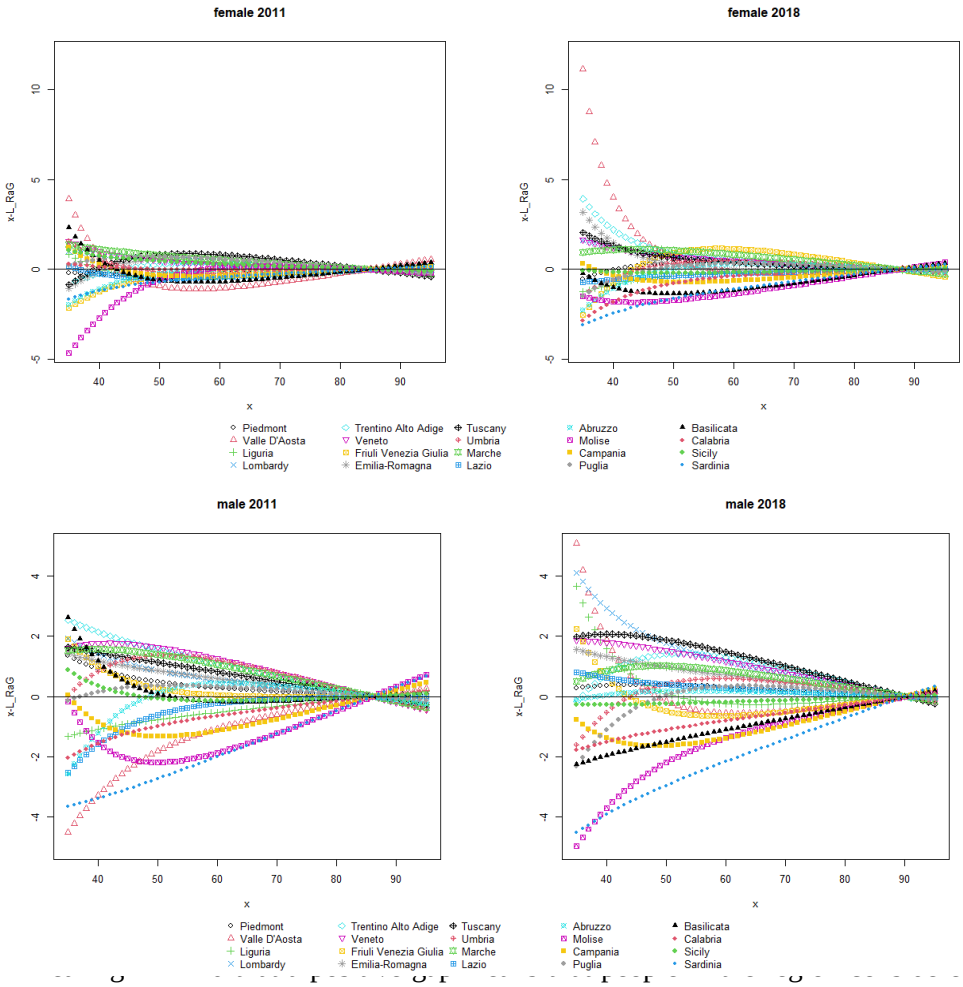
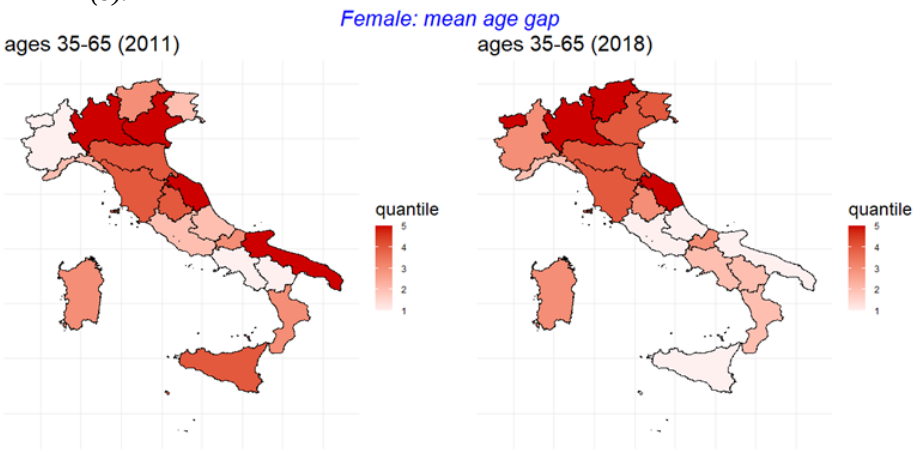


Figure 2 – *Relationship between chronological age (x) and the mean age gap (x-L-RaG) at regional level.*

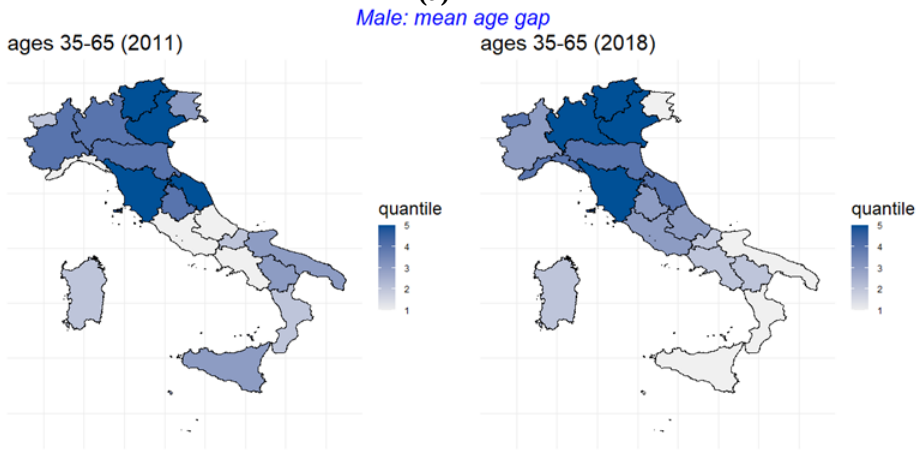


seem younger than their effective age, Figure 3 focuses on the mean age gap at the regional level, divided into quantiles relatived to the adult category (ages 35-65).

Figure 3 – *Mean age gap at regional level divided by quantiles for females (a) and males (b).*



(a)



(b)

Figure 3 shows how the situation worsened on average from 2011 to 2018 for adults (females and males) in the Southern regions (quantiles are lower) marking a distinction between Southern and Northern Italy in terms of the difference between chronological age and L-RaG age.

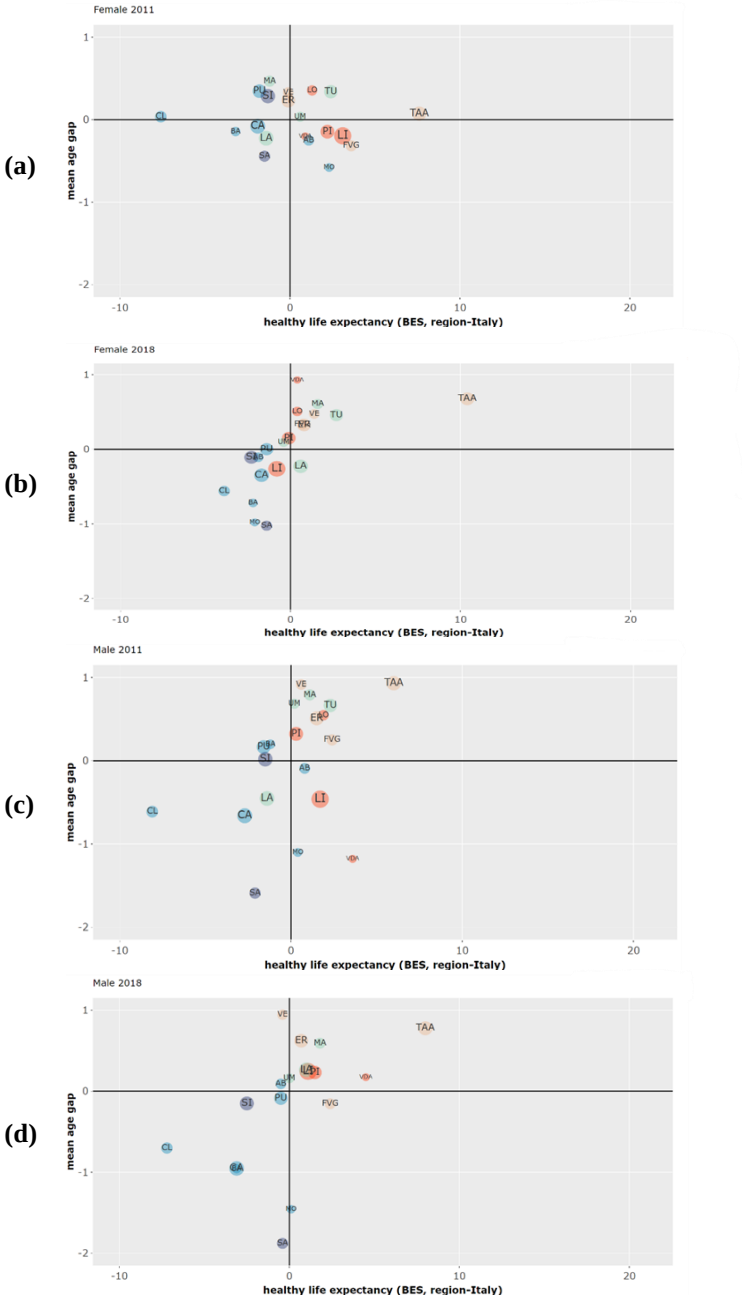
To investigate the age gap in terms of sentiment indicators, we relate the mean age gap (*x*-L-RaG) to healthy life expectancy and personal life satisfaction according to the BES database (ISTAT). We expect that, in the regions with positive health conditions (BES indicators at regional level greater than the Italian values) individuals should feel younger since they live better. Table 1 shows Pearson correlation coefficients between age gap and the indicators.

Table 1 – *Correlation coefficients between mean age gap and BES indicators. Signif. codes: p-value ≤ 0.001 (***); (**) 0.001<p-value≤ 0.01; (*) 0.01<p-value≤0.05; (·) 0.05<p-value ≤ 0.1;(-) 0.1<p-value≤1.*

	Female		Male	
	2011	2018	2011	2018
Healthy life expectancy	-0.13 (-)	0.63 (**)	0.39 (.)	0.51 (*)
Personal life satisfaction	-0.07 (-)	0.59 (**)	0.23 (-)	0.50 (*)

L-RaG is related to sentiment indicators and especially to healthy life expectancy (higher and significant correlations). For this reason, we report only graphs showing regional mean age gaps associated with the difference between the values of healthy life indicator at the regional level and the Italian mean (Figure 4). Node colors refer to the pertinent macro-area: north-west (orange), north-east (light orange), centre (green), south (light blue) and islands (violet).

Figure 4 – *Sentiment indicator and mean age gap at regional level.*



According to Figure 4, regions belonging to the first quadrant represent the group associated with the best living conditions in terms of both perceived age and future perspective of healthy life expectancy. In contrast, regions in the third quadrant represent the group associated with the worst living conditions.

The situation for females changed from 2011 to 2018. In fact, the most recent year is characterized by a clear separation of regions into the first and third quadrants. For males, the two years seem more similar.

4. Conclusion

Mortality is a complex phenomenon affected by many factors such as wellness and healthy lifestyle, geographical area, or sex. However, almost all parametric mortality models assume the mortality rate to be function only of chronological age, neglecting the influence of qualitative, non-measurable factors. Based on these considerations, we hypothesize a relationship between mortality and well-being.

Assuming the mortality described by a GM model, we compute the L-RaG age for males and females among Italian regions for the years 2011 and 2018 and the gap between L-RaG and chronological age. Our empirical findings show that the age gap, especially for females, is highly correlated with a healthy life expectancy, supporting our assumption that sentiment indicators based on surveys may be reflected in indicators based on observed data. Moreover, a joint analysis of the age gap and healthy life expectancy allows Italian regions to be ranked based on health conditions.

References

COLLET D. 2003. *Modelling survival data in medical research*. Florida: Chapman & Hall.

DELLAPORTAS P., SMITH A. F., STAVROPOULOS P. 2001. Bayesian analysis of mortality data, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 164, No. 2, pp. 275-291.

GOMPERTZ B. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies, *Philosophical transactions of the Royal Society of London*, Vol. 115, pp. 513–583.

HELIGMAN L., POLLARD J. H. 1980. The age pattern of mortality, *Journal of the Institute of Actuaries*, Vol. 107, No.1, pp. 49-80.

KAPLAN E. L., MEIER P. 1958. Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, Vol. 53, No.282, pp. 457-481.

KOSTAKI A. 1992. A nine-parameter version of the Heligman-Pollard formula, *Mathematical Population Studies*, Vol. 3, No.4, pp. 277-288.

MAKEHAM W. M. 1867. On the law of mortality, *Journal of the Institute of Actuaries*, Vol. 13, pp. 325-358.

MILEVSKY M. A. 2020. Calibrating Gompertz in reverse: What is your longevity-risk-adjusted global age?, *Insurance: Mathematics and Economics*, Vol. 92, pp. 147-161.

NELSON W. 1969. Hazard plotting for incomplete failure data, *Journal of Quality Technology*, Vol.1, No. 1, pp. 27-52.

PETO R., PETO J. 1972. Asymptotically efficient rank invariant test procedures, *Journal of the Royal Statistical Society: Series A (General)*, Vol. 135, No.2, pp.185-198.

RICHARDS S. J. 2020. A Hermite-spline model of post-retirement mortality, *Scandinavian Actuarial Journal*, Vol. 2020, No.2, pp. 110-127.

TABEAU E., WILLEKENS F., VAN POPPEL F. 2002. Parameterisation as a tool in analysing age, period and cohort effects on mortality: a case study of the Netherlands. *In The Life Table*. Dordrecht: Springer.

SUMMARY

Longevity-risk-adjusted global age as a measure of well-being

In this work, we show how the difference between chronological and biological age can be investigated by sentiment indicators constructed using survey data (healthy life expectancy and personal life satisfaction). We invert the GM mortality model to compute longevity-risk-adjusted global age, detecting how an objective indicator (i.e., L-RaG age) mimics a sentiment indicator, thereby implying that human feeling relies on different lifestyle conditions.

Mattia MEZZELANI, Università Politecnica delle Marche, S1086510@studenti.univpm.it
Gloria POLINESI, Università Politecnica delle Marche, g.polinesi@univpm.it
Francesca MARIANI, Università Politecnica delle Marche, f.mariani@univpm.it
Maria Cristina RECCHIONI, Università Politecnica delle Marche, m.c.recchioni@univpm.it

INCIDENCE, INTENSITY AND INEQUALITY OF POVERTY IN ITALY

Mariateresa Ciommi, Chiara Gigliarano, Francesco M. Chelli

1. Introduction

In the economic literature there is a large consensus on measuring poverty according to three specific aspects: incidence, intensity and inequality among the poor (see Sen, 1976). Looking at more than one single index may allow for useful insights in the poverty analysis as well as for better tailored policies aimed at alleviating poverty.

Policy makers are interested in having efficient and unambiguous measures that could better address local policies for eradicating poverty, especially when resources are scarce.

For these reasons, we decide to investigate poverty at subnational level by means of robust instruments, based on decomposable indices and dominance criteria.

For the analysis, we use IT-SILC data for the years 2005 and 2015 to capture the evolution of poverty before and after the 2007-2009 economic crisis.

For the decomposition analysis, we follow Aristondo and Onadia (2018) approach, which is based on the Shapley decomposition (Shapley, 1953). For the two years of analysis, we compute the second member of the family of Foster-Greer-Thorbecke poverty indices (with parameter alpha equal to 2). Then, by applying the Shapley decomposition, we can establish whether changes in the value of the Foster-Greer-Thorbecke poverty measure are due to variations in the size of poor people (incidence of poverty) rather than in their poverty gaps (intensity) or in the inequality among the poor.

Moreover, we integrate the analysis by adding a robust dominance criterion, based on the so-called TIP curves ("Three I's of poverty") introduced by Jenkins and Lambert (1997). TIP curves provide a graphical representation that summarizes the three aspects of poverty in just a picture, allowing for comparison over time and across countries or regions. We trace the evolution of poverty in Italy and in its macro-regions (North, Center and South) between 2005 and 2015, by comparing the resulting poverty curves.

The rest of the paper is organized as follows. Section 2 presents the methodology adopted, describing the decomposition of the poverty indices as well as the poverty

curves. Section 3 describes the data used and illustrates the main results with a discussion of policy implications. Finally, Section 4 concludes and points the way for future research.

2. Methodology

The proportion of households or individuals having an income below a fixed poverty line (threshold) is the most used way to measure income poverty. The economic literature refers to this proportion as the *Headcount Ratio* (*H*) or poverty incidence. By denoting with $y = (y_1, y_2, \dots, y_n) \in R^n_{++}$ the vector of incomes of a population of size *n* (*n*≥*2*), with $z \in R_{++}$ the fixed poverty line and with $q = q(y; z)$ the number of units (households or individuals) whose income falls below the poverty line, then the index *H* can be written as:

$$H = H(y; z) = \frac{q}{n}. \quad (1)$$

The main advantage of H is that it is very easy to understand and interpret. However, it also presents some drawbacks. First, H does not capture how poor the poor are (poverty intensity). Second, it ignores the shape of the income distribution among the poor, not considering the inequality among the poor. Consequently, two countries could have the same proportion of poor but exhibit different inequality among the poor group.

Thus, to measure properly income poverty, it is compulsory to introduce indices that take into account the distributional aspects of poverty (Sen, 1976; Shorrocks, 1995). In other words, it is important to have measures that reflect three different but complementary aspects of poverty: the intensity, the incidence and the inequality among the poor.

The *Income Gap Ratio (IGR)* is the simplest measure that accounts for poverty intensity, with the main advantage that it gives more emphasis (weight) to the poorest individuals. *IGR* is defined as the mean of the relative gaps from the poverty line among the poor:

$$IGR = IGR(y; z) = \frac{1}{q}\sum_{i=1}^{q} g_i, \quad (2)$$

where $g_i = \max\{\frac{z-y_i}{z}, 0\}$ denotes the relative poverty gap.

To account for inequality, we use the *Coefficient of Variation (CV)*, defined as:

$$CV_p = CV_p(y; z) = \frac{\sqrt{\frac{1}{q}\sum_{i=1}^{q}(y_i - \mu_p)^2}}{\mu_p}, \quad (3)$$

where μ_p denotes the mean income among the poor.

Foster, Greer and Thorbecke (1984) propose a family of poverty indices that includes, as special cases, measures of incidence, intensity and inequality. The FGT family depends on a parameter α that reflects poverty aversion:

$$FGT_\alpha = FGT(y; z) = \frac{1}{n}\sum_{i=1}^{q}(\frac{z - y_i}{z})^\alpha. \quad (4)$$

Assuming α=0, we get $FGT_0 = H$. For α=1, FGT_α reduces to $FGT_1 = H \cdot IGR$, whereas fixing α=2 we get $FGT_2 = \frac{1}{n}\sum_{i=1}^{q}(g_i)^2$. The index FGT_2 can be written in terms of incidence, intensity and inequality, as follows:

$$FGT_2 = \frac{1}{n}\sum_{i=1}^{q}(g_i)^2 = H[IGR^2 + (1 - IGR)^2 CV^2]. \quad (5)$$

Aristondo and Onadia (2018) propose a new decomposition of FGT_2, based on the Shapley method (Shapley, 1953). Shapley decomposition is a method extended from game theory to applied economics. It decomposes the overall poverty change between two periods in terms of the percentage changes due to incidence, intensity and inequality. Shapley decomposition approach consists in evaluating the impact of each determinant by eliminating sequentially each of the contributory factors and computing the corresponding marginal change in the statistic. In other words, it allows to estimate the marginal contribution of each determinant to the overall value. In this paper, we want to understand the contribution to a variation in the FGT_2 from 2005 to 2015 that is due to a variation in the intensity, in the incidence or in the inequality among the poor. Thus, following Aristondo and Onadia (2018)[1] the overall poverty change in FGT_2 can be written as an aggregative function of the three determinants, H, IGR and CV, as follows:

$$f(H_c, IGR_c, CV_c,) = FGT_2(H_1, IGR_1, CV_1) - FGT_2(H_0, IGR_0, CV_0)$$
$$= c(H_c) + c(IGR_c) + c(CV_c) \quad (6)$$

[1] See their Proposition 1 and Proposition 2.

where $H_c = H_1 - H_0$, $IRG_c = IGR_1 - IGR_0, CV_c = CV_1 - CV_0$. Moreover, the components $c(H_c), c(IRG_c)$ and $c(CV_c)$ denote the contribution of the three determinants and are given by:

$$c(H_c) = -\frac{1}{6}(H_0 - H_1)[CV_0^2(2IGR_0^2 - 4IGR_0 + IGR_1^2 - 2IGR_1 + 3) + CV_1^2(IGR_0^2 - 2IGR_0 + 2IGR_1^2 - 4IGR_1 + 3) + 3(IGR_0^2 + IGR_1^2)]; \quad (7)$$

$$c(IGR_c) = -\frac{1}{6}(IGR_0 - IGR_1)[CV_0^2(2H_0 + H_1)(IGR_0 + IGR_1 - 2) + CV_1^2(H_0 + 2H_1)(IGR_0 + IGR_1 - 2) + 3(H_0 + H_1)(IGR_0 + IGR_1)]; \quad (8)$$

$$c(CV_c) = -\frac{1}{6}(CV_0^2 - CV_1^2)[H_0(2IGR_0^2 - 4IGR_0 + IGR_1^2 - 2IGR_1 + 3) + H_1(IGR_0^2 - 2IGR_0 + 2IGR_1^2 - 4IGR_1 + 3)]. \quad (9)$$

Jenkins and Lambert (1997) propose a graphical representation to summarize these aspects of poverty through the "three I's of poverty" (TIP) curves.

TIP curves are cumulative poverty gap curves that plot the cumulated proportion of population (x-axis) versus the cumulated normalized poverty gap among the poor (y-axis):

$$TIP(p,z) = \int_0^{F^{-1}(p)} \left(1 - \frac{y}{z}\right) \mathbf{1}(y \le z) f(y) dy, \quad (10)$$

where $f(y)$ denotes the income density function, $F^{-1}(p)$ is the quantile function and p the proportion of individuals, $p \in [0,1]$. Figure 1 reports an example of TIP curve.

To construct the curve, gaps are ordered from largest to smallest. For values of p (horizontal axis) greater than the poverty incidence, the TIP curve becomes horizontal. At this point the x-axis value corresponds to the incidence of poverty, the y-axis value indicates the poverty intensity, while the curvature indicates the degree of inequality among the poor. If the curve is a straight line, it means that all the poor are equally poor. The more the TIP curve deviates from linearity, the greater is the degree of inequality among the poor.

Jenkins and Lambert (1997) derive also a dominance criterion based on the TIP curves, showing that it is equivalent to a restricted second-order stochastic dominance. Income distribution A TIP dominates income distribution B for a given poverty line z if TIP_A (p, z) $\le$ TIP_B (p, z) for all p $\in$ [0; 1] and < for at least one p. It means that in income distribution B there will be more poverty than in A.

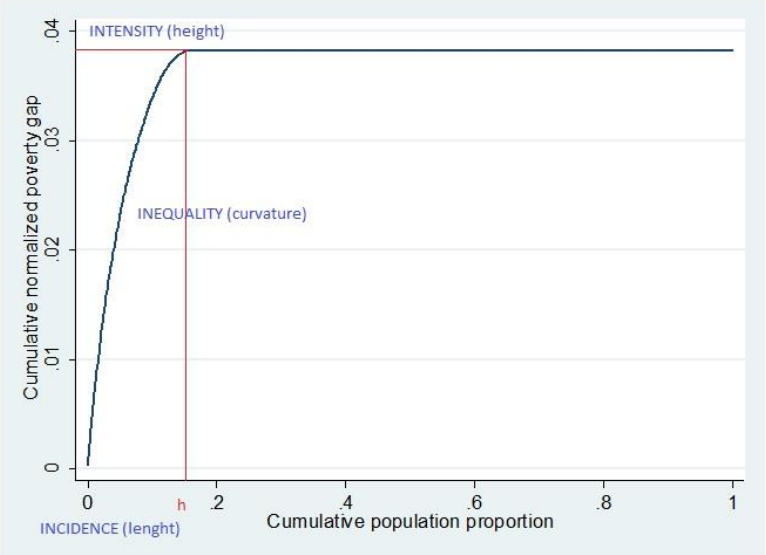
Poverty ranking provided by TIP dominance is robust to the choice of poverty line and to a set of poverty measures. If TIP_A (p, z) is always below TIP_B (p, z) for a

common z, then the ordering is preserved for any common poverty line smaller or equal to z. This implies the ordering of poverty indices similar to FGT1 and FGT2 (intensity and inequality).

Finally, analogous to the Lorenz curves, when TIP curves intersect, no TIP dominance can be assessed.

Statistical inference about TIP dominance have implemented by Berihuete et al. (2018) following the asymptotically distribution-free statistical procedure in Xu and Osberg (1998).

Figure 1 – *TIP curve: an example.*

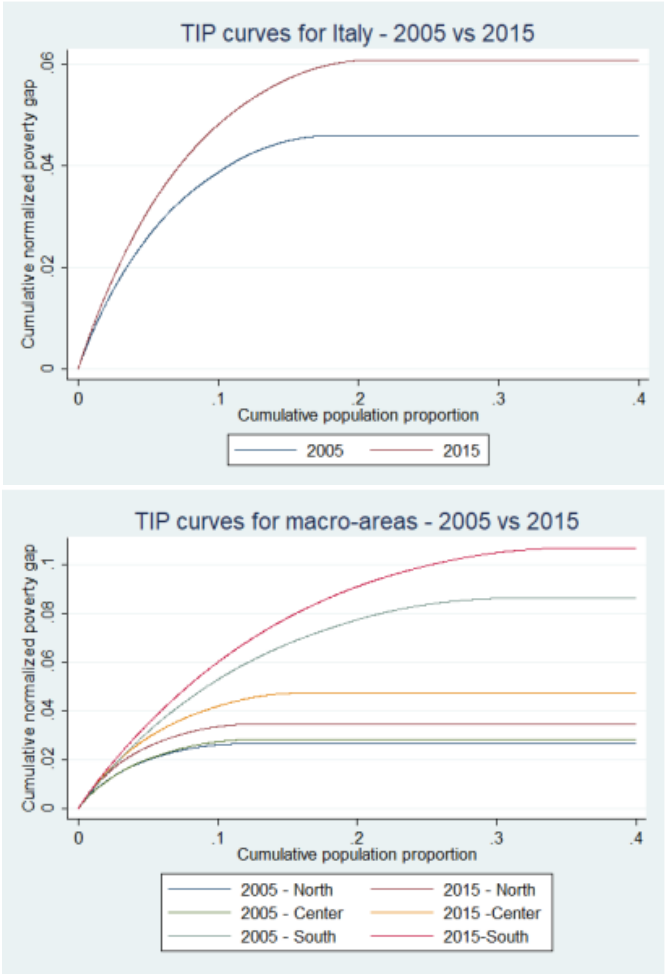


Source: our elaboration.

3. Data and results from empirical application

We use data from the Italian version of European Union Survey on Income and Living Conditions (IT-SILC) referred to the years 2005 and 2015. The variable of interest is the household equivalized disposable income. We compute poverty at household level, considering a relative poverty line equal to the 60% of the median equivalized household income in 2005. Incomes of 2015 have been deflated, so that we keep a fixed poverty line for the two periods of interest. Calculations consider cross sectional sample weights.

Figure 2 – *TIP curves for Italy and macro-regions in 2005 and 2015*



Source our elaboration on IT-SILC data

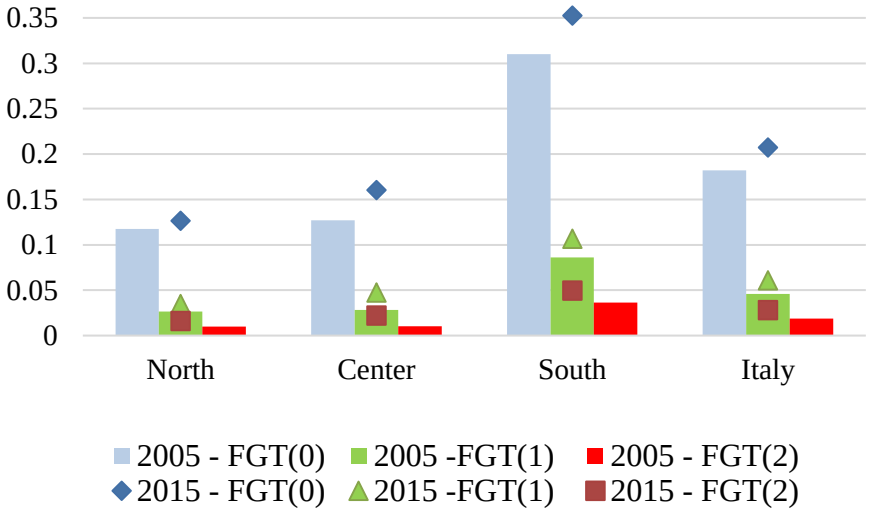
Figure 2 depicts the TIP curves for Italy and its macro-regions (North, Center and South) in the two periods of the analysis.[2]

For Italy, the TIP curve of 2015 dominates the one of 2005, meaning that poverty in 2015 is higher than in 2005. From 2005 to 2015, the incidence coordinate (on

[2] North area includes the following regions: Piedmont, Aosta Valley, Lombardy, Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia and Liguria; the Center includes Emilia-Romagna, Tuscany, Umbria, Marche and Lazio; while regions Abruzzo, Molise, Campania, Apulia, Basilicata, Calabria, Sicily and Sardinia belong to the South.

horizontal axis) has moved to the right and the intensity coordinate (on vertical axis) on the top. Thus, we can conclude that the 2015 the poverty situation is worse than in 2005. The conclusion is analogous for any poverty line smaller than the one chosen.

Figure 3 – *FGT_α for α=0, α=1 and α=2, in 2005 and 2015.*



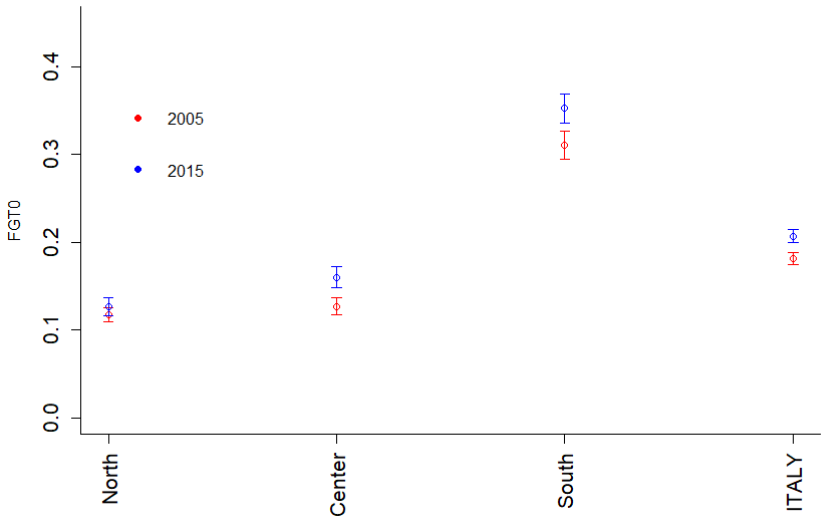
Source: Our elaboration on IT-SILC data.

By looking at the results for the macro-areas (Figure 2 – bottom panel), the first two curves on the top represent the TIP curves for South of Italy in the years 2015 and 2005, respectively. The third curve from above represents the Center in 2015, followed by the North in 2015. The lowest two curves are those of the Center and the North in 2005. It emerges that in all the three macro-areas the situation has worsened over the period of consideration. The level of poverty in the South of Italy remains very far from the rest of the country, while the gap between North and Center increases remarkably.

We now focus on specific indices of poverty, which belong to the FGT_α family, with three values of α, namely α=0, 1 2. Figure 3 compares the values of these indices for Italy and its macro-areas.

Looking at Figure 3, what emerges is that, for all the indices, the poverty values registered in the South are the highest. Moreover, if we compare the results over time, the situation worsened, confirming what already displayed through the TIP curves.

Figure 4 – *95% Confidence Intervals for FGT_0.*



Source: Our elaboration on IT-SILC data.

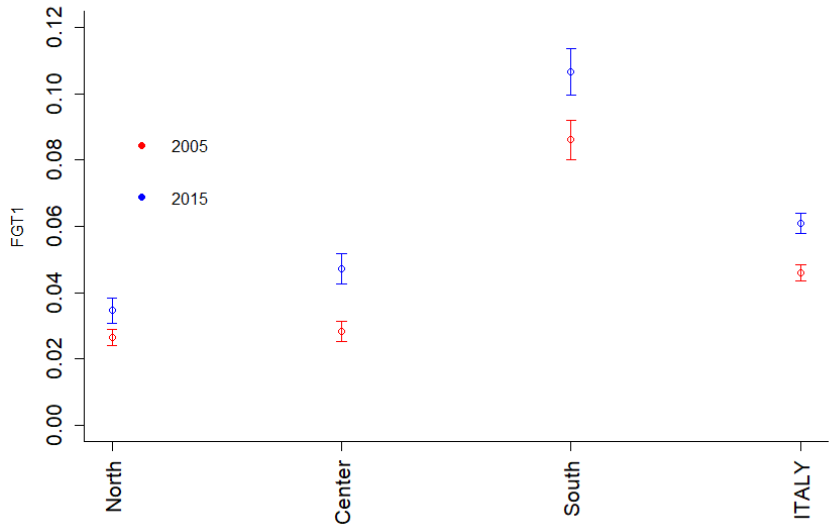
We also compute 95% confidence intervals, to test if the differences are significant. Figure 4, Figure 5 and Figure 6 summarize the results. The confidence intervals overlap only for the Headcount ratio (FGT_0) for North of Italy. For the remaining indices, the increase in FGT indices from 2005 to 2015 is significant for the three values of the parameter α and for all macro-regions and Italy. Looking at FGT_1, the greatest increase is registered in the Center. The same happens for FGT_2

Finally, we compute the Shapley decomposition, according to equations (6), (7), (8) and (9). Table 1 reports the results of the decomposition.

The smallest variation in poverty (last column of Table 1) is registered in the North (0.0058) whereas Center and South display similar results (0.0116 and 0.0128, respectively). All the contributions have a positive sign, confirming that all the aspects concerting poverty - inequality, incidence and intensity - have increased over the period of interest. For Italy, the three components have almost the same weight (31.67%, 35.48% and 32.85%, respectively). The Center behaves in a similar way. North and South show a different picture: in the North the greatest impact in poverty variation is due to the inequality among the poor (44.06%), whereas for South the highest impact is for the poverty incidence (42.32%). Therefore, the Shapley decomposition constitutes an important source of information, which can help policy-maker understand in which direction poverty-oriented policies should be

aimed, whether focused to reduce the number of poor or rather to reduce their inequalities.

Figure 5 – *95% Confidence Intervals for FGT_1.*



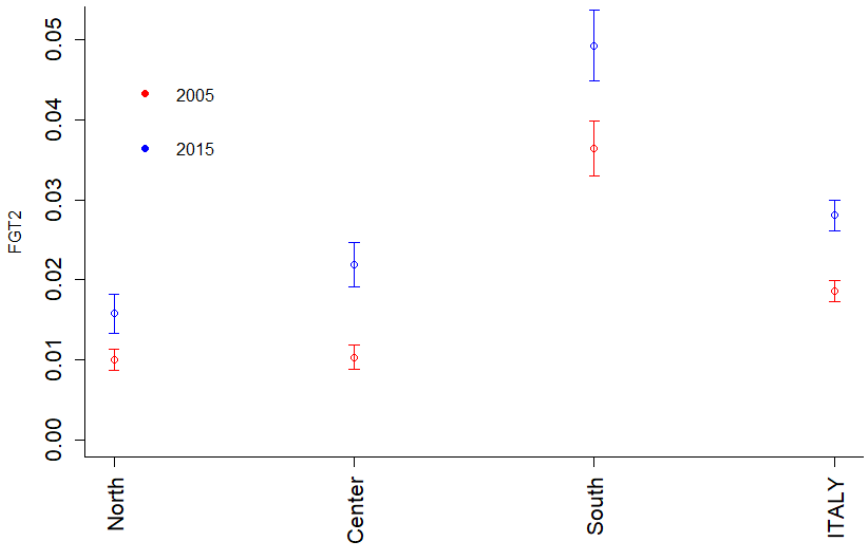
Source: Our elaboration on IT-SILC data.

Table 1 – *Shapley decomposition for the variation of FGT_2.*

	$c(H_c)$	%	$c(IRV_c)$	%	$c(CV_c)$	%	*f(H, IRV, CV)*
North	0.0009	16.17%	0.0023	39.78%	0.0026	44.06%	0.0058
Center	0.0036	31.54%	0.0042	36.38%	0.0037	32.08%	0.0116
South	0.0054	42.32%	0.0038	29.32%	0.0036	28.36%	0.0128
Italy	0.0030	31.67%	0.0034	35.48%	0.0031	32.85%	0.0095

Source: Our elaboration on IT-SILC data.

Figure 6 – *95% Confidence Intervals for FGT_2.*



Source: Our elaboration on IT-SILC data.

4. Concluding remarks

The TIP curves, in combination with the Shapley decomposition, allow to highlight different aspects of poverty that the use of a single poverty index could hide. We believe that the conjunction of these different approaches could be a useful instrument in a period of tight resources. Moreover, the support given by the dominance criterion and inference (using confidence intervals) could add more robustness to the analysis.

There is room for improving the work in several directions. First by considering the local dimension at a deeper degree of granularity: we are planning to extend the results at regional level. Second, we are planning to replicate the analysis to more recent data with the aim of capturing the evolution in poverty intensity, incidence and inequality and help policymakers in identifying the aspects that affect the most the level of poverty. Finally, we are interested in comparing different inferential approaches to TIP curves (see, e.g., Barrett et al., 2016) and in developing inferential results to the Shapley decomposition.

References

ARISTONDO O., ONADIA E. 2020. On measuring the sources of changes in poverty using the Shapley method. An application to Europe, *Fuzzy Sets and Systems*, Vol. 383, pp. 80-91.

BARRETT G. F., DONALD S. G., HSU Y.-C. 2016. Consistent tests for poverty dominance relations. *Journal of Econometrics*, Vol. 191, pp. 360-373.

BERIHUETE A., RAMOS C., SORDO M. 2018. Welfare, inequality and poverty analysis with rtip: An approach based on stochastic dominance, *The R Journal*, Vol 10, pp. 328-341.

FOSTER J., GREER J., THORBECKE E. 1984. A class of decomposable poverty measures, *Econometrica*, Vol. 52, No. 3, pp. 761-766.

JENKINS S. P., LAMBERT P. J. 1997. Three I's of poverty curves, with an analysis of UK poverty trends, *Oxford economic papers*, Vol 49, pp. 317-327.

SEN A. 1976. Poverty: an ordinal approach to measurement, *Econometrica*, Vol. 44, pp. 219-231.

SHAPLEY L. 1953. A value for n-person games. In: Kuhn, H., Tucker, A. (Eds.), *Contributions to the Theory of Games*. Vol. II. Princeton University Press, Princeton, pp. 307-317.

SHORROCKS A. F. 1995. Revisiting the Sen poverty index, *Econometrica*, Vol. 63, No. 5, pp. 1225-1230.

XU K., OSBERG L. 1998. A distribution-free test for deprivation dominance. *Econometric Reviews*, Vol. 17, pp. 415-429.

SUMMARY

Incidence, Intensity and Inequality of Poverty in Italy

Aim of the paper is to analyse unidimensional poverty in Italy and in its macro-regions. In a period of tight resources, the sub-national dimension in measuring poverty is crucial. Indeed, information on household income distribution and poverty at sub-national level may help policymakers focus their efforts and enhance the effectiveness of public interventions. Moreover, high disparities between macro-regions in a given country might undermine national economic growth and lead to ever-increasing regional imbalances over time.

To achieve our aim, we use data from the Italian version of the Statistics on Income and Living Conditions (IT-SILC) for two different years, 2005 and 2015.

The so-called TIP curves, a statistical tool for representing the three different aspects of poverty - incidence, intensity and inequality, provides poverty orderings consistent with a large class of poverty indices and of poverty thresholds.

Finally, we also decompose the variation of poverty index over time to better understand what are the main factors that influence poverty levels.

The main conclusion from the empirical application is an unambiguous increase in poverty levels from 2005 to 2015, both in the entire Italian population as well as its macro-regions.

Mariateresa CIOMMI, Università Politecnica delle Marche, m.ciommi@univpm.it
Chiara GIGLIARANO, Università dell'Insubria, chiara.gigliarano@uninsubria.it
Francesco M. CHELLI, Università Politecnica delle Marche & ISTAT, f.chelli@univpm.it

THE INFLUENCE OF BES TERRITORIAL INDICATORS ON ECONOMIC PERFORMANCE OF MANUFACTURING FIRMS[1]

Romana Gargano, Ferdinando Ofria

1. Introduction

Labour productivity is amongst the most important and influential variables governing economic production activities. In the economic literature, numerous studies have estimated the determinants of productivity growth with an ever-increasing interest both for reasons of equity and social cohesion, due to the existence of large and persistent regional productivity gaps. The argument that productivity differences can be mainly attributed to the structural transformation process (the reallocation of work between sectors) has been empirically demonstrated in several studies (Roncolato and Kucera, 2014). Increasingly, the political objective of reducing disparities by increasing the competitiveness of the least productive regions has been seen as the solution to increase productivity levels and bridge the gap between competing territories. Focusing on productivity within the sector, several studies underline the importance of factors with supply-side implications, such as social and human capital, labour costs, R&D and personal security, (Millemaci and Ofria, 2016).

The study aims to verify whether manufacturing firms (in terms of regional labour productivity) are influenced by socio-environmental context, in addition to some well-known managerial factors (such as "investments per employee", "innovation, research and creativity" and "share of exports") which, in previous studies, have highlighted a significant role in productivity discrepancies between Italian regions. The socio-environmental factors considered are homicide rate (indicator of security), minimum economic conditions (indicator of economic well-being) and the indicator of innovation, research, and creativity. These indicators belong to the 12 dimensions of equitable and sustainable well-being (BES) developed by National Institute of Statistics (ISTAT).

Much research (Ofria and David, 2014; D'Agostino and Scarlato, 2015; Bristow and Healy, 2018; Alesina et al., 2019) has shown that the presence of negative

[1] This article was conceived and prepared by all the authors; however, Romana Gargano is the author of paragraphs 2 and 3, Ferdinando Ofria wrote paragraphs 1 and 4.

externalities, related to the social and institutional variables, impact on the innovative capacity of regional economic systems. Phenomena, such as criminal organizations in areas where they are strongly rooted (Calabria, Campania, Puglia and Sicily) impedes the development of a social fabric founded on trust and sharing (Acemoglu et al. 2020). Their illegal power has spread widely in these societies and has greatly influenced the legal economy, for instance through the phenomenon of corruption, especially in those economic sectors where the government is directly or indirectly involved (Nese and Troisi, 2019; Ofria and Mucciardi, 2021). Organized crime directly produces goods and services for the following reasons: 1) money laundering; 2) territorial control (by social and political consensus), managing labour market shares in labour-intensive sectors (construction, retail, transport and services for families and businesses). Furthermore, the criminal organisation imposes protection rackets and other illegal payments on local firms (Centorrino and Ofria, 2008)

The analysis is based on ISTAT time series data 2012-2016 relating to regional labour productivity, some well-known management factors, and BES indicators. A quantile regression model allowed us to verify the differences in the effects exerted on productivity by independent variables at various quantiles, to identify that labour productivity is heterogeneous and that the relationship between labour productivity, socio-environmental context and managerial characteristics is not constant between quantiles. The empirical results provided by our analysis support the theoretical thesis that the higher the level of uncertainty due to environmental factors, the lower the labour productivity.

The paper is structured as follows: section 2 introduces the data and the methodology adopted; section 3 presents the findings. Finally, Section 4 concludes and discusses the results in the light of some considerations.

2. Data and Methodology

2.1. Data

This paper investigates Italian labour productivity across Nomenclature of Units for territorial statistics level 2 (NUTS 2) to 2012-2016. The data used are part of the "Report on the competitiveness of the productive sectors", produced by ISTAT. For our purpose, the report provides information on labour productivity, investments per employee, share of exports (impact of the sector on the region's total manufacturing exports). Labour productivity is measured as a log of ratio of the value added (output) by number of employees (see e.g., Ahlawat and Renu, 2018 and Mundakkad, 2018).

Investments per employee is a proxy for efficiency investments (Sylos-Labini, 2004), that is, the innovative investments made in response to the growth of the

relative cost of labour. For Kaldor (1967), new investments represent endogenous technical progress. The share of exports is a proxy for competitiveness. Companies that export are stimulated to increase productivity to be competitive. This incentive to be competitive drives larger companies to invest in R&D (Castellani et al., 2017).

BES indicators are useful for assessing the social and environmental progress of society. In this paper, we have used only a few BES indicators capable of providing information on economic well-being, safety and research of a given territory. Minimum economic conditions (MCE) are a composite indicator of economic well-being obtained by summarizing four indicators relating to the condition of serious material deprivation, quality of the home, economic difficulty in making ends meet and very low family work intensity. An increase in the MCE index indicates a reduction in the condition of discomfort. Innovation, research, and creativity (IRS) represents the domain with the same name and considers 3 elementary indicators: research intensity, knowledge workers and employees in creative enterprises. In the composition of this indicator, the indices that best capture social and economic progress were preferred. Homicide rate is one of the indicators that represents the domain of security. It has been standardized in such a way that its dynamics are in line with that of safety. A decrease in homicides corresponds to an increase in the standardized rate and therefore an increase in security, and vice versa. The choice of including this indicator is due to the observation that the safety of citizens is a key dimension in the construction of individual and collective well-being. The sense of insecurity of the population and the fear of being the victim of criminal acts can greatly influence the personal freedoms of each person, the quality of life and the development of the territories. The composite indices calculated for each dimension were obtained by applying Adjusted Mazziotta-Pareto Index (AMPI). It is a partially non-compensatory composite indicator based on a standardization of the individual indicators, at the reference time, that allows comparability of the data across units and over time (Mazziotta and Pareto, 2016).

2.2. Methodology

Quantile regression (Buchinsky, 1998; Koenker and Hallock, 2001) estimates different conditional quantiles of the dependent variables minimizing the sum of absolute residuals. It can be specified by [1]:

$$y_{it} = \alpha + x'_{it}\beta_{\tau} + u_{\tau it} \quad (1)$$

for $0<\tau<1$, and with

$$Quant_\tau = (y_{it}|x_{it}) = x'_{it}\beta_\tau \quad (2)$$

where *y* represents the dependent variable, **x** is a vector of all covariates, α is the term constant, β is the vector of parameters to be estimated and *u* is the vector of residuals. $Quant_\tau = (y_{it}|x_{it})$ specifies the τ-th conditional quantile of y given **x**, with i=1, 2, …, 20 region and t=2012, …, 2016 years.

The τ-th regression quantile solves the following minimization problem for ρ:

$$\min(\beta)[(\sum_{i=1}^{n} \rho_\tau(y_{it} - \beta'_\tau x_{it}))] \quad (3)$$

where $\rho_\tau[\cdot]$ is the check function defined as $\rho_\tau(u\tau_{it}) = \tau u_{\tau it}$ if $u_{\tau it} \geq 0$ and otherwise $(\tau - 1)u_{\tau it}$ if $u_{\tau it} < 0$.

In this paper for log of labour productivity we estimated five different quantile regressions with τ = 0.1, 0.25, 0.5, 0.75 and 0.9. In addition, we addressed heteroscedasticity by means of robust standard errors. Equation 4 specifies the estimated model for our data:

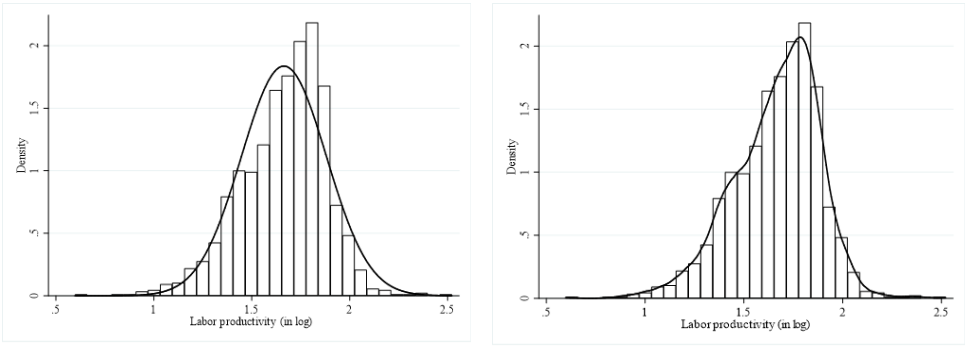
$$\log(LP) = \alpha + \beta_1 \log(IL_{it-1}) + \beta_2(ER_{it}) + \beta_3(IRS_{it}) + \beta_4(MCE_{it}) + \beta_5(HR_{it}) + \varepsilon_{it} \quad (4)$$

where: LP is the labour productivity for each region *i* at time *t*; IL represents the investments per employee for each region *i* at time *t*-1; ER is a proxy for export ratio (incidence of the sector on the total manufacturing exports of the region) for each region at time *t*; IRS is the indicator of innovation, research and creativity for region *i* at time *t*; MCE is the indicator of minimum economic conditions for region *i* at time *t*; HR represents the indicator of homicide in each region *i* to the time *t*.

As summarized by Buchinsky (1998) the quantile regression provides robust estimates of the vector of coefficients, not sensitive to outliers in the values of the dependent variable; in the presence of non-normally distributed error terms, the estimators provided by the quantile regression can be more efficient than least squares estimators. Looking at different estimates for different quantiles it is possible to assess the different influence of covariates on the dependent variable, at the various points of the quantile conditioned distribution. Finally, the estimate, based on a linear combination of estimators of the various quantile regressions, is always more efficient than the estimator of the least squares. The quantile regression parameter estimates the change in a specific quantile of the response variable produced by a one-unit change in the covariate, making it possible to compare if and how covariates influence some percentiles of the dependent variable (Velucchi and Viviani 2011, Velucchi et al. 2014, Holmes et al. 2019). In this paper, we have chosen to estimate models by quantile regression for three main reasons. First, the standard least-squares assumption of normally distributed errors does not hold for

this database because labour productivity in the Italian regions does not display a Gaussian distribution (Figure 1).

Figure 1 – *Histogram and Normal Density Plot (Panel A) and Kernel Density Plot (Panel B) of Labour Productivity (in log).*



Panel A Panel B

In addition, the quantile regressions describe all distributions of the dependent variable and do not focus on the mean (as OLS regression) and their use in the context of this study could be useful since high/low labour productivity regions are of interest for us and are not considered outliers. Finally, using this methodology and avoiding the assumption that the error terms are identically distributed it is possible to consider the regions' heterogeneity and the possibility that estimated slope parameters vary at different quantiles of the conditional distribution.

3. Results

Tables 1 and 2 report respectively descriptive statistics for selected variables in and the same descriptive statistics disaggregated on labour productivity quantiles. In the Appendix, in figures I-VI, we show the cartograms of the labour productivity median (in log) and of all variables considered in this study. The gap between the regions of the Centre-North and those of the South of the country in labour productivity is clear in manufacturing sectors, with the former showing better situations than the latter. There were similar results for all indicators considered.

Table 1 – *Descriptive Statistics for Manufacturing*.

Var.	Mean	Std. Dev.	Min	Max
ln_LP	3.755	0.478	1.380	5.050
ln_IL	1.502	0.965	-9.210	4.120
ER	4.697	7.713	0.000	82.770
HR	102.132	7.425	72.300	113.800
IRS	98.407	9.600	79.300	124.100
MEC	95.324	11.061	65.800	109.100

Table 2 – *Descriptive Statistics for Manufacturing by labour productivity quantiles*.

Var.		10%	25%	50%	75%	90%
ln_PL	Mean	3.045	3.300	3.637	3.950	4.192
	sd	*0.064*	*0.092*	*0.100*	*0.087*	*0.059*
ln_IL	Mean	0.738	0.854	1,353	1.721	2.045
	Sd	*0.538*	*1.162*	*0.759*	*0.652*	*0.619*
ER	Mean	1.935	2.969	3.780	5.296	7.312
	Sd	*4.897*	*5.720*	*6.804*	*7.329*	*10.164*
HR	Mean	96.675	99.219	102.393	103.342	103.957
	Sd	*8.552*	*9.745*	*6.327*	*5.929*	*5.640*
IRS	Mean	93.625	94.962	98.668	99.398	102.014
	Sd	*8.090*	*8.115*	*9.851*	*9.396*	*9.382*
MEC	Mean	86.304	89.417	95.515	97.631	99.241
	Sd	*9.948*	*12.005*	*10.438*	*9.733*	*9.725*

Table 3 reports the regression estimates for five different quantiles of the regional labour productivity distribution. Successively, in order to evaluate the importance of the differences in the quantile parameter estimates we test for the equality of coefficients between any two quantiles as well as jointly for all quantiles. The tests were performed using the F-statistic, the computation of which requires an estimate of the variance-covariance matrix of the quantile coefficients (table 4).

The results indicate that there are statistically significant differences in the coefficients and among the various quantile regression estimates for most explicative variables. In particular, the coefficient of ln(IL) (investments per employee) varies significantly from 0.176 to 0.268 as we move from the lower quantile (0.10) to the upper quantile (0.90) of the labour productivity conditional distribution. Probably reflecting the fact that the most productive regions are more sensitive to investments intensity while the less productive ones are more indifferent.

Table 3 – *Estimation results quantile regression model.*

	0.10	0.25	0.50	0.75	0.90
ln_IL	0.176**	0.236**	0.264**	0.268**	0.225**
	(0.023)	(0.020)	0.012	0.022	0.029
ER	0.008**	0.007**	0.008*	0.008**	0.006*
	0.002	0.002	0.002	0.002	0.002
HR	0.005*	0.010**	0.007*	0.004	0.003
	0.003	0.003	0.002	0.003	0.003
IRS	0.012**	0.009**	0.005**	0.002	-0.001
	0.003	0.002	0.001	0.002	0.002
MEC	0.010**	0.008**	0.009*	0.007*	0.007*
	0.002	0.002	0.001	0.002	0.002
Cons	0.833**	0.672*	1.263**	2.168**	2.933**
	0.038	0.270	0.212	0.255	0.317
Pseudo R^2	0.228	0.279	0.286	0.221	0.208

*Note: Standard errors (SEs) in parentheses. SEs for quantile regressions are derived via bootstrap techniques for 1000 replications. Significance levels: * p<0.05, ** p<0.01.*

Table 4 – *Test for coefficient equality between pairwise quantiles and across all quantiles.*

Quantile Group	*ln*_IL	ER	HR	IRS	*MEC*
Panel A					
0.10-0.25	0.002	0.315	0.126	0.216	0.160
0.10-0.50	0.000	0.848	0.676	0.006	0.821
0.10-0.75	0.001	0.964	0.986	0.001	0.296
0.10-0.90	0.141	0.180	0.611	0.000	0.403
0.25-0.50	0.083	0.501	0.231	0.022	0.160
0.25-0.75	0.183	0.426	0.040	0.004	0.917
0.25-0.90	0.720	0.301	0.049	0.000	0.867
0.50-0.75	0.830	0.842	0.574	0.054	0.165
0.50-0.90	0.127	0.144	0.041	0.001	0.351
0.75-0.90	0.038	0.074	0.487	0.082	0.744
Panel B					
Joint	0.000	0.315	0.049	0.000	0.271

Note: The null hypothesis is that the coefficients are equal between pairwise quantiles (panel A) and across all quantiles (panel B). Tests statistics are based on the variance-covariance matrix of the quantile coefficients estimated and reported in table 3. The table repots the p value foe the F-values; if the p-value is less than the level of significance (0.005), the hypothesis of equal coefficients is rejected.

The opposite picture prevails in the IRS and HR coefficients, which seem to be a more important productivity enhancing factors for the less productive areas. The exportation rate and MEC are statistically significant to estimate the regional labour productivity but not vary across quantile distribution.

It is interesting to notice that the in-between coefficient differences of HR and IRS are significant in the joint test among all five quantiles. We can interpret these results as evidence that environmental factors do not matter among the very productive regions. It is in the less- middle productivity range that these factors confirm their superior efficiency by causing a productivity shift.

4. Conclusions

As stated, organized crime influences the market economy in the South, influencing its development in a negative way (Calamunci and Drago, 2020). Mafia activity within legitimate markets "confuses" other competitors, as it creates barriers which prevent numerous industries from entering both production markets and work markets. In many respects, these markets are much less competitive in those regions affected by organized crime in comparison to other regions. In some extreme cases, where the mafia manages to control both the supply and demand of goods supplied by the State, the markets (both corrupt and normal markets) are suppressed and there is a hierarchic economic organization, in which those businesses outside the cartel, or those potential candidates for entry, are forced to deal with very high transaction costs. This institutional environment is a source of inefficiency and low productivity growth (Felli and Tria, 2000).

According to Centorrino and Ofria (2008), crime is interested in productive sectors that are directly or indirectly reached by State interventions, that are not very open to external competition (since they are non-tradable sectors), with high labour intensity rather than capital intensity, in such a way as to leave wider margins for money laundering, and also to guarantee some forms of social consensus, through the distribution of work opportunities.

Such circumstances discourage entrepreneurs to invest in these territories. (Detotto and Otranto, 2010, Daniele and Marani, 2011, Brown and Hibbert, 2019).

Starting from this premise, the aim of research is to examine whether some territorial indicators of equitable and sustainable well-being (BES), proxies of the socio-environmental context, can contribute to influencing the Italian regional labour productivity in manufacturing sector. As proxies for socio-environmental factors we used some equitable and sustainable well-being composite indicators that ensuring temporal and territorial comparability. The choice to use these indicators stems from the awareness that measuring well-being, also and above all, from an economic point

of view, requires a multifaceted statistical approach since no single measurement can summarize the multidimensional value of something as complex as the well-being of society. By adopting quantile regression approach, we highlighted that labour productivity is heterogeneous and that the relationship between labour productivity, socioeconomic context, managerial factors, and a firm's characteristics is not constant across quantiles. In particular, the empirical answers provided by our analysis support the theoretical proposition that the higher the social and environmental well-being of a territory, the more efficient the production is. We can interpret these results as evidence that environmental factors do not matter among the regions with firms very productive. It may be that companies are negatively influenced to invest and operate in areas of the Italian territory where the population insecurity sense and the fear of being a victim of criminal acts are high. In our opinion, to increase economic performance in less productive Italian regions, it is essential to invest in safety and to improve the minimum economic condition to remove all the factors that hinder the will or the ability to invest of a firm. Furthermore, it is important to continue investing in innovation and research which has proved to be a growth factor in areas where labour productivity is lower but also in investments per employees which, on the contrary, is a factor whose importance increases as it grows productivity.

Appendix

Cartograms of all variables considered in Italian regions.

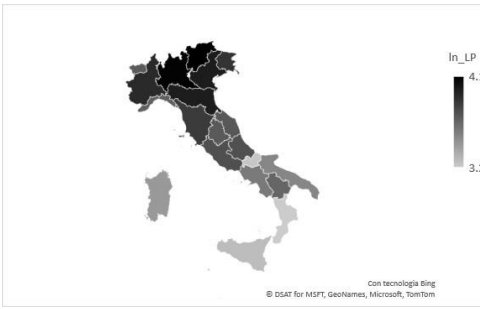


Figure I - Maps of ln(LP) median

Figure II - Maps of ER Median

Figure III - Maps of ln(IL)

Figure IV - Maps of HR median

Figure V - Maps of MEC median

Figure VI - Maps of IRS median

References

ACEMOGLU D., DE FEO G., DE LUCA G.D. 2020. Weak states: Causes and consequences of the Sicilian mafia. *The Review of Economic Studies*, Vol. 87, pp. 537-581.

AHLAWAT V., RENU. 2018. An analysis of growth and association between labour productivity and wages in Indian textile industry. *Management and Labour Studies*, Vol. 43, No. 1-2, pp. 78-87.

ALESINA A., PICCOLO S., PINOTTI P. 2019. Organized crime, violence, and politics. *The Review of Economic Studies*, Vol. 86, No. 2, pp. 457-499.

BRISTOW G., HEALY A. 2018. Innovation and regional economic resilience: an exploratory analysis. *The annals of regional science*, Vol. 60, No. 2, pp. 265-284.

BROWN L, HIBBERT K, 2019. The Incidence of crime on industry-level foreign direct investment: An assessment of OECD member countries. *Social Science Quarterly*, Vol. 100, No. 4, pp. 1228-1240.

BUCHINSKY M., 1998. Recent advances in quantile regression models: A practical guide for empirical research, *Journal of Human Resources*, Vol. 33, pp. 88-126.

CALAMUNCI F., DRAGO F. 2020. The economic impact of organized crime infiltration in the legal economy: Evidence from the judicial administration of organized crime firms. *Italian Economic Journal*, Vol. 6, pp. 275-297.

CASTELLANI D., MONTRESOR S., SCHUBERT T., VEZZANI A. 2017. Multinationality, R&D and productivity: Evidence from the top R&D investors worldwide. *International Business Review,* Vol. 26, No. 3, pp. 405-416.

CENTORRINO M., OFRIA F. 2008. Criminalità organizzata e produttività del lavoro nel Mezzogiorno: un'applicazione del "modello Kaldor-Verdoorn". *Rivista economica del Mezzogiorno*, Vol. 22, No. 1, pp. 163-188.

D'AGOSTINO G., SCARLATO M. 2015. Innovation, socio-institutional conditions and economic growth in the Italian regions. *Regional Studies*, Vol. 49, No. 9, pp. 1514-1534.

DANIELE V., MARANI U. 2011. Organized crime, the quality of local institutions and FDI in Italy: A panel data analysis. *European Journal of Political Economy*, Vol. 27, No. 1, pp. 132−142.

DETOTTO C., OTRANTO E. 2010. Does crime affect economic growth? *Kyklos*, Vol. 63, pp. 330−345.

FELLI E., TRIA G. 2000. Produttività e crimine organizzato: un'analisi delle regioni italiane, *Sviluppo Economico*, Vol. 1, pp. 79-101.

GANAU R., RODRÍGUEZ-POSE A. 2018. Industrial clusters, organized crime, and productivity growth in Italian SMEs. *Journal of Regional Science*, Vol. 58, pp. 363-385.

HOLMES M., DOAN T., HASSAN G. 2019. Does foreign investment enhance domestic manufacturing firms' labour productivity? Evidence from a quantile regression approach. *Economic Change and Restructuring*, pp. 1-18.

KALDOR N. 1967. *Causes of the slow rate of economic growth in the United Kingdom: An inaugural lecture*. Cambridge University Press. Cambridge.

KOENKER R., HALLOCK K.F. 2001. Quantile regression. *Journal of Economic Perspectives*, Vol. 15, No. 4, pp. 143-156.

MAZZIOTTA M., PARETO A. 2016. On a generalized non- compensatory composite index for measuring socio-economic phenomena. *Social Indicators Research*, Vol. 127, pp. 983–1003.

MILLEMACI E., OFRIA F. 2016. Supply and demand-side determinants of productivity growth in Italian regions. *Structural Change and Economic Dynamics*, Vol. 37, pp. 138-146.

MUNDAKKAD P. 2018. Firms' leverage and labour productivity: A quantile regression approach. *Economics Bulletin*, Vol. 38, No. 4, pp. 2331-2344.

NESE A., TROISI R. 2019. Corruption among mayors: Evidence from Italian court of cassation judgments. *Trends in Organized Crime*, Vol. 22, No. 3, pp. 298-323

OFRIA F., DAVID P. 2014. *L'economia dei beni confiscati*, Franco Angeli, Milan.

OFRIA F., MUCCIARDI M. 2021. Government failures and non-performing loans in European countries: a spatial approach, *Journal of Economic Studies*, ahead-of-print. https://doi.org/10.1108/JES-01-2021-0010

RONCOLATO L., KUCERA D. 2014. Structural drivers of productivity and employment growth: A decomposition analysis for 81 countries. *Cambridge Journal of Economics,* Vol. 38, No. 2, pp. 399-424.

SYLOS LABINI P. 2004. *Torniamo ai classici. Produttività del lavoro, progresso tecnico e sviluppo economico*. Laterza. Rome-Bari.

VELUCCHI M., VIVIANI A. 2011. Determinants of the Italian labor productivity: A quantile regression approach, *Statistica,* Vol. 71, No. 2, pp. 213-238.

VELUCCHI M., VIVIANI A., ZELI A. 2014. Italian manufacturing and service firms labor productivity: A longitudinal quantile regression analysis. *Statistica,* Vol. 74, No. 3, pp. 267-293.

SUMMARY

The influence of BES territorial indicators on economic performance of manufacturing firms

The research aims to verify whether some BES territorial indicators influence the Italian regional labour productivity in the manufacturing sector. The quantile regression approach allows us to highlight that labour productivity is heterogeneous that the relationship between labour productivity and environmental and firm's characteristics is not constant across quantiles. Our results show that labour productivity is affected by negative externalities such as the homicide rate and minimum economics conditions and that these indicators have a greater influence in regions with lower labour productivity.

Romana GARGANO, Department of Economics, rgargano@unime.it
Ferdinando OFRIA, Department of Economics, ofriaf@unime.it

RISK HABITS AMONG ITALIAN UNIVERSITY STUDENTS: SENSATION SEEKING AND SEXUAL RISK-TAKING BEHAVIOURS [1]

Giuseppe Gabrielli, Giancarlo Ragozini, Antonietta Bisceglia, Anna Paterno

1. Introduction

Abuse of alcohol and tobacco and use of illegal drugs are common during adolescence. Indeed, a behaviour may be risky when it appears in combination with other behaviours (Vasilenko et al., 2015). Having multiple risk habits at the same time increases the risk a sexually transmitted disease. During young ages, the risk habits are generally linked with sexual risk-taking behaviours (Castilla et al., 1999; Kipping et al., 2012; Pirani, Matera, 2020). There is wide evidence that these risk habits may be associated to risk behaviours in sexual life and leaded to risk taking profile and to trait of personality (Ellickson et al., 2001; Stueve, O'Donnell, 2005).

Generally speaking, sexual risk behaviours, i.e. unprotected sexual intercourse, casual sex, etc., can lead to a number of negative consequences, including damage to young couple relationships, family conflicts, financial concerns, damage to social reputations, health problems, and legal disputes. However, the two most addressed outcomes are unintended pregnancies and sexually transmitted infections (STIs), including HIV/AIDS.

Many scientific studies show that the habit of risk behaviours during adolescence and youth can be often maintained even during adulthood. Risk behaviours at young ages have not only short-term consequences on health, but also along the life-course, especially in relation to development of diseases and wrong lifestyles (Connell et al., 2009). These behaviours significantly affect the state of health and psychophysical well-being at both adult ages and at old ages (Dembo et al., 1992).

We aim at exploring the patterns of relationships among risk habits, in particular the sex related ones, among Italian young people, in order to identify different risk profiles of individuals, according to their characteristics, attitudes and behaviours.

[1] This work is the result of a close collaboration between the authors. As for this version, sections 1 and 2 has been written by Antonietta Bisceglia; sections 3 and 4 has been written by Giuseppe Gabrielli and Giancarlo Ragozini; section 5 by Anna Paterno.

2. Theoretical background

Risk taking attitude can be an expression of personality trait, denoted as sensation seeking (SS). This is an individual's inclination to engage in risky and emotional behaviours in order to obtain new and exciting sensations and experiences (Dunlop, Romer, 2010). SS is "the tendency to prefer exciting, optimal, and novel levels of stimulation or arousal" (Kalichman et al., 1994: 386). Zuckerman (1979) defined four dimensions of SS. *Thrill and Adventure Seeking* is "the desire to engage in sports or other activities involving speed or danger". *Experience Seeking* is a measure of "the seeking of new experiences through the mind and senses, and through an unconventional, non-conforming life-style". *Boredom Susceptibility* assesses "the dislike of repetition of experience, routine work, predictable dull or boring people, and restlessness when things are boring". *Disinhibition* measures "the desire to find release through social disinhibition, drinking, going to parties and having a variety of sexual partners".

It is necessary to adopt a multidimensional approach aimed at analyzing how the different levels of different sexual risk-taking behaviours (SRB) are associated (Potard et al., 2019). Hoyle et al. (2000) considered three attitudes among the SRB: unsafe sex, multiple partners, and high risk intercourse.

Several factors can influence independently or interactively SS and SRB especially during the adolescents (Spitalnick et al., 2007; Pirani, Matera, 2020). Different studies showed a not univocal gender-role in SS and SRB: some of them outlined no differences among men and women (Hendershot et al., 2007), while others showed that men have an higher propensity in SS and SRB (Gaither, Sellbom, 2003). SS and SRB have also unclear patterns among homosexuals and heterosexuals (Doll et al., 1991; Zuckerman, Neeb, 1980): both studies tend to exclude any difference, even though they are focused only on males and they are not intended as comparative studies. On the contrary, the religiosity is one of the factors that in literature appear more often to protect an individual from risky behaviours (Zuckerman, Neeb, 1980; Ameri et al., 2017), because it disciplines moral and sets social rules (Yonker et al., 2012). If we consider the negative scholastic experiences, the school dropout, instead, is considered a risk factor for SRB (Flisher at al., 1995), especially related to the outcomes of unintended pregnancies. Parental monitoring and family characteristics are very important when examining SS and SRB among adolescents (Huebner et al., 2003). With respect to SS attitude studies are mainly focused on the use of drugs. Parental monitoring is considered a mediator factor on SS with respect to possible negative outcomes related to the drug abuse.

3. Data and methods

We examine data from the Sexual and Emotional LiFe of Youths (SELFY) survey, which was carried out in 2017 in 28 Italian universities with the aim of drawing an updated picture of sexual and emotional attitudes and behaviours among almost 8,000 (20% of the total) young Italian university students, attending undergraduate courses in economics and statistics in Italian state universities (Minello et al., 2020; Dalla-Zuanna, Vignoli, 2021). Data are weighted to make them representative of this group of university students at the national level.

The main limitation is that we are not dealing with a representative sample of the whole youth population. Nonetheless, a sample comprised of university students has many advantages: mainly, a high number of respondents who are well disposed towards filling in a relatively long, although not complex, questionnaire. Within this group, there exists relevant heterogeneity regarding both sexuality and affective behaviours, which makes university students relevant subjects of research (Berntson et al., 2014; Stinson, 2010; Weeden, Sabini, 2007).

In our analyses, we defined two different score scales of SS and of SRB. We excluded from the analysis students that had missing information on the variables that in the following we use for the construction of the two scales. All sampled students (N=7,841) are considered when analysing SS, while we restricted the sample to 6,068 young people who have had at least one full sexual intercourse when analysing SRB.

Although the questionnaire was not built to measure the SS, there were the necessary items to measure two dimensions defined so far by Zuckerman (1979), namely: experience seeking and disinhibition. We make use of the following 12 items to obtain a final scale od SS: reading sex magazine, watching pornographic DVDs, go clubbing, go into a sexy shop, having exchanged in sexting, have seen a striptease, visit to porno sites, drunkenness, ever used ecstasy, ever used marijuana, high speed driving, self-pleasuring.

As for SRB, we exploit the following items to measure the three attitudes proposed by Hoyle et al. (2000): level risk of the contraceptive method used during first sexual relationship and the level risk of the contraceptive method used during last sexual encounter for the unsafe sex, the number of sexual partner for the tendency of having multiple partners, and having paid for sex, having sexual intercourse with a person when you're with another, unsafe sex with non-stable partner (sex without condom) for the high risk intercourses.

Aiming at obtaining behavioural profiles, we adopt the so-called tandem analysis in order to find the profiles that consists in a sequential two-step analysis (Lebart, 1994), i.e. first, we first analyse by a Multiple Correspondence Analysis (MCA) the association among the selected categorical variables of SS and SRB along with other variables able to define individual profiles, and after we use a hierarchical clustering

algorithm on the derived factorial scores. Because the to scales clearly contributed in defining profile, we tested the one-dimensionality of SS and SRB scales through the Cronbach's Alpha procedure. In the end, we estimated liner regression models, that assume as dependent variables the two SS and SRB scales defined so far, and observed three groups of factors associated to them, namely: individual characteristics and experiences, family characteristics and relations with parents, and individual perception and self-confidence.

According to the literature, we considered the following individual characteristics: gender, religion, sexual identity, negative scholastic experience, the first sexual experience. The second group of factors included the variables concerning parental monitoring and parenting style. Last, a number of individual perception and self-confidence are included in the models. Our findings provide mixed support for the role of both set of variables in high scores of SS and in the prediction of high SRB.

4. Results

Looking at the variation in the between-group and within-group inertia, as well as at the dendrogram (here not shown for space reasons), results highlight the presence of eight different target groups (or clusters) in terms of risk taking propensity and behaviours. Considering the selected characteristic items, it is possible to order these clusters according to a continuum scale, where at one hand we find groups characterized by high risk-taking behaviours, and on the other hand, we find groups with low levels of risk behaviours.

More specifically, in the first cluster (7.2% of the sample) there are the young "sensation seekers". This personality trait manifests itself in a variety of risky behaviours. In literature, the sensation seeker are characterized by the generalized tendency to seek varied, novel, complex, and intense sensations and experiences and the willingness to take risks for the sake of such experiences (Zuckerman, 2007). They do not show awareness about their risky behaviours.

In the second group (17.7% of the sample) there are the young "sensation seekers who are vulnerable to pornography". This group characterized by the vulnerability to pornography and an explicit denial of religion. Furthermore, they have a more correct level of information about sexual behaviour at risk of HIV, and an awareness of risks of their sexual personal life. In this group, there is a significant correlation between the indicators of used pornography and sexual risk behaviours. Although this young people are aware of the preventive efficacy of condoms, even though most of them do not use any protection on a consistent basis.

The sexual activity of both above-mentioned groups extends to multiple partners, casual and unsafe sex. They show also the significant propensity to conduct other risk behaviours like the abuse of alcohol and drugs as well as tobacco.

The third cluster (17.7% of the sample) is composed by "reticent" young people with a sexual active life. This group answered negatively to almost all questions on the deviant behaviours and on their sexual habits. They are all heterosexual males and enough informed about the risk of HIV and STD, but take risks in sexual life.

In the fourth cluster (23.4% of the sample), there are young people with a sexual active life, very informed about the sexual risks and mostly responsible in their sexual life. They know that the correct and consistent use of condoms is a highly effective means of reducing the risk of sexual transmission of HIV and other STDs as well as unintended pregnancy.

Cluster five (15.9% of the sample) includes heterosexual young girls with active and prudent sex life. They know sexual risks and sexually transmitted diseases. They have no risk experiences for health motivations, and know that limiting sexual activity to a single, uninfected partner is another highly effective mean of managing the risks associated with sexual activity.

Cluster six (5.9% of the sample) is made up of young people worried about HIV. They are highly spooked about sexually transmitted diseases and HIV. For these reasons, they use the condom as prevention method of sexually transmitted infections, including HIV.

Interviewees with inactive sexual life are included in the last two clusters (respectively 3.3% and 19.5% of the sample). The two groups differ by the gender composition. The cluster seven is mostly composed of males, while, cluster eight is mainly made up of females.

Cluster analysis shows the presence of the latent factors and also a certain degree of one-dimensionality. For this reason, we proceed to define and test the two scales of SS and SRB. We report in Table 1 and 2 the Cronbach's Alphas, also considering the item deletion. The values obtained are quite satisfying for both scales (0.734 for the SS and 0.624 for SRT) denoting an adequate level of one-dimensionality. Considering the variations in the alpha values when the item is deleted, all the items used to measure the two scales are necessary.

The scores of the two scales can be used within a linear regression framework that showed interesting results on individual characteristics and experiences that are associated to SS and SRB (Table 3). Males and young people who had early sexual debut had a greater propensity to SS and SRB. SS and SRB are positively associated to young people: who declared to be omosexual or bisexual; who had both coercive sexual relations and disease after one sexual experience; who experienced failed at school; and who did not accept school-related rules. Those who are fully satisfied of their physical appearance are also more positively associated in having SRB than the others.

Table 1 - *Scale of Sensation Seeking by selected items. Cronbach's Alpha.*

Item	Cronbach's Alpha if Item Deleted
Reading sex magazine (in the last 12 month)	.722
Watching a pornographic Dvds (in the last 12 month)	.717
Go Clubbing (at present time)	.731
Go into a sexy shop (in the last 12 month)	.718
Having exchanged in sexting (in the last 12 month)	.689
Have seen a striptease (in the last 12 month)	.717
Visit to porno sites (in the last 12 month)	.677
Drunkenness – (at present time)	.694
Ever used ecstasy- (at present time)	.727
Ever used marijuana - (at present time)	.700
High speed driving- (at present time)	.707
Self pleasuring (at present time)	.691
Cronbach's Alpha	*.734*

Source: our elaboration on SELFY data, 2017.

Table 2 - *Scale of Sexual Risk-Taking Behaviours by selected items. Cronbach's Alpha.*

Item	Cronbach's Alpha if Item Deleted
Level risk of the contraceptive method used during first sexual relation	.579
Number of sex partner	.527
Level risk of the contraceptive method used during last sexual encounter	.540
Having for sex	.593
Having sexual intercourse with a person when you're with another	.548
Unsafe sex with non-stable partner (sex without condom)	.476
Cronbach's Alpha	*.624*

Source: our elaboration on SELFY data, 2017.

Conversely, we observed lower significant levels of SS among those young people who consider religion as important (the same is not significant for SRB). When considering family characteristics and relations with parents, young people who had divorced parents or who had negative relationships with their parents (more significantly with mother than with father) had a greater propensity to SS and SRB. As well, more liberal and permissive parental behaviours (that allow their children to return late at home or to live sexual intimacy at home) are positively associated to SS and SRB. Individual perception and self-confidence are related to sexuality. SS significantly increases for those who live the body very intensely.

Table 3 - *Determinants of Sensation Seeking (SS) and Sexual Risk Behaviours (SRB). Linear regression models. Coefficients and p-values.*

Variables	Categories	SS		SRB	
		coef.	p-val.	coef.	p-val.
Gender (ref. Boy)	Girl	-0.313	***	-0.100	***
Age at interview (ref. less than 21 years old)	21 yrs. old +	-0.004		0.100	***
Age at first sexual intercourse (ref. No experience)	<16	0.137	***	---	
	16	0.103	***	-0.136	***
	17	0.082	***	-0.198	***
	18	0.071	***	-0.249	***
	>18	0.063	***	-0.309	***
Dimension of Municipality (ref. <20 thou.)	20-100 thou.	0.011	*	0.004	
	100 thou. +	0.003		0.025	*
Are you omosexual or bisexual? (ref. No)	Yes	0.093	***	0.174	***
Failed at school (ref. No)	Yes	0.014		0.076	***
Follow the rules at school (ref. No)	Yes	0.068	***	0.071	***
Is religion important for you? (ref. No)	Yes	-0.042	***	-0.018	
Are you satisfied with your physical appearance? (ref. No)	Almost yes	-0.010	*	0.010	
	Definitely yes	-0.003	*	0.052	***
Have you had sexual experiences against your will? (ref. No)	Yes	0.106	***	0.192	***
Have you had diseases after one sexual experience? (ref. No)	Yes	0.071	***	0.183	***
Have at least one brother (ref. No)	Yes	0.010	*	0.004	
Have at least one sister (ref. No)	Yes	0.008		0.030	**
Divorced parents (ref. No)	Yes	0.015	**	0.040	**
Type of relation with father (ref. OK)	Deceased	0.002		0.049	
	Negative	0.004		0.061	***
	Positive	-0.013	**	0.013	
Type of relation with mother (ref. OK)	Deceased	0.050		0.191	**
	Negative	0.042	***	0.098	***
	Positive	-0.015	**	0.019	
Do your parents allow you to return late at home at 16-18 years old? (ref. No)	Sometimes	0.031	***	0.026	*
	Often	0.074	***	0.065	***
Do your parents allow you to live sexual intim. at home at 16-18 years old? (ref. No)	Sometimes	0.010		0.007	
	Often	0.057	***	0.027	*
Are you agree with the following sentences? To live the body very intensely (ref. Definitely not agree)	Almost not agree	0.008		0.081	*
	No position	-0.016		-0.019	
	Almost agree	0.031	***	-0.028	***
	Definitely	0.099	***	0.076	***

(continues)

*Legend: * p<0.05; **p< 0.01; ***p<0.001*
Source: our elaboration on SELFY data, 2017.

Table 3 - *(continues) Determinants of Sensation Seeking (SS) and Sexual Risk Behaviour (SRB). Linear regression models. Coefficients and p-values.*

Variables	Categories	SS		SRB	
		coef.	p-val.	coef.	p-val.
Are you agree with the following sentences?					
To follow fashion	Almost not agree	-0.046	***	-0.038	
(ref. Definitely not agree)	No position	0.002		-0.026	
	Almost agree	0.013	**	0.021	
	Definitely agree	0.042	***	0.026	
To live without purpose	Almost not agree	-0.069	***	-0.054	***
(ref. Definitely not agree)	No position	-0.033	***	-0.030	*
	Almost agree	-0.013		-0.034	*
	Definitely agree	0.010		0.001	*
It is not enough for me to live	Almost not agree	0.036	***	0.048	**
(ref. Definitely not agree)	No position	0.007		-0.011	
	Almost agree	-0.027	***	-0.005	
	Definitely agree	-0.020	*	0.027	
Constant term		1.456	***	1.214	***
N		7,841		6,068	
R-squared		0.439		0.230	

*Legend: * p<0.05; **p< 0.01; ***p<0.001*
Source: our elaboration on SELFY data, 2017.

5. Discussion and conclusion

Despite data present some limitations, mainly due to the characteristics of the sample (that includes only university students), the present paper has the advantage of trying to fill the existing gap, especially in the Italian research, on the related topic, by simultaneously examining co-occurrence of two different risky behaviours in a multidimensional perspective that includes individual, familiar and social behaviour, besides attitudes of youths. Moreover, the novelty of this study is that, despite the cross-sectional structure of data, we used, where applicable, retrospective information to preserve the casual effect between sexual and risks behaviours and their determinants.

Our analyses showed that Italian university students present very heterogeneous profiles on their risk habits. In this perspective, our descriptive analyses showed that the eight clusters of observed students offered a very nuanced perspective on their behaviors and risk taking propensity.

The analyses performed through a multidimensional approach showed that a wide range of factors are associated to SS and SRB. In particular, the role played by the individual characteristics about respondents and their families, also including a

number of individual perception and opinions, emerged. In other words, SS can have interactive effects on SRB, being an important factor in determining the habit to have risky sexual behaviour).

Both the descriptive analyses and the regression model showed that different behaviours characterize men with respect to women. Confirming previous researches, gender is configured as one of the most important determinant of attitudes and behaviors, as it translate in different notions of appropriate behaviors for men and women. Our findings outlined also that another important determinant of risk profiles are the familiar and institutional contexts (e.g. relationship with the mother and the observance of school rules). Lastly, it is notably the effects of having had sexual experiences against own will, that positively influences both SS and SRB.

These empirical evidences obtained should stimulate future research to further clarify mechanisms and causality patterns that affect attitudes and behaviours of youths. Among the others, scholars showed that sensation seeking is highly linked with performing sexual risk behaviours (Spitalnick et al., 2007). In other words, SS can have interactive effects on SRB, being an important factor in determining the habit to have risky sexual behaviour.

Furthermore, our study indicates that positive family relationships and parental monitoring should be enforced to reduce youth engaging in sexual and behavioral risk (Huang et al., 2011). On the other hand prevention and health-risk reduction programs should be realized to capitalize on the protective role of the school and/or university context against youth risk behaviours.

Acknowledgements

Preliminary results included in this paper were presented at Workshop “Uguali o diversi? I giovani di fronte alla sessualità” [Similar or different? Young people facing sexuality] held in Florence (September 17-18, 2018) and at the XIII edition of Pop-Days held in Milan (January 24-26, 2019). We would like to thank participants for useful insights on previous versions of this manuscript. As usual, all errors remain ours. Antonella Bisceglia and Giancarlo Ragozini are partially funded by the grant “Osservatorio Regionale delle Politiche Giovanili – Università di Napoli Federico II”, POR Campania FSE 2014-2020.

References

AMERI Z., MIRZAKHANI F., NABIPOUR A. R., KHANJANI N., SULLMAN M. J. 2017. The relationship between religion and risky behaviours among Iranian university students, *Journal of religion and health*, Vol. 56, No. 6, pp. 2010-2022.

BERNTSON M.A., HOFFMAN K.L., LUFF T.L. 2014. College as context: Influences on interpersonal sexual scripts, *Sexuality & Culture*, Vol. 18, No. 1, pp. 149-165.

CASTILLA J., BARRIO G., BELZA M.J., DE LA FUENTE L. 1999. Drug and alcohol consumption and sexual risk behaviour among young adults: results from a national survey, *Drug and alcohol dependence*, Vol. 56, No. 1, pp. 47-53.

CONNELL C.M., GILREATH T.D., HANSEN N.B. 2009. A multiprocess latent class analysis of the co-occurrence of substance use and sexual risk behavior among adolescents, *Journal of studies on alcohol and drugs*, Vol. 70, No. 6, pp. 943-951.

DALLA-ZUANNA G., VIGNOLI D. 2021. *Piacere e fedeltà. I millennials italiani e il sesso,* Bologna: il Mulino.

DEMBO R., WILLIAMS L., SCHMEIDLER J., BERRY E. 1992. A structural model examining the relationship between physical child abuse, sexual victimization, and marijuana/hashish use in delinquent youth: A longitudinal study, *Violence & Victims*, Vol. 7, pp. 41-62.

DOLL L.S., BYERS R.H., BOLAN G.A., DOUGLAS JR I.L., MOSS V.P., WELLER P.D., ..., HARRISON J.S. 1991. Homosexual men who engage in high-risk sexual behavior. A multicenter comparison, *Sexually Transmitted Diseases*, Vol. 18, No. 3, pp. 170-175.

DUNLOP S., ROMER D. 2010. Adolescent and young adult crash risk: Sensation seeking, substance use propensity and substance use behaviours. *Journal of Adolescent Health*, Vol. 46, pp. 90–92.

ELLICKSON P.L., TUCKER J.S., KLEIN D.J. 2001. High-risk behaviours associated with early smoking: results from a 5-year follow-up, *Journal of Adolescent Health*, Vol. 28, pp. 465–73.

FLISHER A.J., CHALTON D.O. 1995. High-school dropouts in a working-class South African community: selected characteristics and risk-taking behaviour, *Journal of adolescence*, Vol. 18, No. 1, pp. 105-121.

GAITHER G.A., SELLBOM M. 2003. The sexual sensation seeking scale: Reliablity and validity within a heterosexual college student sample, *Journal of Personality Assessment*, Vol. 81, pp. 157.

HENDERSHOT C.S., STONER S.A., GEORGE W.H., NORRIS J. 2007. Alcohol use, expectancies, and sexual sensation seeking as correlates of HIV risk behavior

in heterosexual young adults, *Psychological Addict Behavior*, Vol. 21, No. 3, pp. 365-72.

HOYLE R.H., FEJFAR M.C., MILLER J.D. 2000. Personality and sexual risk taking: A quantitative review, *Journal of personality*, Vol. 68, No. 6, pp. 1203-1231.

HUANG D.Y., MURPHY D.A., HSER Y.I. 2011. Parental monitoring during early adolescence deters adolescent sexual initiation: Discrete-time survival mixture analysis, *Journal of Child and Family Studies*, Vol. 20, pp. 511-520.

HUEBNER A.J., HOWELL L.W. 2003. Examining the relationship between adolescent sexual risk-taking and perceptions of monitoring, communication, and parenting styles, *Journal of adolescent health*, Vol. 33, No. 2, pp. 71-78.

KALICHMAN S.C., JOHNSON J.R., ADAIR V., ROMPA D., MULTHAUF K., KELLY J.A. 1994. Sexual sensation seeking: Scale development and predicting AIDS-risk behavior among homosexually active men, *Journal of personality assessment*, Vol. 62, No. 3, pp. 385-397.

KIPPING R.R., CAMPBELL R.M., MACARTHUR G.J., GUNNELL D.J., HICKMAN M. 2012. Multiple risk behaviour in adolescence, *Journal of Public Health*, Vol. 34, No. 1, pp. 1-2.

LEBART L. 1994. Complementary use of correspondence analysis and cluster analysis. In Greenacre M.J. and Blasius J. (Eds.) *Correspondence Analysis in the Social Sciences*, London: Academic Press, pp. 162-178.

MINELLO A., CALTABIANO M., DALLA-ZUANNA G., VIGNOLI D. 2020. Catching up! The sexual behaviour and opinions of Italian students (2000–2017), *Genus*, Vol. 76, No. 16, pp. 1-22.

PIRANI E., MATERA C. 2020. Who is at risk? Gendered psychological correlates in Italian students' sexual risk profiles, *Genus*, Vol. 76, No. 1, pp. 1-21.

POTARD C., LANCELOT C., COURTOIS R. 2019. Examining relationships between sexual risk–safety behaviours and physical selfconcept by gender: A cluster analytical approach, *Emerging Adulthood*, Vol.7, No. 1, pp. 31-44.

SPITALNICK J.S., DICLEMENTE R.J., WINGOOD G.M., CROSBY R.A., MILHAUSEN R.R., SALES J.M., ET AL. 2007. Brief report: Sexual sensation seeking and its relationship to risky sexual behaviour among African-American adolescent females, *Journal of adolescence*, Vol.30, No. 1, pp. 165–173.

STINSON R.D. 2010. Hooking up in young adulthood: A review of factors influencing the sexual behavior of college students, *Journal of College Student Psychotherapy*, Vol. 24, No. 2, pp. 98-115.

STUEVE A., O'DONNELL L.N. 2005. Early alcohol initiation and subsequent sexual and alcohol risk behaviours among urban youths, *American Journal of Public Health*, Vol. 95, pp. 887-93.

WEEDEN J., SABINI J. 2007. Subjective and objective measures of attractiveness and their relation to sexual behavior and sexual attitudes in university students. *Archives of sexual behavior*, Vol. 36, No. 1, pp. 79-88.

YONKER J.E., SCHNABELRAUCH C.A., DEHAAN L.G. 2012. The relationship between spirituality and religiosity on psychological outcomes in adolescents and emerging adults: A meta-analytic review, *Journal of adolescence*, Vol. 35, No. 2, pp. 299-314.

ZUCKERMAN M. 1979. *Sensation seeking: beyond the optimal level of arousal*. Hillsdale, New York: L. Erlbaum Associates.

ZUCKERMAN M. 2007. *Sensation seeking and risky behavior*. Washington, DC: American Psychological Association.

ZUCKERMAN M., NEEB M. 1980. Demographic influences in sensation seeking and expressions of sensation seeking in religion, smoking and driving habits. *Personality and individual Differences*, Vol. 1, No. 3, pp. 197-206.

SUMMARY

Risk habits among Italian university students: sensation seeking and sexual risk-taking behaviours

Using data coming from the survey on "Sexual and Emotional Life of Youths", we aim at exploring the patterns of relationships among risk habits, in particular the sex related ones among Italian university students to identify different risk profiles by using a sequential two-step analysis. Results highlight the presence of eight different target groups of students in terms of risk habits. Among them, the 24.9% of students can be defined as "young sensation seekers" who have the highest risk habits. Conversely, the 22.8% of students have inactive sexual life. Subsequently, linear regression models defined those characteristics, attitudes and behaviours associated to these risk habits. School and sexual experiences as well as family characteristics and relations with parents result to be significantly associated to risk habits.

Giuseppe GABRIELLI, Università di Napoli Federico II, giuseppe.gabrielli@unina.it
Giancarlo RAGOZINI, Università di Napoli Federico II, giragoz@unina.it
Antonietta BISCEGLIA, Università di Napoli Federico II, antonietta.bisceglia@unina.it
Anna PATERNO, Università di Bari "A. Moro", anna.paterno@unina.it

THE ROLE OF SOCIAL CAPITAL IN ECONOMIC PERFORMANCE ACROSS EUROPEAN REGIONS

Mariateresa Ciommi, Francesca Mariani, Maria Cristina Recchioni, Riccardo Zallocco

1. Introduction

Economic growth is determined by a large range of aspects and the literature focuses on a variety of factors. Among these factors, scholars have also introduced social capital in recent decades. Social capital is defined by Poteyeva (2009) as the process of relationships, trust and reciprocal exchanges of benefits within social networks that ensure various advantages, also in problem solving. Considering these various characteristics, it could be a resource beyond economic aspects for a community and its inhabitants. This type of capital is composed of intangible elements, for example, personal values, norms, perceptions about behaviour, and opinions, as well as more concrete and objective measures.

Pioneers in the study of social capital include Coleman (1988) and Putnam (1993). They show that social capital has a (positive) impact on the sense of citizenship, a well-functioning society, and the point of view of people and firms in the development of collaboration and cohesion. Focusing on Italian regions, Putnam (1993) depicts civic culture as a determinant of economic growth, defining social capital as unique, composed of interdependent elements, such as trust, social norms, and networks. Moreover, he argues that non-hierarchical and spontaneous associative activities facilitate communication and relationships play a decisive role. These cultural characteristics are mostly inherited. Thus, any changes in a population could lead to a societal transform only in the long run.

The concept of social capital divides researchers. Some scholars maintain that the clear distinctions among different aspects of social capital do not allow everyone to work concurrently (Bjørnskov, 2006), nor is the impact unique. In fact, using different statistical techniques, studies provide diverging results. Moreover, with regard to territorial analysis (country or regional levels), there are also contrasting points of view, confirming the unique aspects of this field of study. Knack and Keefer (1997) state that trust, civic norms, and associational activities influence economic development in 29 market economies. These results were confirmed and strengthened in Zak and Knack (2001) and Beugelsdijk *et al*. (2004). Results show that the growth is affected mainly by variables related to economics rather than social

capital (Schneider *et al*., 2000) and others in which the incidence of trust and social capital indicators do not have a generalized effect on different economies (Peiró-Palomino and Tortosa-Ausina, 2013).

Therefore, the varied results of research in recent decades based on distinct statistical techniques, moving from descriptive statistics to Bayesian and non-linear regression analysis as in Forte *et al*. (2015) and Peiró-Palomino (2016), highlight the need for further investigation, encouraging a change in perspective.

Our idea is therefore to conduct an analysis of social capital on the local level, mixing objective and subjective measures. We analyse 20 European countries belonging to the OECD, collecting data for 194 European regions in 2014. We believe that the local dimension is central to the analysis of social capital because differences emerge not only between countries, but also and above all within countries. Moreover, regions of different countries could display similar performance and share common paths, and country-level analysis could mask these aspects.

The rest of the paper is organized as follows. Section 2 illustrates the adopted methodology and describes the dataset. Section 3 is devoted to the discussion of main results and Section 4 presents a conclusion.

2. Method and description of the data

We collect data Eurostat and OECD databases for 194 regions (NUTS2[1]) representing 20 countries that are both European and OECD countries in 2014[2]. We collect 12 variables, namely gross domestic product per capita (GDP), share of R&D total expenditure[3] (ERD), people with tertiary education and/or employed in science and technology (HC), education participation (ED), population growth (GPOP), economic activity rates (EAR), voter turnout (VOT), life expectancy (LFE), perceived social network support (SNS), perception of corruption (COR), unemployment rate (UNR), and life satisfaction (SAT).

[1] We use OECD territorial levels, which present some differences with the Eurostat nomenclature. See, https://www.oecd.org/regional/regional-statistics/.

[2] France is the sole exception since its GDP is from 2015 because Eurostat did not show regional values before this year.

[3] Territorial data of 2014 for Austria, Germany, Finland, and Sweden is obtained through the average between the 2013 and 2015 values; instead, for Greece, France, Slovenia, and the Netherlands, the national value is also assigned to the regions because these countries do not have local distinctions. At the end, Italy and United Kingdom have estimates made by OECD.

Table 1 – *Variables and descriptive statistics*

Variable	Tag	Description	Source	Minimum	Maximum	Mean
Gross Domestic Product per capita	GDP	Individual gross domestic product at current market prices by regions	Eurostat	6,694	90,643	27,655
Share of R&D total expenditure	ERD	Percentage of R&D total expenditure of regional GDP	OECD	0.180	4.980	1.652
HRST by category and regions	HC	People with tertiary education (ISCED) and/or employed in science and technology (percentage of active population)	Eurostat	22.40	65.40	40.65
Education participation	ED	Participation rate of group from 15 to 24 years in education	Eurostat	33.50	100.00	62.17
Population growth	GPOP	Growth of regional population in 2014	Eurostat	-1.3367	2.4156	0.2037
Economic activity rates	EAR	Economic activity percentage of people from 15 to 74 years	Eurostat	44.10	75.80	63.70
Voter turnout	VOT	Percentage voter turnout in general election	OECD	40.40	91.10	69.76
Life expectancy	LFE	Years of life expectancy at birth	OECD	74.20	84.80	80.86
Perceived social network support	SNS	Percentage of perceived social network support	OECD	68.80	100.00	90.99
Perception of corruption	COR	Percentage of perception of corruption	OECD	13.40	93.30	61.62
Unemployment rate	UNR	Unemployment rate in regions	OECD	2.20	34.80	10.59
Life satisfaction	SAR	Index of self-evaluation of life satisfaction	OECD	4.500	7.800	6.516

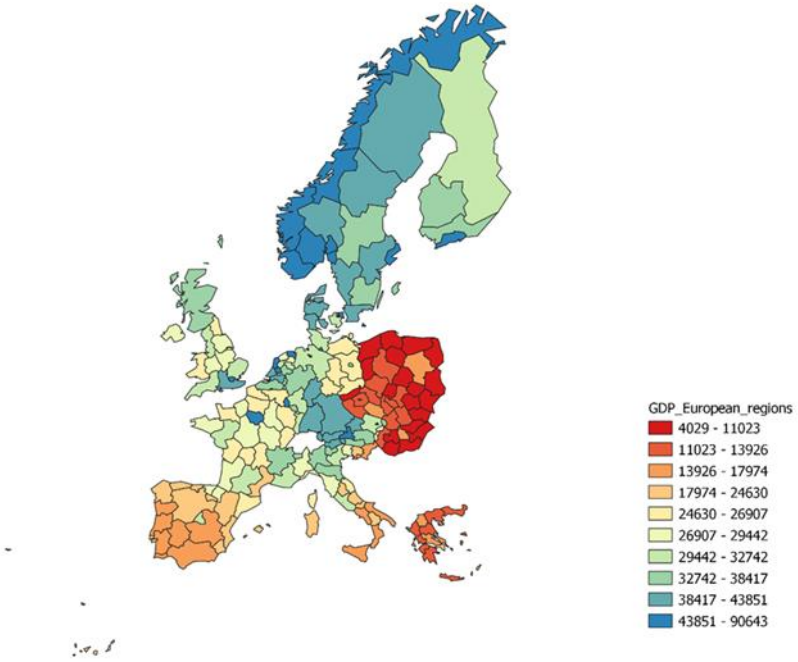
Source: Eurostat, OECD and preparation by the authors.

Table 1 reports the name of the variable, its acronym and a brief description, the source, and basic statistics (minimum, maximum and average value).

To obtain information about the role of social capital in economic growth, we initially calculate some basic descriptive statistics, focusing on GDP per capita. Figure 1 shows the GPD level for the regions analysed. Mediterranean regions (belonging to Southern Italy, Spain, and Portugal) display similar performance. The poorest countries are located in the Eastern and South-Eastern Europe. In contrast, the richer regions are found in Scandinavia.

From the overall database, what emerges is a significant heterogeneity in the various aspects in different parts of the continent. For instance, the perception of corruption (COR) registers a minimum value of 13.40 and a maximum of 93.30, covering practically the entire range of possible values.

Figure 1 – *Map of European regions based on GDP per capita levels*



Source: Prepared by the authors.

Thus, to investigate the relationships between economic variables, namely the gross domestic product per capita and the different features of social capital, we perform a multidimensional stepwise regression analysis. The model is defined in Equation 1:

$$GDP = \beta + \beta_1 ERD + \beta_2 HC + \beta_3 ED + \beta_4 GPOP + \beta_5 EAR + \beta_6 VOT + \beta_7 LFE + \beta_8 SNS + \beta_9 COR + \beta_{10} UNR + \beta_{11} SAT + \varepsilon \quad (1)$$

where β is the intercept, βi are the coefficients and ε is the statistical error. Looking at the selected variables, we expect that the perception of corruption and the unemployment rate are the only determinants with negative coefficients. We run a stepwise regression instead of a simple multiple regression analysis for two reasons. First, since the stepwise regression is an iterative procedure that removes predictor variables via a series of F-tests, that is, by testing for statistical significance after each iteration, the algorithm selects the best grouping of predictor variables that account for the greatest variance in the outcome (measured through R-squared). Consequently, the procedure will reduce the variance by estimating unnecessary terms. Second, since we have 11 independent variables, we can deal with multicollinear problems, and the output of stepwise regression will reduce this problem.

As a complementary step in the analysis and following the outcomes of the statistics, we perform a cluster analysis to highlight the discrepancies in the territories and to discover the similarities. We perform several cluster methods, both hierarchical and non-hierarchical cluster analysis, to find the most suitable method.[4] The next section is devoted to a discussion of the results.

To summarize the results of our analysis and conclude it, we create a composite indicator for social capital (SC) that summarizes all the information collected on social capital in a single number and also allows countries to be compared. For this step, we focus on seven variables that most affect social capital, namely 1) people with tertiary education and/or employed in science and technology (HC), 2) economic activity rates (EAR), 3) voter turnout (VOT), 4) life expectancy (LFE), 5) perceived social network support (SNS), 6) perception of corruption (COR), and 7) life satisfaction (SAT). Among the large range of composite measures, we adopt the Mazziotta and Pareto (2016) method (hereafter, MPI). MPI concerns the non-substitutability of the variables and the introduction of a type of penalty in determinate cases; this method also defines the normalization, using classical z-scores but with a mean equal to 100 and standard error equal to 10. The penalty is assigned after the arithmetic mean, considering the differences with the average (coefficient of variation), and the implications are the assignment of equal weights and the elimination of variability. This can be adapted for all cases because the penalty can be positive or negative (De Muro *et al*. 2011). Among the seven variables

[4] The results of the different methods implemented as well as a detailed description of the results of the cluster analysis are available upon request.

involved, only the perception of corruption has a negative polarity, meaning that an increase in corruption will lead to a drop in social capital. We account for this when we apply the aggregation method. Equation (2) reports the method:

$$SC_i = M_{z_i} \pm s_{z_i} \cdot cv_i \tag{2}$$

where SC_i denotes the value of the index for region *i* computed according to the MPI method, M_{z_i} and s_{z_i} represent the mean and standard deviation of standardized indicators[5], and cv_i is the coefficient of variation of the i-th unit, namely:

$$M_{z_i} = \frac{\sum_{j=1}^{m} z_{ij}}{m};\ s_{z_i} = \sqrt{\frac{\sum_{j=1}^{m}(z_{ij}-M_{z_i})^2}{m}};\ cv_i = \frac{s_{z_i}}{M_{z_i}} \tag{3}$$

This method focuses on horizontal variability, in which the regions with good performance in each factor are favoured and the normalization of z-scores is suitable for our single-year analysis.

3. Results

The results of the regression analysis demonstrate the important position held by factors constituting the social capital in determining the GDP per capita. Table 2 shows the results of the stepwise regression. From the initial model, the procedure suggests the elimination of two variables: participation in education and life satisfaction, neither of which are statistically significant in evaluating the impact on GDP per capita. The best predictors obtained, according to their relative levels, are HC, GPOP, VOT, LFE, SNS, COR, and UNR. The share of expenditure in research and development yields a value near significance.

As for the coefficients, corruption (-145.16) confirms the negative sign. This can be interpreted as a negative impact of this aspect of social capital on GDP. Similarly, unemployment rate (-245.42) reports the expected sign, confirming the initial hypothesis. The innovative results with respect to social capital are obtained in the coefficients of social network support and voter turnout. Indeed, the first has a low significant negative value (-302.53) while the second displays a positive incidence

[5] The standardized values are obtained by applying the usual z-score procedure and then rescaling the obtained values in order to obtain new values with mean equal to 100 and standard deviation equal to 10. See De Muro *et al*. (2011) for more details.

(202.61) with higher significance. The particularity lies in the fact that, as with civic engagement and distrust, the social network support should foster the development of cooperation and cohesion, which is useful for collaboration and the spread of innovation. It is also important to notice the coefficients of other variables in the model — life expectancy, population growth, and people with tertiary education or employed in science and technology — all of which have positive effects on economic growth. The multiple R-squared (about 0.81) and F-statistic (85.05) show that the regression has a good fit, and the entire model is statistically significant.

Table 2 – *Stepwise regression: dependent variable GDP per capita*

Variables	Estimates
Intercept	-94,722.48***
ERD	-1,279.74
HC	387.54***
GPOP	7,760.27***
EAR	176.74
VOT	202.61**
LFE	1,494.38***
SNS	-302.53*
COR	-145.16***
UNR	-245.42*
R-squared	0.8062
F-statistic	85.05

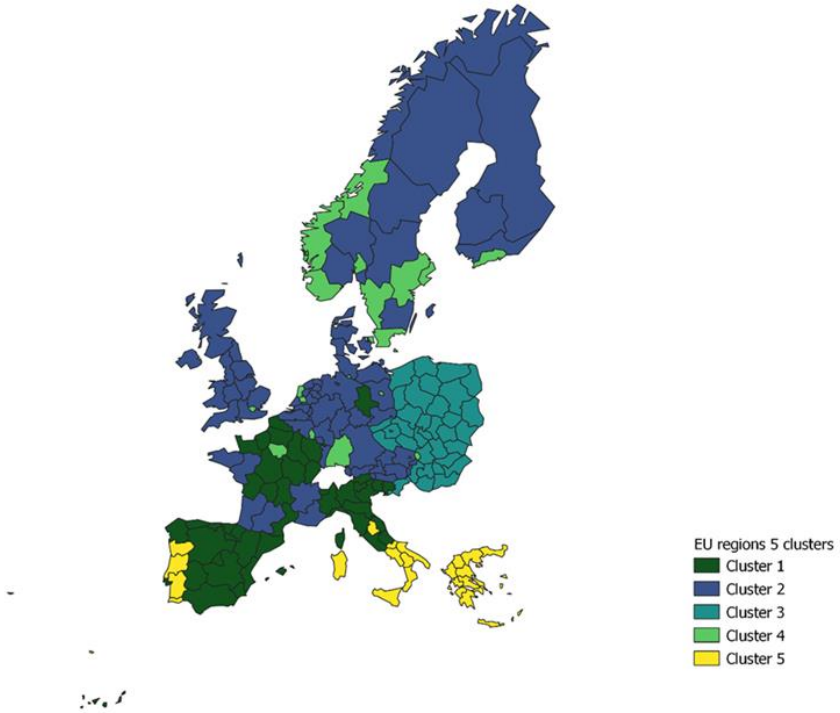
*Significance at *10%, **5% and ***1% levels.*

As second step in the analysis, we perform a cluster analysis to detect similarities in the regions according to the predictors selected. We implement both hierarchical and non-hierarchical methods, and, after comparing the results, we adopt the K-means method. We recall that the number of clusters must be chosen for the K-means method and, after computing statistical diagnostics, we fix K=5.[6] Figure 2 reports the results of the cluster analysis. Cluster 1 is composed of 47 regions in Central Europe (Figure 2 in dark green). Looking at the average values of variables of regions belonging to this cluster, we confirm that this cluster is made up of regions at an intermediate level in all indicators. Cluster 2 (blue) accounts for 64 developed regions in Northern and Central Europe. Cluster 3 (light blue) contains 35 regions,

[6] To be more precise, statistical tests suggested choosing K = 2. The result was a substantial division into regions of Central and Northern Europe versus the Mediterranean and Eastern European regions. However, this subdivision concealed all the differences. For this reason, we look at the resulting scree plot. Of course, as the number of clusters increases, the within-group sum of squares (variance) decreases, and we find a second peak at five clusters. This point represents the best balance between minimizing the number of clusters and minimizing the variance within each cluster.

mainly located in Eastern Europe, characterized by the lowest levels for all the indicators and, consequently, with the poorest performance in terms of social capital. Cluster 4 (green) is composed of 22 regions located in Western Europe and Southern Scandinavia and containing the capital cities (i.e. L'Île-de-France with Paris). This cluster is the richest group. Finally, Cluster 5 (yellow) contains 26 regions in Southern Europe, basically Portugal, Southern Italy, and Greece, summarizing more stagnant economies.

Figure 2 – *Map of groups according to the cluster analysis*



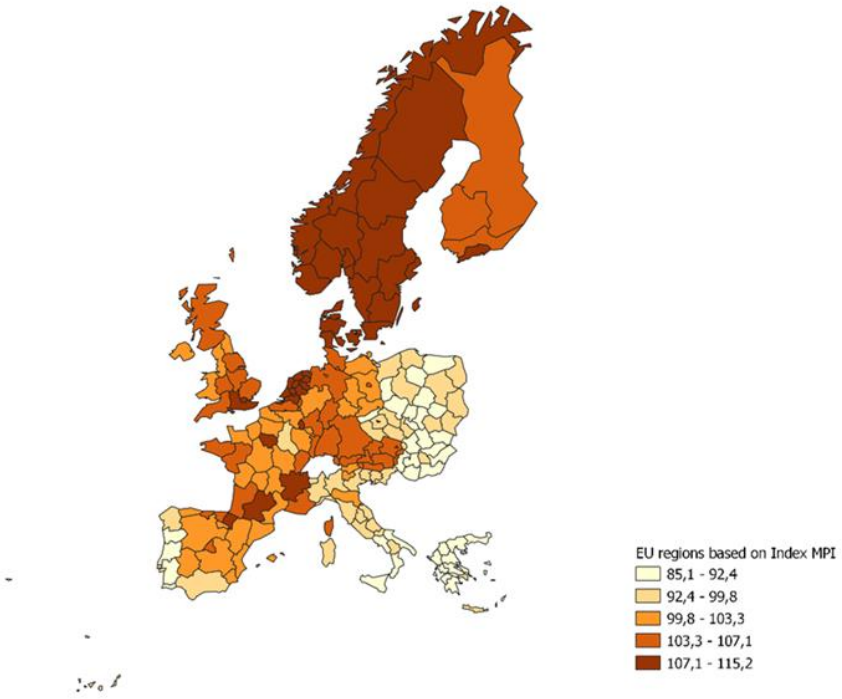
Source: Prepared by the authors.

As discussed in the Method Section, with the aim of expanding knowledge about social capital on European regions, we compute a composite index to summarize the seven variables related to social capital — namely HC, EAR, VOT, LFE, SNS, COR, and SAT — in a single number that simplifies comparisons.

Figure 3 shows the map of European regions based only the results of the social capital (SC) index. The map shows the values of SC per quantile; the darker the colour, the better the performance. What is evident is that the regions of Eastern Europe, with Greece, Portugal, and Southern Italy, have the lowest level of social

capital. On the contrary, the most developed in social capital are concentrated in Northern Europe. The map highlights interesting aspects: some regions of Central Europe (Spain and France) are in the highest positions of development of social capital while the United Kingdom is not at the same level as the other Northern countries as occurred for economic performance.

Figure 3 – *Map of regions according to SC index*



Source: Prepared by the authors

Another important result from the construction of the composite indicator is that regions belonging to the same country behave differently. We look at Spain as an example. This country exhibits at least three different levels of social capital. More specifically, regions in Northern Spain show a performance similar to Southern France or the South of Great Britain. Central regions are similar to Central France, whereas the Southern regions share a common path with Central Italy. Furthermore, Spain is just an example: France, Germany, Great Britain, just to name a few, have the same behaviour, displaying a huge variety in the values of the SC index. This means that an analysis made on a country level could hide various important differences, confirming the need for local analysis.

In our opinion, the formation of a regional social capital indicator that includes subjective information, for example, the perception of corruption, and objective measures, such as voter turnout in general elections, better represent the concept of Social Capital.

4. Conclusions

In the wider economic context, the relevance of social capital when analysing the performance of European regions is clear. Indeed, distrust, represented by the perception of corruption and civic commitment, have significant impacts on gross domestic product per capita.

Social capital has several facets, and, in a huge number of situations, it is difficult to measure them, basically due to the problem of uncertainty and the lack of available data, especially when moving from the country level to the regional scale. However, the result of the analysis concerning the construction of the composite indicator reveals the need to perform local analysis. In fact, we find that regions belonging to the same country behave differently, meaning that an analysis made for the country as a whole could hide several important aspects.

This work is a first attempt to analyse social capital by mixing several statistical methods. However, further work is necessary. First, more recent data must be obtained to evaluate the trend in composite indicators. More in general, an improvement in the available data specific to measuring social capital in terms of space and time is necessary. Second, to improve the analysis, the introduction of spatial components could add more appeal to the analysis. Finally, a robustness check on the construction of the composite indicator could round out the analysis.

References

BEUGELSDIJK S., DE GROOT H. L., VAN SCHAIK A. B. 2004. Trust and economic growth: a robustness analysis, *Oxford Economic Papers*, Vol. 56, No. 1, pp. 118–134.

BJØRNSKOV C. 2006. The multiple facets of social capital, *European Journal of Political Economy*, Vol. 22, No. 1, pp. 22–40.

COLEMAN J. 1988. Social capital in the creation of human capital, *American Journal of Sociology*, Vol. 94, pp. 95–120.

DE MURO P., MAZZIOTTA M., Pareto A. 2011. Composite indices of development and poverty: An application to MDGs, *Social Indicators Research*, Vol. 104, pp. 1–18.

FORTE A., PEIRÓ-PALOMINO J., TORTOSA-AUSINA E. 2015. Does social capital matter for European regional growth?, *European Economic Review*, Vol. 77, pp. 47–64.

KNACK S., KEEFER P. 1997. Does social capital have an economic payoff? A cross country investigation, *Quarterly Journal of Economics*, Vol. 112, No. 4, pp. 1251–1288.

MAZZIOTTA M., PARETO A. 2016. On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena, *Social Indicators Research*, Vol. 127, pp. 983–1003.

PEIRÓ-PALOMINO J. 2016. Social Capital and Economic Growth in Europe: Nonlinear Trends and Heterogeneous Regional Effects, *Oxford Bulletin of Economics and Statistics*, Vol. 78, No. 5, pp. 717-751.

PEIRÓ-PALOMINO J., TORTOSA-AUSINA E. 2013. Can trust effects on development be generalized? A response by quantile, *European Journal of Political Economy*, Vol. 32, pp. 377–390.

POTEYEVA M. 2009. Sentencing disparities, native Americans. In H. T.

PUTNAM R. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton: Princeton University Press.

SCHNEIDER G., PLUMPER T., BAUMANN S. 2000. Bringing Putnam to the European regions: on the relevance of social capital for economic growth, *European Urban and Regional Studies*, Vol. 7, No. 4, pp. 307–317.

ZAK P. J., KNACK S. 2001. Trust and growth, *The Economic Journal*, Vol. 111, No. 470, pp. 295–321.

SUMMARY

The role of social capital in economic performance across European regions

In recent decades, there has been increasing interest in analysing social capital. However, the analysis is commonly conducted on the country scale, hiding the regional dimension and, consequently, neglecting regional disparities.

In this work, we investigate the diffusion of social capital across European regions, defining the different aspects through objective measurements as well as personal opinions and values. At the same time, the analysis also focuses on the effects that social capital has on economic performance, especially on gross domestic product per capita, underlining the relevant impacts. At the end, in order to highlight the stock of social capital in the European Union, particularly in 194 regions in 20 European countries in 2014, a composite index is developed to better evaluate the spread. More in detail, the variables defining social capital are summarized using the Mazziotta-Pareto method.

The results of the analysis show the positive impact of social capital on economic growth, mainly due to the development of trust, collaboration, and cohesion, which is also decisive for innovation. Moreover, the study exhibits remarkable differences among regions, even within countries. This suggests that region-specific policies could be more effective than country-specific policies.

Mariateresa CIOMMI, Università Politecnica delle Marche, m.ciommi@univpm.it
Francesca MARIANI, Università Politecnica delle Marche, f.mariani@univpm.it
Maria Cristina RECCHIONI, Università Politecnica delle Marche, m.c.recchioni@univpm.it
Riccardo ZALLOCCO, Università Politecnica delle Marche, s1091643@studenti.univpm.it

TRADE SURVEYS: QUALITATIVE AND QUANTITATIVE INDICATORS[1]

Maria Rita Ippoliti, Luigi Martone, Fabiana Sartor

1. Introduction

Qualitative business surveys on enterprises first appeared in the 1970s, within a harmonised project of the European Commission, in order to track the economic evolution and the short-term dynamics of the economic activity promptly.

The survey questionnaire asks enterprises to express their prospect of the business evolution and their evaluation, based upon opinion surveys about developments of business variables. Question types are mainly qualitative and closed-ended (i.e. "high", "low", "increasing", "decreasing", "steady", "above standard", "standard", "below standard").

Collecting these data allows measuring the economic evolution, by recording the different types of responses given by the interviewees over time (UNECE, 2015; European Commission, 2017).

Qualitative surveys are extremely relevant for the short-term economic analysis and exceptionally useful in the building process of cyclical indicators of the economic development as they gather information quickly and do not involve revisions.

This approach underlines the important role of assessments and expectations for the economic decision makers.

Each response is converted into a weighted balance, given by the difference between positive and negative responses, ignoring the neutral answers[2].

[1] Though the article is the result of a joint work, the single paragraphs are attributed as follows: paragraph 1 to Fabiana Sartor; paragraph 2 to Maria Rita Ippoliti; paragraph 3 to Luigi Martone. The published articles are exclusively expressing the authors' opinions; Istat shares no responsibility for the published contents.

[2] Weighted balances can vary from -100, when all the respondents choose the negative answer, to +100 when all the respondents choose the positive answer. Thus, qualitative series typically do not detect the long-term trend.

Weighted balances are used to calculate the business confidence indicators, using arithmetic means of variable sets, which are relevant for the short-term economic evolution of a specific economic sector[3].

As qualitative series are so important for policy makers, it is necessary to assess the reliability of results of the qualitative surveys, measuring the closeness with official quantitative series and estimating the similarity of signals coming from qualitative surveys and short-term movement in the economic activity (Koopmans, 1947; Zarnowitz, 1992). Therefore, qualitative variables are often considered to be complementary to official macroeconomic data and they are both used for analysing the short-term economic development (Bergstrom 1995; Bruno e Lupi, 2004; Croux *et al.* 2005; Claveria *et al.* 2007; Bruno, 2009; Cesaroni, 2011; Cesaroni *et. al.* 2011; Conti e Rondinelli, 2015; Girardi *et al.* 2016). Several studies have also documented the business cycle analysis through a qualitative approach (Altissimo *et al.* 2000; Bruno and Otranto 2008; McNabb and Taylor 2007; Cesaroni and Iezzi 2017; Bruno *et al.* 2019).

There has been much debate on this topic in the economic literature, mainly with regard to the manufacturing sector, while the discussion remained poor concerning the other economic sectors, such as services, trade and construction (Crosilla and Leproux, 2007; Crosilla *et al.* 2009; Martelli and Rocchetti, 2007).

This paper investigates the trade sector, comparing the evolution of business confidence in retail trade with two quantitative indicators, which measure the different sectors of section G: retail trade (Division 47) and turnover in services (Division 45) from 2010 until 2019. Business confidence in retail trade involves enterprises classified in the NACE Rev. 2[4] section G, apart from the wholesale trade classified in Division 46, the itinerant retailers classified in Group 47.8 and non-store retailers, stalls and markets classified in the 47.9.

The aim of the project is to assess the relation between qualitative and quantitative components concerning the retail trade sector: analysing the trend in time series and testing the forecasting ability of the qualitative index on quantitative data.

This paper is structured as follows: paragraph 1 presents the background to the study, it introduces and illustrates the qualitative and quantitative surveys involved. Paragraph 2 provides a description of methods and tools used, it presents a first graphic analysis comparing the quantitative and qualitative indicators and it introduces a new "ad hoc" indicator covering all sectors involved in the business confidence. Finally, paragraph 3 presents our conclusions and it shows that among

[3] See Moore and Shiskin (1967).

[4] Section G includes Division 45 (Wholesales trade, retail trade and maintenance of motor vehicles), Division 46 (Wholesales trade, except wholesales trade of motor vehicles) and Division 47 (Wholesales trade, except wholesales trade of motor vehicles).

the quantitative indices, the "ad hoc" index returns the highest correlation with its equivalent qualitative indicator.

1.1. Business Confidence Survey in Retail Trade

Business Confidence Survey in Retail Trade is part of a joint project harmonised at European level and coordinated by the European Commission[5]. It allows to have information on the economic evolution of retail trade (NACE Division G, except for Division 46 - Wholesale trade, except of motor vehicles and motorcycles and for Group 47.9 - Retail trade not in stores, stalls or markets including retail sales via mail order or via Internet).[6] The survey asks enterprises to express their opinions (judgements and expectations over the next 3 months) about the main economic variables (orders, prices, employment), giving therefore an updated overview on the evolution of the sector. Respondents are requested to state their consideration on their total sales in the last three months, on their current volume of stock and on prices charged by their suppliers. They are also invited to express how they expect the volume of orders, the employment, the prices they charge and total sales to change in the next three months. Information about enterprises of the Business Confidence Survey in Retail Trade are taken from a panel of approximately 1,000 commercial enterprises. The theoretical sample is stratified by enterprise employment size class (1-2 employees, 3-5; 6-999; at least 1,000 employees), by geographical area (North-West, North-East, Centre, South and the Islands) and by main activity (45.1 sales of motor vehicles; 45.2-45.4 maintenance of motor vehicles and sales of accessories; 47.1, 47.2 retail sales of food, drinks and tobacco; 47.3 retail sales of automotive fuel; 47.4-47.7 retail sales of other goods). The sampling scheme depends upon randomisation for enterprises with less than 1.000 employees and upon a census placement for all units with 1,000 employees and above. The data processing method sets out the estimate of the frequency percentages of each reply option relating to each question of the questionnaire. For this purpose, the processing of the micro data is based on a double weighting system: a) the frequencies of each reply option are firstly weighted using the number of employees declared by the enterprise at the time of the interview (internal weight); b) subsequently fixed weights reflecting the distribution of the added value of the reference sector (external

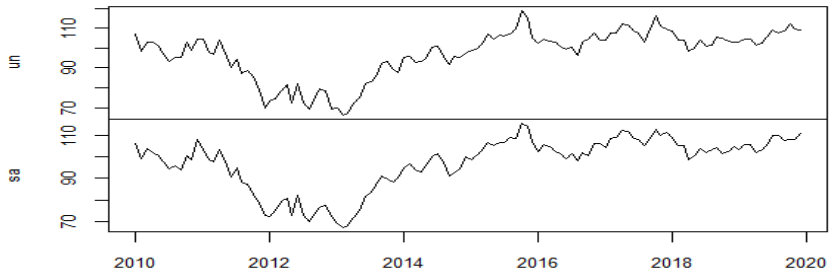
[5] European Commission. 2016. The Joint Harmonised EU Programme of Business and Consumer Surveys.

[6] Divisions of NACE Section G involved in the survey are Division 45 (Wholesale and retail trade and repair of motor vehicles and motorcycles) and Division 47 (Retail trade, except of motor vehicles and motorcycles).

weight) are used. Since March 2015, the aggregation procedure uses an external weighting structure derived from the added value at factor cost referred to 2012.

The index of business confidence in retail trade is calculated as the arithmetic mean of seasonally adjusted balances based upon opinions and expectations on sales and upon judgments on volume of stocks (the above-mentioned values have inverse signs). Weighted balances are equal to the difference between favourable and unfavourable responses given for each observed variable. Figure 1 shows the monthly evolution of the business confidence in retail trade, unadjusted and seasonally adjusted for years 2010-2019.[7].

Figure 1 – *Retail trade confidence (monthly, unadjusted and seasonally adjusted data, years 2010-2019).*



Source: Elaboration based on ISTAT data.

1.2. Retail Trade Survey

Monthly Retail Trade Survey collects data from enterprises that mainly operate in the retail trade sector (except for sale of motor vehicles and motorcycles and sale of automotive fuel). Therefore, the survey covers the retail sales sector only partially (NACE Rev. 2, G 47 - Retail trade, except of motor vehicles and motorcycles not including automotive fuel). Estimates of retail trade survey provide useful information on consumer spending.

Monthly indices on retail trade are released at national level, consistently with the European Union Regulations concerning short-term statistics (see European Regulations n. 1165/98 and n. 1158/2005)[8].

Retail trade data are collected from a sample of about 8,000 enterprises having at least a legal unit that operates in Italy. The sample is stratified considering the

[7] Seasonally adjusted data are available at http://dati.istat.it/en starting from January 2003.
[8] See https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1998R1165:20120621:EN:PDF

following variables: main activity according to NACE Rev. 2 and employment size class (1-5, 6-49 and at least 50 persons employed). According to the sampling scheme, enterprises with less than 50 persons employed are selected at random, while any enterprise with equal or more than 50 persons employed is included in the survey. All the enterprises employing 50 or more persons add up to more than 1,.100 units. The sampling design of the survey rotates some units out and rotates new units in each year (belonging to employment size classes 1-5 and 6-49 only) to share burden and refresh the sample. Within the weighting structure of monthly retail trade index (base=2015), large scale-distribution accounts for 46.4% of total turnover, while small-scale distribution reaches 48.0% of total turnover.

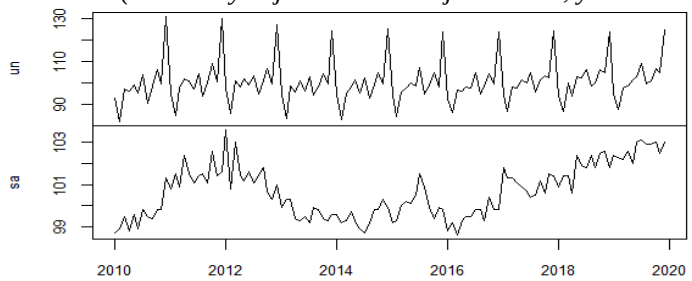
Retail trade indices are calculated as weighted means of the sub-indices of each stratum. To calculate aggregate indices up to the retail trade total, the Laspeyres index is used. The weights are based on turnover data from SBS of the year 2015.

Value of sales indices measures the retail trade turnover over time at current prices and, therefore, incorporates the effects in changes of quantity sold and prices. In order to determine estimates on the volume of sales, value of sales indices are processed to allow removing price effects on turnover, using the Harmonised index of consumer prices (HICP).

Monthly data are first revised in the following month after publication (which occurs 38 days past the reference period). Estimates are then subject to a second revision, which occurs on annual basis and replaces the provisional estimates with the final indices.

Figure 2 shows the monthly evolution of the retail trade index, years 2010-2019[9].

Figure 2 – *Retail trade (seasonally adjusted and unadjusted data, years 2010 – 2019).*



Source: Elaboration on Istat data. Retail Trade Survey

[9] Seasonally adjusted data on retail trade are available at http://dati.istat.it/en starting from January 2003

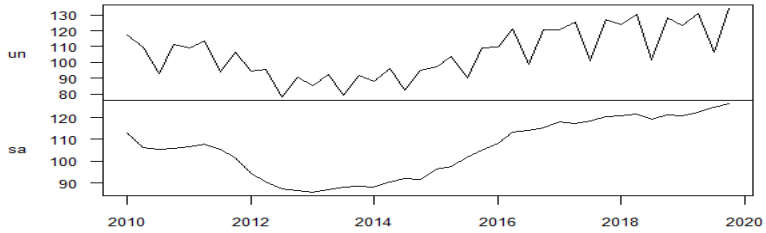
1.3. Turnover in services survey

Quarterly Turnover in Services Survey focuses on short-term dynamics of value of services sold by enterprises, which operate in the services field as a main economic activity.

These dynamics incorporate growth rates of volumes and prices, giving a prompt information. The methodology, the breakdown and the frequency of releases are defined by the European Regulations on short-term statistics (see European Regulations n. 1165/98 and n. 1158/2005 and footnote 9). This is a sample survey, enterprises are the units of observation, while turnover and average number of persons employed by the economic unit in the reference quarter are the variables of interest. Quarterly indicators on turnover in services are calculated for each economic sector, setting 2015 as the base year; these indices are then aggregated according to the Laspeyres formula that uses a weighting structure reflecting the proportion of turnover by economic sector in the base year 2015. The survey does not include retail trade, however it includes retail trade of motor vehicles and motorcycles (NACE Rev. 2 Sections from G to N apart from retail trade as mentioned before)[10]. The sample for this survey includes 1.627 enterprises, starting from 2010.

Figure 3 shows the monthly evolution of the quarterly turnover in services index for Division 45 (seasonally adjusted, base 2015=100), from April 2010 until December 2019[11].

Figure 3 – *Turnover in services – Ateco Division 45 (quarterly data, seasonally adjusted and unadjusted, years 2010-2019).*



Source: Elaboration on Istat data. Turnover in Services Survey.

[10] Economic sectors covered by the survey are represented by the following Sections: G - Wholesale and retail trade; repair of motor vehicles and motorcycles (except of G 47 Retail trade), H - Transporting and I –Accommodation and food service activities, J – Information and communication, L – Real estate activities, M – Professional, scientific and technical activities, N – Administrative and support service activities.

[11] Seasonally adjusted data on turnover in services Division 45 are available at http://dati.istat.it/en starting from 2010.

2. Data analysis: tools and methods

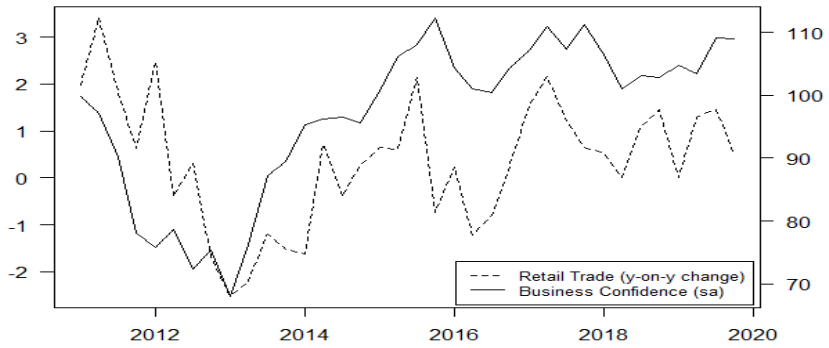
Generally speaking, comparing qualitative and quantitative series can be challenging as quantitative indices indicate quantities expressed in value or volume, while qualitative indices use an ordinal scale of measurement. The balance between the percentage of positive and negative answers can be presented as a diffusion index, therefore the business confidence index can be seen as a diffusion index capturing the movement of the different components. This common component is not necessarily connected to the trend nor to the seasonal component (qualitative data do not have a trend component by definition) and can be interpreted as the cyclical component[12]. Comparing graphics of data from Business Confidence Survey and the other two quantitative surveys was the first step of our analysis. Since business confidence is a cyclical index, it must be properly transformed first in order to be compared to a quantitative index. Concerning our analysis, we applied the year-on-year change, a simple method often used to eliminate the trend and the seasonal component.

2.1. A first graphic analysis

Our first analysis displays the relationship between business confidence in retail trade, covering part of NACE Rev. 2 Division 47 (from 47.1 to 47.7 only), and retail trade index, covering the whole Division 47 (data here considered are those forwarded to Eurostat including NACE 47.3 fuel sector, while at national level Istat disseminates data concerning NACE G 47 excluding Group 47.3): the graph (fig. 4) shows that the two indices have a similar pattern and appear to have a good closeness. The following graphic analysis was focused on comparing the business confidence in retail trade and the quantitative index representing Division 45 of the turnover in services. Fig. 5 shows the relationship between the indices of business confidence in retail trade and turnover in services: in this case, they seem to be closer, although distance between them starts to increase since 2017. Dissimilarity between the two indices can be caused by the different sectoral coverage, by the different sampling scheme of the three surveys and by the different features of the seasonal component, which is higher in the retail trade index.

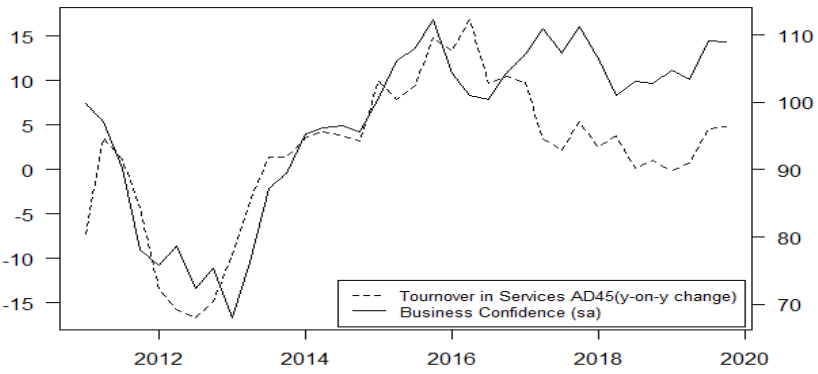
[12] An observed time series can be decomposed into three components: the trend (long term direction that may also be best described by a curved model like a polynomial in *t*); the seasonal component (systematic, calendar related movements); the cyclical component (when data exhibit rises and falls that are not of fixed period and the average length of cycles is between 2 and 10 years) and the irregular component (unsystematic, short term fluctuations, it's the residual time series component after the other components have been removed).

Figure 4 – *Comparison between Business confidence (seasonally adjusted) and Retail trade (year-on-year change) – quarterly data, year 2011 – 2019.*



Source: Elaboration based on ISTAT data. Business Confidence Survey in Retail Trade and Monthly Retail Trade Survey.

Figure 5 – *Comparison between Business Confidence (seasonally adjusted) and Turnover in services (year-on-year change) - quarterly data, year 2011 – 2019.*



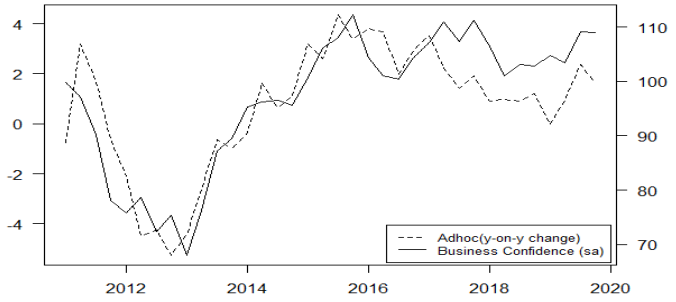
Source: Elaboration based on ISTAT data.

2.2. The new quantitative indicator: an "ad hoc" index

The different sectoral coverage made it necessary to create an ad hoc index involving both the surveyed sectors (Divisions 45 and 47), calculated as a weighted average of the two indices (retail trade index and turnover in services index) where weights reflect the distribution of turnover across the different economic sectors.

Figure 6 shows the new index has a closer relationship and the gap seem to be reduced.

Figure 6 – *Comparison between Business Confidence (seasonal adjusted) and the "ad hoc" index (year-on-year change) - quarterly data, years 2011-2019.*



Source: Elaboration on ISTAT data.

In order to verify previous results, we calculated the correlation between business confidence in retail trade and the three considered quantitative indices.

The values in table 1 show that the correlation between business confidence and the new ad hoc index is higher (0.85) than the correlation between the indicators of each quantitative survey and Business Confidence Survey (Retail trade 0.41 and Turnover in services -Div. 45 0.78).

Table 1 – *Correlation index between qualitative data and quantitative data.*

Confidence – Retail Trade	Confidence – Turn. in Serv. (Div 45)	Confidence – 'Ad hoc' index
0.41	0.78	0.85

Source: Elaboration on ISTAT data.

3. Conclusions and future perspective

Business confidence surveys provide timely information on the short-term dynamics of the economic activity; qualitative surveys are extremely relevant for the short-term economic analysis and exceptionally useful in the building process of cyclical indicators of the economic development as they gather information quickly and do not involve revisions. Moreover, these variables allow anticipating turning points in the economic activity and exploring issues often ignored by quantitative indicators, producing crucial information for policy-makers. There has been much debate on this topic in the economic literature, mainly with regard to the manufacturing sector, while the discussion remained poor concerning the retail trade. This paper investigates the trade sector from 2010 until 2019, comparing the

evolution of business confidence in retail trade with two quantitative indicators, measuring the different sectors of NACE Section G: Retail trade (Division 47) and Turnover in services concerning Division 45 only (Wholesale trade). The graphic analysis shows that the two indices appear to have a good closeness even though they follow different routes towards the end of the reference period. The different sectoral coverage made it necessary to create an ad hoc index involving both the surveyed sectors (Divisions 45 and 47), calculated as a weighted average of the two indices where weights reflect the distribution of turnover across the different economic sectors. This analysis shows that the new index has a closer relationship and the big gap towards the end of the reference period appears to be reduced, also the correlation between business confidence and the new ad hoc index is higher than the correlation between the indicators of each quantitative survey and Business Confidence Survey.

The method we used could also be applied to investigate the association between business confidence index concerning NACE Division G 47 and 45 and the quantitative indicators relating to the same sectors. In addition, further studies could use the most suitable transformations of quantitative indicators (for instance the cyclical component or the seasonal logarithmic difference) to identify the turning points and the expansion and recession phases of the economic cycle. Future works could also verify whether the turning points of the qualitative series of business confidence tend to predict the future movement, to be coincident or to follow the fluctuation in the reference series (therefore additional studies could help classify these indices as leading, coincident or lagging indicators as compared to quantitative series).

References

ALTISSIMO F., MARCHETTI D.J., ONETO G.P. 2000. The Italian business cycle: Coincident and Leading indicators and some stylized facts. *Giornale degli Economisti e Annali di Economia.* Vol. 60, No. 2, pp. 147-220.

BERGSTROM R. 1995. The relationship between manufacturing production and different business survey series in Sweden 1968-1992, *International Journal of Forecasting,* Vol. 11, No. 3, pp. 379-393.

BRUNO G. 2009. Non-linear relation between industrial production and business survey data, *ISAE working papers 119*, Rome.

BRUNO G., LUPI C. 2004. Forecasting industrial production and the early detection of turning points, *Empirical Economics, Springer*, Vol. 29, No. 3, pp. 647-671.

BRUNO G., OTRANTO E. 2008. Models to date the business cycle: The Italian case, *Economic Modelling,* Vol. 25, No. 5, pp. 899-911.

BRUNO G., CROSILLA L. e MARGANI P. 2019. Inspecting the Relationship Between Business Confidence and Industrial Production: Evidence on Italian Survey Data, *Journal of Business Cycle Research*, Vol. 15, No. 1, pp. 1-24.

CESARONI T. 2011. The cyclical behavior of the Italian business survey data, *Empirical Economics*, Vol. 41, No. 3, pp. 747-768.

CESARONI T., MACCINI L., MALGARINI M. 2011. Business cycle stylized facts and inventories behavior: New evidence for the euro area, *International Journal of Production Economics*, Vol. 133, No. 1, pp. 12-24.

CESARONI T., IEZZI S. 2017. The Predictive Content of Business Survey Indicators: Evidence from SIGE, *Journal of Business Cycle Research*, Vol. 13, No.1, pp. 75-104.

CLAVERIA O., PONS E., RAMOS R. 2007. Business and consumer expectations and macroeconomic forecasts, *International Journal of Forecasting*, Vol. 23, No.1, pp, 47-69.

CONTI A.M., RONDINELLI C. 2015. Easier said than done: the divergence between soft and hard data, *Bank of Italy Occasional papers n. 258*, Roma.

CROSILLA L., LEPROUX S. 2007. Leading Indicators on construction and retail trade sectors based on ISAE survey data, *ISAE working papers 77*, Roma.

CROSILLA L., LEPROUX S., MALGARINI M. e SPINELLI F. 2009. Model-based Confidence Indicators for the Italian Economy, *ISAE working papers 116*, Roma.

CROUX C., DEKIMPE M. G., LEMMENS A. 2005. On the predictive content of production surveys: A pan-European study, *International Journal of Forecasting*, Vol. 21, No. 2, pp. 363-375.

EUROPEAN COMMISSION 2016. The Joint Harmonised EU Programme of Business and Consumer Surveys.

GIRARDI A., GAYER C. e REUTER A. 2016. The Role of Survey Data in Nowcasting Euro Area GDP Growth, *Journal of Forecasting*, Vol. 35, No. 5, pp. 400-418.

KOOPMANS T. 1947. Measurement without theory. *Review of Economic Statistics* Vol. 29, No. 3, pp. 161–72.

MARTELLI B.M. e ROCCHETTI G. 2007. Cyclical features of the ISAE business services series, *ISAE working papers 89*, Roma.

McNABB R., TAYLOR K. 2003. Business cycle and the Role of Confidence: Evidence for Europe, *Oxford Bulletin of Economics and Statistics*, Vol. 69, No. 2, pp. 185-208.

MOORE G.H. e SHISKIN J. 1967. Indicator of business expansion and contractions, *NBER Occasional Paper*, No.103.

UNECE. 2015. Handbook of Economic Tendency Surveys, Economic e Social Affairs, *Statistical papers*, Series M, No. 96, United Nations, New York.

ZARNOWITZ V. 1992. *Business Cycles: Theory, History, Indicators, and Forecasting*. Chicago: The University of Chicago Press.

SUMMARY

The surveys on retail trade: comparing qualitative and quantitative indicators

Qualitative survey data, being designed to provide early information which are not subject to subsequent revisions, are extremely useful among the short-term statistics. Hence, qualitative and macroeconomic time series are often considered to be complementary, therefore both approaches are used to analyse the short-term economic evolution. This paper compares the retail trade confidence index (NACE Rev. 2, G 45 and G 47), a business sentiment indicator, along with quantitative data from retail trade statistics (NACE Rev. 2 G 47) and turnover in services data (NACE Rev. 2 G 45) from 2010 to 2019.

First step was merging seasonally adjusted data on retail trade confidence, year-over-year percentage change indices on retail trade and turnover in services indices (division 45 only). Due to lack of accuracy in comparing the above-mentioned series, an "ad hoc" index has been produced, including both economic activities concerning quantitative surveys. The new index has been calculated as a weighted average where weights reflect the distribution of turnover across the different economic sectors. In order to test the reliability of the study, this paper proceeds with calculating a correlation between the two indicators.

Maria Rita IPPOLITI, Istat, marippoliti@istat.it
Luigi MARTONE, Istat, martone@istat.it
Fabiana SARTOR, Istat, sartor@istat.it

DATA NORMALIZATION FOR AGGREGATING TIME SERIES: THE CONSTRAINED MIN-MAX METHOD[1]

Matteo Mazziotta, Adriano Pareto

1. Introduction

In recent years, composite indices have been increasingly adopted by many institutions for providing a measurement of multidimensional phenomena, such as well-being, development, poverty and so on. Examples of well-known composite indices are the Human Development Index (HDI) created by the United Nations Development Programme (UNDP, 1990; 2010) and the Canadian Index of Well-being (CIW) produced by the University of Waterloo (Michalos *et al.*, 2011).

In both cases, a set of time series of individual indicators with different units of measurement and ranges (e.g., "Life expectancy at birth" and "Gross national income per capita") are aggregated into a single composite index for various geographical areas. This procedure involves several problems. In particular, it is necessary to normalize the data to make them comparable and the method used must be time-independent, so as not to change the past data every time a new observation is added.

In the HDI individual indicators are converted to a common scale with range [0, 1] by the Min-Max method, whereas in the CIW individual indicators are converted to a common scale where the 1994 value (base value) is set to 100 by 'indicization' (i.e., transformation in index numbers).

In this paper, we show that both these normalization methods have some weaknesses and we use an alternative method, the "Constrained Min-Max Method", that combines the strengths of the two methods, without having their limitations. The method is a generalization of the normalization formula used in the Adjusted Mazziotta-Pareto Index (Mazziotta and Pareto, 2016). An empirical comparison among the three normalization methods is also presented, where the time series of two well-being indicators of the CIW are normalized and then aggregated.

[1] The paper is the result of the common work of the authors: in particular M. Mazziotta has written Sections 1-2 and A. Pareto has written Sections 3-5.

2. The traditional methods

Let x_{ij}^t be the value of individual indicator *j*, for unit *i*, at time *t* (*j*=1, …, *m*; *i*=1, …, *n*; t=1, ..., *k*). We want to build, for each unit, a composite index CI_i^t that summarizes the trend of the individual indicators over time. If individual indicators have different measurement units, the data must be normalized in order to make them comparable both across units and over time.

The first solution that might come to mind is to transform the data in *z*-scores by the classical standardization. However, if a standardization is done over time, adding a new observation would change the mean and variance of the time series and then the data would have to be standardized every time. Furthermore, the data concerning different units would not be easily comparable since the mean of the time series is a reference difficult to interpret. For this reason, the Min-Max method or indicization are usually used.

The Min-Max method is most used by sociologists. It normalizes indicators to have an identical range [0, 1]. For a generic unit *i* and indicator *j* at time *t*, the normalized value is:

$$y_{ij}^t = \frac{x_{ij}^t - \min_{it}(x_{ij}^t)}{\max_{it}(x_{ij}^t) - \min_{it}(x_{ij}^t)}$$

where $\min_{it}(x_{ij}^t)$ and $\max_{it}(x_{ij}^t)$ are, respectively, a minimum and a maximum that represent the possible range of indicator *j* (*goalposts*). They can be calculated across all the units over time or can be fixed by the researcher. However, if new data exceed the selected range, the normalization parameters should be updated in order to avoid values out of the range [0, 1], and normalized values would have to be recalculated for all series.

The Min-Max method normalizes the range of indicators, but it does not 'centre' them with respect to a base value and this leads to the loss of a common reference, such as the mean (Mazziotta and Pareto, 2021). Indeed, the normalized value 0.5 is the mean of the range, but not of the distributions, and then it cannot be used as a reference for reading results (e.g., if the normalized value of a given unit is 0.3., we cannot know if its original value is above or below the mean).

Indicization[2] is the method most used by economist. It measures the relative position of a given value from a reference (Tarantola, 2008). For a generic unit *i* and indicator *j* at time *t*, the normalized value (also called index number) is:

[2] This method is also known as 'Distance from a reference' (OECD, 2008).

$$y_{ij}^t = \frac{x_{ij}^t}{x_{oj}} 100$$

where x_{oj} is the reference (or base) value for indicator *j* (e.g., a mean or the value for a given year).

Indicization 'centre' normalized indicators with respect to a common reference (set equal to 100), but it does not normalize their range, because they have the same CV[3] of original indicators. Therefore, it introduces implicit weights which can affect the aggregation (Mazziotta and Pareto, 2020). The wider the minimum and maximum values are apart, the higher the implicit weighting and vice versa (Booysen, 2002). Therefore, if indicators are normalized by indicization with base 100, but a normalized indicator ranges between 99 and 101 and other ranges between 50 and 200, the composite index will be dominated by the second indicator.

3. The constrained Min-Max method

The Constrained Min-Max method is an alternative method that 'normalizes' indicators – similarly to the Min-Max method – but uses a common reference that allows to 'centre' them – like indicization. It converts indicators to a common scale where a reference is set equal to 0 e the range is 1. For a generic unit *i* and indicator *j* at time *t*, the normalized value is:

$$y_{ij}^t = \frac{x_{ij}^t - x_{oj}}{\max_{it}(x_{ij}^t) - \min_{it}(x_{ij}^t)} \tag{1}$$

where $\min_{it}(x_{ij}^t)$ and $\max_{it}(x_{ij}^t)$ are, respectively, a minimum and a maximum that represent the possible range of indicator *j* (*goalposts*) and x_{oj} is the reference value for indicator *j*.

Normalized indicators have the same reference (e.g., the value for a given year) and equal range. This allows to have the advantages of indicization (normalized values are easier to interpret) without introducing implicit weights. Moreover, transformed scores may be further adjusted if calculations yield awkward values. Finally, if new data exceed the selected range, the comparability across time is

[3] The coefficient of variation (CV) is a measure of dispersion, often expressed as a percentage, defined as the ratio between standard deviation and mean.

maintained and the parameters of formula (1) do not need to be updated.

4. An empirical comparison

As is known, the CIW is a composite index calculated annually, composed of eight domains that measures change in the wellbeing of Canadians over time (Michalos *et al*., 2011). Let us consider the following two indicators taken from the "Education" domain of the CIW for the 1994 to 2008[4] period:

- "Ratio of childcare spaces to children aged 0 to 5 years of age" (X_1) with a mean of 15.4 and a standard deviation of 2.72 (CV of 17.7%);
- "Average of 5 social and emotional competence scores for 12 to 13 year olds" (X_2) with a mean of 3.2 and a standard deviation of 0.05 (CV of 1.5%).

We chose these two indicators because they have opposite trends over time and very different CVs.

Table 1 – *Comparing normalization methods.*

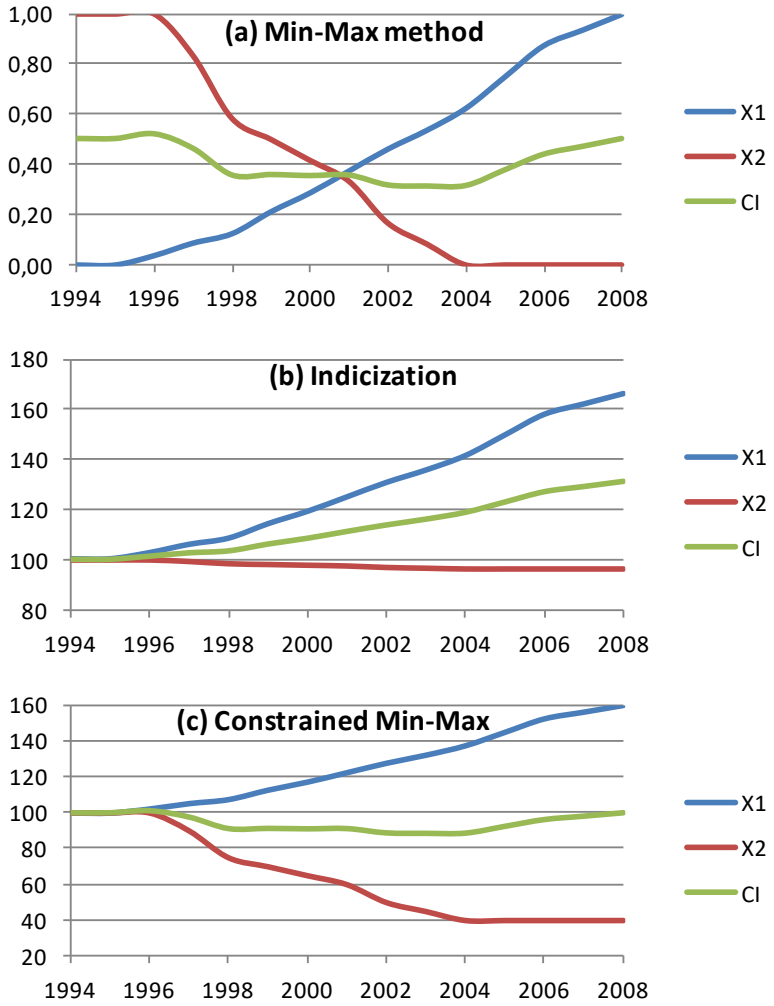
Year	Original data		(a) Min-Max method			(b) Indicization			(c) Constrained Min-Max		
	X_1	X_2	X_1	X_2	CI	X_1	X_2	CI	X_1	X_2	CI
1994	12.0	3.25	0.00	1.00	0.50	100.0	100.0	100.0	100.0	100.0	100.0
1995	12.0	3.25	0.00	1.00	0.50	100.0	100.0	100.0	100.0	100.0	100.0
1996	12.3	3.25	0.04	1.00	0.52	102.5	100.0	101.3	102.3	100.0	101.1
1997	12.7	3.23	0.09	0.83	0.46	105.8	99.4	102.6	105.3	90.0	97.6
1998	13.0	3.20	0.13	0.58	0.35	108.3	98.5	103.4	107.5	75.0	91.3
1999	13.7	3.19	0.21	0.50	0.36	114.2	98.2	106.2	112.8	70.0	91.4
2000	14.3	3.18	0.29	0.42	0.35	119.2	97.8	108.5	117.3	65.0	91.1
2001	15.0	3.17	0.38	0.33	0.35	125.0	97.5	111.3	122.5	60.0	91.3
2002	15.7	3.15	0.46	0.17	0.31	130.8	96.9	113.9	127.8	50.0	88.9
2003	16.3	3.14	0.54	0.08	0.31	135.8	96.6	116.2	132.3	45.0	88.6
2004	17.0	3.13	0.63	0.00	0.31	141.7	96.3	119.0	137.5	40.0	88.8
2005	18.0	3.13	0.75	0.00	0.38	150.0	96.3	123.2	145.0	40.0	92.5
2006	19.0	3.13	0.88	0.00	0.44	158.3	96.3	127.3	152.5	40.0	96.3
2007	19.5	3.13	0.94	0.00	0.47	162.5	96.3	129.4	156.3	40.0	98.1
2008	20.0	3.13	1.00	0.00	0.50	166.7	96.3	131.5	160.0	40.0	100.0

Table 1 shows the original values of the indicators and the normalized values by: (a) Min-Max method, (b) indicization, (c) constrained Min-Max method. In (a) and (c) the goalposts are the minimum and the maximum over time; in (b) and (c)

[4] Since 2009, these indicators have been replaced and then there is no more recent data.

the base/reference is the 1994 value. In order to make it easier to compare (b) and (c), we multiplied formula (1) by 60 and we added 100[5].

Figure 1 – *Comparing normalization methods.*



[5] Note that this is the normalization method used in the Adjusted Mazziotta-Pareto Index (Mazziotta and Pareto, 2016).

Finally, for each normalization method, a composite index (CI) was calculated by a simple arithmetic mean.

As we can see, from 1994 to 1998, X_2 decreases by 1 standard deviation (from 3.25 to 3.20), whereas X_1 increases by 0.35 standard deviation (from 12 to 13), thus the variation of X_2 is wider than that of X_1.

However, if indicators are normalized by indicization, X_1 changes from 100 to 108.3 (+8.3%) and X_2 changes from 100 to 98.5 (-1.5%). So, normalizing by indicization, the variation of the indicator with greater CV (X_1) is considered more important than the variation of the indicator with less CV (X_2).

In contrast, if indicators are normalized by the constrained Min-Max method, X_1 changes from 100 to 107.5 (+7.5%) and X_2 changes from 100 to 75 (-25%). consistently with the different variability of the two indicators. The same happens with the Min-Max method, but in this case the reference (or base value) is lost and reading the results is more difficult.

In Figure 1, the line-plots of X_1, X_2 and CI are reported for the three normalization methods. The effect of the different normalization methods on the trend of the composite index is evident. With indicization, CI is increasing over time, since X_1 has an implicit weight greater than X_2. With the classical and the constrained Min-Max method, CI is more stable over time, since X_1 and X_2 have the same weight and the increase of X_1 is offset by the decrease of X_2.

Nevertheless, if a new year of data becomes available, it may happen that the minimum or maximum across units over time, for one or more indicators, changes. In such case, if the indicators are normalized by the classical Min-Max method using the existing goalposts, some values can fall below 0 or above 1. To prevent this, the goalpost should be updated, and CI should be is recalculated across the past years. If instead indicators are normalized by the constrained Min-Max method, CI maintains comparability between the existing and the new data (similarly to indicization).

5. Conclusions

Normalizing data to make them comparable both across units and over time is not a trivial task. The matter can get complicated if new observations are added every year. The most used methods are the Min-Max method and Indicization. The Min-Max method normalizes the variability of indicators, but do not use a common reference to compare them. Indicization uses a common reference to compare indicators but does not normalize their variability.

The constrained Min-Max method combines the advantages of the two methods, as it normalizes the variability of indicators and uses a common reference to

compare them. Furthermore, normalized data maintain comparability even when new data are added.

The method is particularly recommended for the normalization of time series of 'unbounded' indicators[6], when implicit weighting is not desired.

Particular attention must be paid to the choice of normalization parameters (*goalposts* and reference value). They must be kept fixed over time and must not be recalculated every time new data are added. In addition, to facilitate reading, the reference value should be within the range defined by the *goalposts*. When the *goalposts* or the reference value will be considered obsolete, it will be sufficient to define new parameters and normalize all the time series again, similarly to the rebasing of index numbers.

References

BOOYSEN F. 2002. An overview and evaluation of composite indices of development, *Social Indicators Research*, Vol. 59, pp. 115-151.

MAZZIOTTA M., PARETO A. 2016. On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena, *Social Indicators Research*, Vol. 127, pp. 983-1003.

MAZZIOTTA M., PARETO A. 2017. Synthesis of Indicators: The Composite Indicators Approach, in F. MAGGINO (eds.), *Complexity in Society: From Indicators Construction to their Synthesis*, Cham: Springer.

MAZZIOTTA M., PARETO A. (a cura di) 2020. *Gli indici sintetici*. Torino: Giappichelli.

MAZZIOTTA M., PARETO A. 2021. Everything you always wanted to know about normalization (but were afraid to ask). *Italian Review of Economics, Demography and Statistics*, Vol. LXXV, 1, pp. 41–52.

MICHALOS, A.C., SMALE, B., LABONTÉ, R., MUHARJARINE, N., SCOTT, K., MOORE, K., SWYSTUN, L., HOLDEN, B., BERNARDIN, H., DUNNING, B., GRAHAM, P., GUHN, M., GADERMANN, A.M., ZUMBO, B.D., MORGAN, A., BROOKER, A.-S., HYMAN, I. 2011. *The Canadian Index of Wellbeing. Technical Report 1.0,* Canadian Index of Wellbeing and University of Waterloo, Waterloo.

OECD 2008. *Handbook on Constructing Composite Indicators. Methodology and user guide*, Paris: OECD Publications.

[6] Indicators can be divided in 'bounded' and 'unbounded'. An indicator is 'bounded' when it ranges between fixed values (e.g., 'Employment rate'). An indicator is 'unbounded' when there are no predetermined upper or lower limits (e.g., 'GDP per capita') (Mazziotta and Pareto, 2017).

TARANTOLA, S. 2008. *European Innovation Scoreboard: strategies to measure country progress over time*. Luxembourg: Office for Official Publications of the European Communities.
UNDP 1990. *Human Development Report 1990: Concept and Measurement of Human Development*. Oxford: Oxford University Press.
UNDP 2010. *Human Development Report 2010*. New York: Palgrave Macmillan.

SUMMARY

Data Normalization for Aggregating Time Series: the Constrained Min-Max Method

This paper presents a method for normalizing data in time series, when variables have different measurement units and they must be aggregated (e.g., for constructing a composite index). The proposed method, denoted as "Constrained Min-Max Method", normalizes the range of variables, similarly to the Min-Max method, but uses a common reference that allows to 'centre' them, as in the case of index numbers. A comparison with the traditional normalization methods is also shown.

Matteo MAZZIOTTA, Istat, mazziott@istat.it
Adriano PARETO, Istat, pareto@istat.it

PUBLIC OPINION TOWARDS A UNIVERSAL BASIC INCOME IN EUROPE

Paolo Emilio Cardone

1. Introduction

The idea of a Universal Basic Income (UBI) has gained increasing attention in public debates and among policymakers across Europe (De Wispelaere and Stirton, 2004; OECD, 2017).

In the decade since the great recession, the issue of whether a universal basic income (UBI) can provide a guaranteed, basic living standard for all in society has increased in prominence in political and academic discourse.

Obtaining a sufficient citizenship-based income without work obligations is fundamentally opposing the foundations of the welfare systems that are in place nowadays.

In particular, providing a sufficiently high income for all, regardless of their need for support and without work obligations, is fundamentally at odds with the foundations of European welfare systems, where reciprocity and need play a crucial role.

The introduction of a universal basic income would provide a flat-rate cash transfer to all citizens irrespective of their previous contributions and present situation.

The aim of a universal basic income scheme is to gradually replace all other forms of transfer and presently several experiments are underway in European countries.

Less clear, however, are the factors influencing support for a UBI.

Proponents hold that a UBI can provide a flexible protection against poverty and destitution in light of increasingly fragmented labour markets and the threat of automation (Standing, 2011; Van Parijs and Vanderborght, 2017).

Advocates stress its universalism, which reduces the gaps in coverage of existing welfare state policies. Its unconditional characteristic could also decommodify labour more fully, thereby increasing the bargaining power of workers to push for better working conditions and wages, especially at the low-skill end of the labour market.

Notable advocates have argued that UBI has the potential to fully "emancipate" its recipients and allow them to pursue the life they desire unencumbered (see, for example, Van Parijs, 1991; Van Parijs and Vanderborght, 2017).

Critics, meanwhile, consider the UBI economically inefficient or as posing a disincentive to work.

Since a UBI continues to be paid regardless of whether people are in jobs or not, they also emphasise its low adverse effects on work incentives.

Many liberal economists argue that a UBI generous enough to achieve its objectives would be too expensive. It is also inefficient, they reason, as it targets resources to those who may not need them most (Kay, 2017). Others on the Marxist left see a UBI as a politically dangerous legitimation of capitalism, while social democrats worry that the UBI represents an implicit abandonment of the full employment objective (Hassel, 2017).

An investigation into sources of support for a UBI is necessary to further understand the political feasibility of the policy.

Starting from the large and excellent literature on the UBI, this paper analyses empirically a broad range of explanatory individual and contextual factors that may affect popular support for a Universal Basic Income, using a recently conducted wave of the cross-national European Social Survey (ESS).

The ESS Round 8 module "*Welfare Attitudes in a Changing Europe: Solidarities under Pressure*" makes it possible to shed scientific light on these debates.

2. Data and methods

The analysis is carried out using microdata from the quantitative research "*European Social Survey*" (ESS Data, 2016)[1].

The ESS source questionnaire contains a "core" module, which largely remains the same each round[2].

In each round, there are also two short "rotating" modules, which are developed by competitively selected, multinational questionnaire design teams in collaboration with the Core Scientific Team.

In Round 8 these modules focus on:

- Public Attitudes to Climate Change, Energy Security and Energy Preferences;
- Welfare Attitudes in a Changing Europe: Solidarities under Pressure (repeat module with a number of new items).

[1] For more details: https://www.europeansocialsurvey.org.

[2] For more details: www.europeansocialsurvey.org/methodology/questionnaire.

In particular, the core and rotating modules that form the backbone of the ESS questionnaires have addressed multiple topics, including attitudes toward the media, social trust, politics, democracy and citizen involvement; subjective well-being and human values; attitudes towards immigration; family, work and well-being, the timing of life and gender roles; economic morality, welfare attitudes and justice; public attitudes toward climate change.

More in details, the inclusion of the Welfare Attitudes in Europe module during Round 8 of the ESS, first of all allowed attitudes towards these services to be assessed in 23 countries, but also it addresses new solidarity questions fielded for the first time, most notably items assessing the introduction of a Universal Basic Income (UBI) scheme and the implementation of a European Union-wide social benefit scheme[3].

Round 8 of the ESS (about 44,000 individuals aged 15 or older) was fielded in 23 countries: Austria, Belgium, Czech Republic, Estonia, Finland, France, Germany, Hungary, Iceland, Ireland, Israel, Italy, Lithuania, Norway, the Netherlands, Poland, Portugal, Russia, Slovenia, Spain, Sweden, Switzerland and the United Kingdom.

Studying support for UBI, especially in a comparative perspective, was long hindered by the lack of availability of high-quality survey data.

Moreover, many questions hinting at a UBI were ambiguous, for example, asking about support for a "guaranteed minimum income", which could be interpreted not only as support for a BI, but also as support for social assistance.

This is the first survey that introduces a comprehensive idea of UBI and its different aspects: its guaranteed minimum income, its universal character, its unconditionality, that it replaces other benefits and services and that it is paid for by taxes[4].

Specifically, respondents are asked whether they are "against or in favour of the UBI scheme" being introduced in their respective country, which "some countries are currently talking about", with the following characteristics and framing the question of UBI support in the following way:

"The government pays everyone a monthly income to cover essential living costs. It replaces many other social benefits. The purpose is to guarantee everyone a minimum standard of living. Everyone receives the same amount regardless of whether or not they are working. People also keep the money they earn from work

[3] Public support for an EU-wide social benefit scheme is widely described and analysed in *Cardone* (2021) and *Cardone et al.* (2019).

[4] The European Social Survey (ESS8) data wave is the first international academic survey to directly pose a question on UBI, thus allowing for a comparative inquiry into the determinants of UBI support while controlling for a range of individual and country level variables. By contrast, many previous studies have been limited to one or a few countries (Andersen, 1996; Andersson and Kangas, 2005) or the result of *ad hoc* surveys.

or other sources. This scheme is paid for by taxes. Overall, would you be against or in favour of having this scheme in [your country]?"

Survey participants select from a four-item scale to indicate whether they are "strongly in favour", "in favour", "not in favour" or "strongly not in favour" for the UBI (item E36).

As mentioned, the definition used in the ESS8 provides an accurate starting point for analysis, defining UBI as an income (1) paid by government to everyone on a monthly basis to cover living costs, (2) financed by taxes, (3) replacing many other social benefits to (4) guarantee a minimum standard of living, (5) with no variation depending on whether recipients are working and (6) allowing people to keep money earned from work or other sources.

3. Results

We first present descriptive findings on countrylevel support for a UBI. Figure 1 presents mean values of support across the 23 countries examined. Here, we separate and stack the share of respondents proclaiming to be "in favour" and "strongly in favour" of a UBI. Norway, Switzerland and Sweden stand out as the countries with the least level of support for a UBI. In either country, around one-third of respondents indicated favourable attitudes towards a UBI.

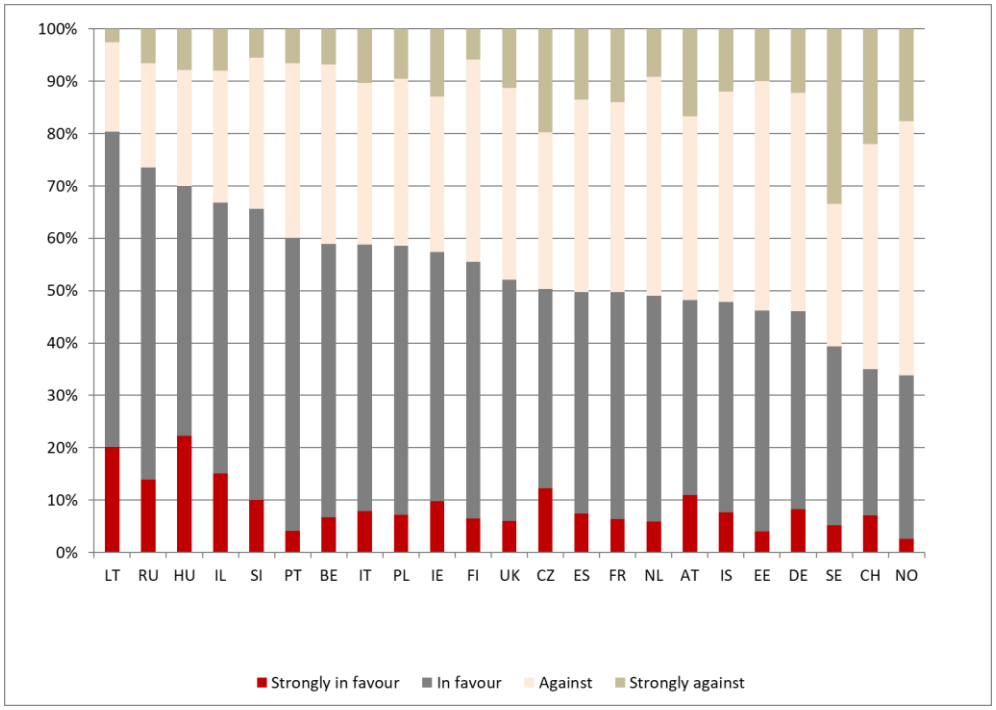
Conversely, more than two-thirds of respondents in Lithuania, Russia, Hungary, Israel and Slovenia are in favour of a UBI.

As shown in figure 1, we see ample variation in support for UBI with the highest level of support in Lithuania (over 80% indicates to be in favour of a UBI) and the lowest level of support in Norway (over 66% indicates to be against a UBI).

Nevertheless, it is remarkable that support for such a radical alternative to the current welfare system gains so much support in European countries. In total, about 56% of the respondents in this selection of countries indicates support, while 44% is against a UBI. As figure 1 illustrates, sizable crossnational differences exist, but in 20 out of 23 countries, support is higher than 45%. Overall, it seems that populations in Eastern European countries are more in favour than their counterparts in Nordic and Western European countries, although there are some exceptions (for instance, Estonia being more against UBI and Finland more in favour compared with their respective region).

Over 50% of Russian and Lithuanian respondents think that they are not likely to have enough money for their household necessities in the next 12 months. In Norway and Sweden, less than 10% of respondents share this fear.

Figure 1 - *Public support for a Universal basic Income, by country (% values).*

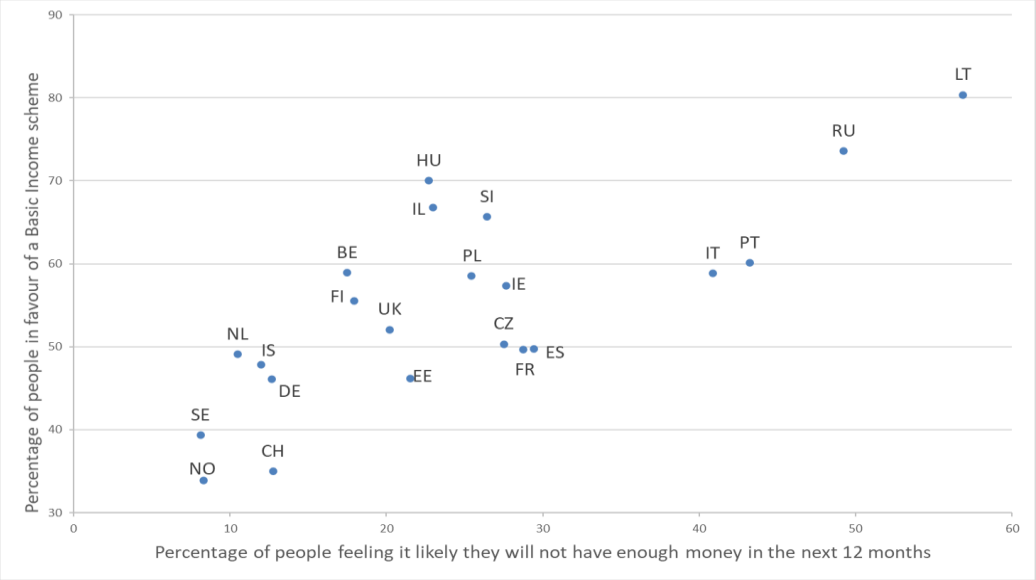


Source: own elaboration on ESS data Round 8.
Note: Design weights are applied.
Legend: LT=Lithuania; RU=Russia; HU=Hungary; IL=Israel; SI=Slovenia; PT=Portugal; BE=Belgium; IT=Italy; PL=Poland; IE=Ireland; FI=Finland; UK=United Kingdom; CZ=Czech Republic; ES=Spain; FR=France; NL=Netherlands; AT=Austria; IS=Iceland; EE=Estonia; DE=Germany; SE=Sweden; CH=Switzerland; NO=Norway.

This relationship generally holds true for the sample of European countries: the stronger the concern for future unstable economic conditions in a country, the stronger the preference for a basic income scheme (figure 2).

With a few exceptions, we can identify the different Macro-Regions in the figure: the Nordic countries are distinguished by low to medium support for a basic income while the fear for not having enough money is low; in Western European countries, the fear is slightly stronger as is the support for a basic income; in Eastern Europe, the fear is relatively high as well as the support for a basic income; while the Southern European countries are characterized by having a strong fear but a comparatively low, but still strong, support for a basic income scheme. Russia and Lithuania form a group of their own, distinguished by a strong fear of not having enough money in the future, as well as strong support for a basic income scheme.

Figure 2 - *Support for basic income scheme and perceived unstable financial future (% values).*



Source: own elaboration on ESS data Round 8.
Note: Design weights are applied.
Percentage of respondents in favour or strongly in favour of a basic income scheme, answered on a 4-point scale: "strongly against", "against", "in favour" or "strongly in favour" (item E36, N=40.592).
Percentage of people feeling it likely or very likely they will not have enough money to cover household necessities in the next 12 months, answered on a 4-point scale: "not at all likely", "not very likely", "likely" or "very likely" (item E40, N=40.612).

The ESS asks respondents about their history of unemployment: over 63% of respondents who have experienced a period of unemployment and work seeking within the last five years are favourable to a UBI compared to 54% for those who have not (table 1).

The difference between temporary and permanent workers in their support for a UBI is not trivial: 56 % of respondents with "limited duration" contracts favour a UBI compared to 49.4 % for those with "unlimited duration" employment contracts (table 2).

Table 1 - *Any period of unemployment and work seeking within last 5 years and support for a UBI.*

Period of unemployment?	Strongly against	Against	In favour	Strongly in favour	Total
Yes	9.93	26.90	50.31	12.86	100.00
No	10.58	35.25	46.08	8.08	100.00

Source: own elaboration on ESS data Round 8.
Note: Design weights are applied.

Table 2 - *Employment contract and support for a UBI.*

Employment contract unlimited or limited duration	Strongly against	Against	In favour	Strongly in favour	Total
Unlimited	11.85	34.67	44.99	8.48	100.00
Limited	10.45	33.33	45.96	10.26	100.00

Source: own elaboration on ESS data Round 8.
Note: Design weights are applied

Using multivariate analysis (logistic regression models with Stata software) it was possible to estimate the different attitudes among countries for UBI support more accurately. The model includes, first of all adults' socio-demographic characteristics (age, gender, number people living in the household, citizenship, domicile, education level, voted or not), secondly, economic and work-related (worked or not, total household income).

In order to achieve this goal, the dependent variable of this study is the "*basic income scheme*": we recode respondents' answers into a binary outcome variable which receives a value of 1 if the respondent supports or strongly supports a UBI and takes a value of 0 if the respondent does not support or strongly does not support a UBI.

Concretely, in the study analyzed variables are:

- *Gender*. Categorical. Dummy variable: Female, Male (reference cat.).
- *Country*. Categorical. Twentythree countries. Netherlands (reference cat.), Portugal, Spain, Slovenia, Lithuania, Italy, Hungary, Poland, Ireland, Belgium, Estonia, Germany, France, Czech Republic, Sweden, United Kingdom, Finland, Austria, Russia, Israel, Iceland, Norway and Switzerland.
- *Domicile*. Categorical. Four levels. A big city/Suburbs or outskirts of big city (reference cat.); Town or small city; Country village; Farm or home in countryside.
- *Work*. Categorical. Dummy variable: Yes, No (reference cat.).

- *Income*. Categorical. Ten levels: 1st decile (reference cat.), 2nd decile, 3rd decile, 4th decile, 5th decile, 6th decile, 7th decile, 8th decile, 9th decile, 10th decile.
- *Household*. Categorical. Five levels: 1 individual (reference cat.), 2 ind., 3 ind., 4 ind., 5 ind. or more.
- *Vote*. Voted in the last election. Categorical. Three levels: No, Not eligible to vote, Yes (reference cat.).
- *ISCED*. Categorical. Three levels: Low (Isced 0-1-2), Medium (Isced 3-4), High (Isced 5-6, reference cat.).
- *Age group*. Categorical. Three intervals. From 15 to 40; 40 to 60; over 60 (reference cat.).

First of all, we test the goodness-of-fit using a postestimation tool, the Hosmer-Lemeshow statistic. This test follows a chi-square distribution with the degrees of freedom equal to the number of groups minus 2. A not significant *p value* indicates that the model fits the data well since there is no significant difference between the observed and expected data (Liu, 2016). In this case, the Hosmer-Lemeshow chi-square test has a value of 12.68 with the degrees of freedom equal to 8. The associated *p value* is 0.1235 which is not significant. Therefore, the model fits the data well.

Logistic model for "*basic income scheme*", goodness-of-fit test:
Number of groups = 10
Hosmer-Lemeshow chi2(8) = 12.68
Prob > chi2 = 0.1235

Table 3 shows odds ratios of logistic model and this means that the coefficients (*Beta*, not showed) in logistic regression are in terms of the log odds because the coefficients can be expressed in odds by getting rid of the natural log. This is done by taking the exponential for both sides of the equation, because there is a direct relationship between the coefficients produced by logit and the odds ratios produced by logistic: a logit is defined as the natural log (base *e*) of the odds.

Table 3 - *Logistic regression model.*

Number of obs = 33,576; LR chi2(46) = 2,151.21; Prob > chi2 = 0.0000
Log likelihood = -22,106.42 Pseudo R2 = 0.0464

Variables		ODDS	Sign.
• Gender	Male	1 (base)	
	Female	0.96	0.041

Table 3 - *Logistic regression model (continued).*

• Country	AT	0.75	0.000
	BE	1.36	0.000
	CH	0.51	0.000
	CZ	1.07	0.341
	DE	0.90	0.107
	EE	0.81	0.003
	ES	0.93	0.378
	FI	1.27	0.001
	FR	0.89	0.117
	UK	0.95	0.496
	HU	2.05	0.000
	IE	1.17	0.036
	IL	1.72	0.000
	IS	0.83	0.045
	IT	1.04	0.637
	LT	3.78	0.000
	NL	1 (base)	
	NO	0.47	0.000
	PL	1.29	0.002
	PT	1.42	0.000
	RU	2.43	0.000
	SE	0.63	0.000
	SI	1.82	0.000
• Domicile	A big city/Suburbs	1 (base)	
	Town or small city	0.92	0.003
	Country village	0.87	0.000
	Farm or home in countryside	0.82	0.000
• Work	No	1 (base)	
	Yes	0.88	0.000
• Income	J - 1st decile	1 (base)	
	R - 2nd decile	0.91	0.063
	C - 3rd decile	0.94	0.252
	M - 4th decile	0.88	0.017
	F - 5th decile	0.74	0.000
	S - 6th decile	0.73	0.000
	K - 7th decile	0.73	0.000
	P - 8th decile	0.69	0.000
	D - 9th decile	0.59	0.000
	H - 10th decile	0.54	0.000
• Household	Single person / lone parent	1 (base)	
	2	1.10	0.005
	3	1.19	0.006
	4	1.12	0.009
	5 or more	1.10	0.058
• Vote	Yes	1 (base)	
	No	1.11	0.001
	Not eligible to vote	1.12	0.015
• ISCED	Low	1.03	0.424
	Medium	0.93	0.008
	High	1 (base)	
• Age group	15 – 40	1.44	0.000
	40 – 60	1.23	0.000
	Over 60	1 (base)	
	cons.	1.22	0.008

Source own elaboration on ESS data Round 8.

This fitted model says that, holding covariates at a fixed value, the odds of being in favour of a public support for a UBI scheme for female over the odds of being in favour of a public support for a UBI scheme for male (reference category) is 0.96. In terms of percent change, we can say that the odds for female are 4% lower than the odds for male. In other words, the hazard to be in favour of a public support for an EU-wide social benefit scheme is slightly higher for male rather than female.

Regarding the citizenship, the odds of being in favour of a public support for a UBI scheme for ten countries (Belgium, Finland, Hungary, Ireland, Israel, Lithuania, Poland, Portugal, Russia and Slovenia) are higher over the odds of being in favour of a public support for a UBI scheme for The Netherlands (reference category). In particular, the odds for Lithuania are almost four times higher than the odds for The Netherlands (OR=3.78) and the odds for Russia and Hungary are double (OR=2.43 and OR=2.05). On the other side, the odds for six countries (Austria, Switzerland, Estonia, Iceland, Norway and Sweden) are lower than the odds for The Netherlands. In terms of percent change, the odds for Slovenia (OR=1.82) are 82% higher than the odds for The Netherlands and the odds for Switzerland are 49% lower (OR=0.51). Please note that the odds for Czech Republic, Germany, Spain, France, Italy and United Kingdom are not significant (*p value* > 0.05).

The hazard to be in favour of a public support for a UBI scheme is higher for young people (younger ones have more confidence than the elderly, "over 60" reference cat.) and lower for those who have a medium education level (Isced 5-6 reference cat.). Moreover, it decreases with household income (1st decile reference cat.), for those who live in a small town, country village or farm (big city reference cat.) and for workers (not workers reference cat.). On the contrary, the hazard to be in favour increases for those who do not vote or are not eligible (those who vote reference cat.) and for individuals who belong to families of 2 or more people (single persons/lone parents reference cat.).

4. Conclusions

This study contributes to the wider debate surrounding "Social Europe". The main conclusion is that support for UBI is high across Europe but it can observe important cross-national variation in support for a UBI. Support for basic income seems to be lower in more affluent countries in Northern and Western Europe, and higher in the less wealthy welfare states in the East. Countries with more limited and less generous welfare states (southern, liberal and central and eastern European welfare regimes) tend to exhibit higher support. Belgium and Finland are outliers in this cluster, probably due to their long standing unemployment problems in their labour market. Moreover, adults' socio-demographic, economic and work-related

characteristics play an important role. Basic income is favoured more by the young than the olders while high incomes tends to reduce support for UBI if compared with low incomes. As previously seen, labour market status significantly predicts basic income support: people who have experienced a period of unemployment and work seeking within the last five years are more favourable to a UBI compared to those who have not.

Finally, research community need more studies to ask people about their exact understandings of UBI, which elements of the policy they support or reject, and their argumentations to be in favour or against it.

Acknowledgements

I would like to take this opportunity to thank anonymous reviewers and participants at the 2021 SIEDS Scientific Meeting for their helpful comments and suggestions in the course of writing this paper.

References

ANDERSEN J.G. 1996. Marginalization, Citizenship and the Economy: The Capacities of the Universalist Welfare State in Denmark. In ERIKSEN E.O., LOFTAGER J. (Eds.) *The Rationality of the Welfare State*, Oslo: Universitetsforlaget, pp. 155-202.

ANDERSSON J.O., KANGAS O. 2005. Popular Support for Basic Income in Sweden and Finland. In STANDING G. (Ed.) *Promoting Income Security as a Right: Europe and North America*. London: Anthem Press, pp. 289-301.

CARDONE P.E. 2021. Public support for an Eu-wide social benefit scheme: evidence from round 8 of European Social Survey (ESS). *RIEDS (Rivista Italiana di Economia, Demografia e Statistica)*, Vol. LXXV, No. 1, pp. 73-84. ISSN 0035-6832.

CARDONE P.E., DEIDDA M., MAROCCO M. 2019. Le opinioni sulla condizionalità: i risultati in Italia dell'European Social Survey. *Inapp Paper* n. 21, Roma, Inapp, p. 17.

DE WISPELAERE J., STIRTON L. 2004. The many faces of universal basic Income. *The Political Quarterly*, Vol. 75, No. 3, pp. 266-274.

EUROPEAN SOCIAL SURVEY Round 8 DATA 2016. *Data file edition 2.1*. NSD - Norwegian Centre for Research Data, Norway - Data Archive and distributor of ESS data for ESS ERIC.

HASSEL A. 2017. Unconditional basic income is a dead end. In VAN PARIJS P. (Ed.) *Basic Income and the Left: A European Debate*. Social Europe.

KAY J. 2017. The basics of basic income. *InterEconomics*, Vol. 52, No. 2, (March/April), pp. 69–74.

LIU X. 2016. *Applied Ordinal Logistic Regression using Stata. Sage Publications*, pp. 121-122.

OECD 2017. *Basic income as a policy option: Can it add up? Policy brief on the future of work*. Paris: OECD.

STANDING G. 2011. *The Precariat: The New Dangerous Class*. London: Bloomsbury.

STATA, *Software for Statistics and data Science*: https://www.stata.com.

VAN PARIJS P. 1991. Why Surfers Should Be Fed: The Liberal Case for an Unconditional Basic Income, *Philosophy and Public Affairs*, Vol.20, No.2, pp. 101-31.

VAN PARIJS P., VANDERBORGHT Y. 2017. *Basic Income - A Radical Proposal for a Free Society and a Sane Economy*. Cambridge, MA: Harvard University Press.

SUMMARY

Europeans' attitudes towards the idea of a Universal Basic Income

The idea of a Universal Basic Income (UBI) has gained increasing attention in public debates and among policymakers across Europe. The inclusion of the Welfare Attitudes in Europe module during Round 8 (2016/17) of the European Social Survey (ESS) - for the first time in academic cross-national research - allowed attitudes towards the introduction of a Universal Basic Income to be assessed in 23 countries. Aim of this article is to explore and understand the different aspects and predictors of UBI support. In particular, this paper analyses a broad range of explanatory individual and contextual factors that may affect popular support for a UBI. Main findings shows that the stronger support for a European minimum income benefit in less generous welfare states is explained by more optimistic expectations about the EU's domestic impact and lower socioeconomic status groups are more supportive of this policy proposal. Moreover, the analysis reveals that diverging national experiences and expectations are crucial in understanding why Europeans are widely divided on the implementation of such a benefit scheme. Secondly, in almost all countries, the younger age group is more supportive of activation and a universal basic income than the older age group. Finally, using logistic regression model it is possible to estimate the different attitudes among countries for a UBI more accurately.

Paolo Emilio CARDONE, INAPP, p.cardone@inapp.org

THE WELLBEING OF SINGLE PARENTS IN ITALY BEFORE AND AFTER THE COVID-19 PANDEMIC

Raffaele Guetto, Elena Pirani, Patrizio Lodetti

1. Introduction

In the last two decades, Italian society has been experiencing great transformations in its family demography, partly bridging the gap with other Western European countries in terms of the diffusion of "new" family forms (Castiglioni and Dalla Zuanna, 2009; De Rose and Vignoli, 2011). Marriage does not represent anymore the unique context for childbearing, non-marital unions are becoming more and more popular not only as a prelude to marriage, unsatisfied partners increasingly opt for divorce, and re-partnering is becoming more common, even in presence of children from the previous union (Pirani *et al.*, 2021). An important consequence of these changes is the strong increase of children experiencing the dissolution of their parents' union and being raised in a non-intact family, namely a household where either the mother or the father is absent.

In this paper, we first describe the recent evolution of Italian families and their characteristics, with a special focus on single parents, considering both those living with their children and those without them. In doing so, we discuss the risks of economic deprivation and social exclusion to which these households are exposed, and that make them particularly vulnerable to the negative consequences of the recent COVID-19 pandemic outbreak.

Many theoretical and empirical studies on the socio-economic consequences of the pandemic have highlighted the difficult situation faced by working women with children, in terms of increasing unpaid work burden and the related difficulties in reconciling work and family duties (Power, 2020; Profeta, 2020). However, the specific situation of single parents, which are, in most cases, single mothers, has been scarcely considered, especially in the Italian context. To fill this gap in the literature, we present the first results of an *ad hoc* survey aimed at evaluating the consequences of the pandemic on the wellbeing and living conditions of single parents in Italy. Data collection occurred in April 2021 and targeted single parents by taking the variety of family situations in which they are involved into account. We present results concerning the psychological and economic wellbeing of a sample of 715 single parents before and after the pandemic.

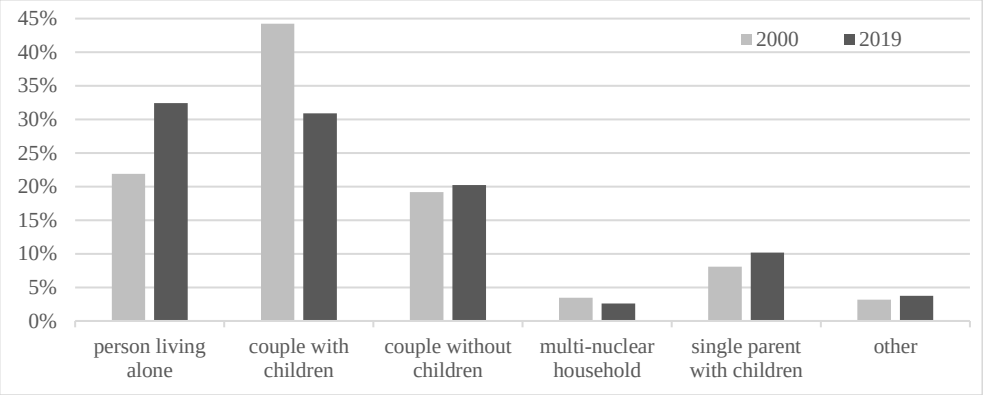
2. Italian families and their characteristics

2.1. Two decades of increasing family complexity

In 2019 Italy counted almost 25.7 million households, with an increase of more than 4 million units compared to the beginning of the new Millennium. Such fast and extraordinary growth in less than 20 years reflects the "exceptionalism" of Italian demography in recent decades (Billari and Tomassini, 2021).

This quantitative change has been accompanied by even more relevant qualitative transformations (Pirani *et al.*, 2021). For instance, if in the early 2000s the so-called "traditional" families – namely couples with at least one child – represented the 44% of Italian households (Figure 1), nowadays they are slightly above a third of the total. This decrease is similar to the one recorded in Southern and Eastern European countries, but more marked compared to that of Northern Europe (Eurofound, 2019).

Figure 1 – *Distribution of Italian households, by type, in 2000 and 2019*



Source: Author's elaborations on the surveys on Aspects of Daily Life, ISTAT, 2000 and 2019.

A second major transformation compared with the beginning of the new Millennium refers to the substantial increase in single-person households, which in 2019 account for about one-third of Italian households (about 8.6 million), and to which the largest part of the overall increase in the total number of households can be ascribed to. This sharp increase is largely due to adults who remain alone after union dissolution, whereas widowhood impacts less than in the past, and mainly at very old ages.

The increasing trend in union dissolution has also induced an increase in the number and percentage of single-parent households, which shifted from less than 1.8 million in 2000 to approximately 2.6 million in 2019, 10% of the total number of households. Interestingly, also those in which a single father lives with at least one child are on the rise (from 350 thousand in the early 2000s to almost 514 thousand households in 2019),

although the large majority of single-parent households (about 80%) continue to be composed by single mothers and one or more children.

It is worthwhile to recall that the distribution of household types is not homogeneous across Italian territories. North-western regions and, to a lesser extent, those of the North-East and the Center, record a higher-than-average number of single-persons, single parents, and childless couples, with values close to those of Northern and Central European countries; on the contrary, in the South and Islands the couple with at least one child continues to represent the most widespread type of household, despite the sharp decline recorded in the period considered (from 50% in 2000 to 36% in 2019). This is the result of the different paces of family changes that still characterize the North and South of the country.

2.2. After a union dissolution: the characteristics of single parents with cohabiting and non-cohabiting children

A relevant change in contemporary Italian families is the continuous increase of persons experiencing the dissolution of a union even in presence of (young) children, resulting in an increasing share of single parents either cohabiting or not cohabiting with their children. To have a more detailed insight on this population segment, the focus of this paper, we selected households with a young-adult reference person, i.e. a person aged between 25 and 54.

Among single parents with at least one cohabiting child (first column in Table 1), the proportion of women is strongly overrepresented (83%), and in the vast majority of cases, this living arrangement is the result of a union (cohabitation or marriage) breakdown. 44% of single parents live with two children, who are minor children in 70% of cases.

Single-parent households are only one side of the coin, however, as single-living is one of the possible outcomes following the dissolution of a union with children. In fact, about one-quarter of people who live alone, in the 25-54 age group, have at least one child with whom they do not live. Their characteristics are shown in the last column of Table 1. In this case, the gender proportions are reversed – two out of three are men – and the large majority come from a separation or a divorce (note that if the previous union which originated the child was a non-marital one, even stable and long-lasting, the person's marital status is "single"). The territorial distribution shows a larger prevalence in Northern regions. In a similar way to the group of single parents living with their children, 46% of single parents without cohabiting children has 2 or more children, often of very young age (21% in pre-scholar age and 25% between 6 and 13 years old).

Table 1 – *Characteristics of single parents with cohabiting and non-cohabiting children in Italy, reference person aged 25-54.*

		Single parents with cohabiting children (2019)	Single parents with non-cohabiting children (2016)
Gender	male	16.9	66.0
	female	83.1	34.0
Marital status	single	26.2	19.3
	separated/divorced	65.9	77.2
	widowed	7.9	3.5
Area of residence	north	43.9	48.8
	center	20.3	26.4
	south/islands	35.9	24.8
Number of children	1	55.9	54.4
	>=2	44.1	45.7
Age of the youngest child	0-5	15.8	21.1
	6-13	35.2	24.9
	14-17	18.9	13.4
	>=18	30.1	40.6

Source: Authors' elaborations on Aspects of Daily Life (2019) and Family and Social Subjects (2016) surveys.

2.3. The wellbeing of single parents with cohabiting and non-cohabiting children before and after the COVID-19 pandemic

The increasing complexity of Italian families is not only relevant from a demographic point of view, but it also entails important consequences in terms of social inequalities (McLanahan, 2004). Previous studies showed that different family arrangements are associated with different risks of poverty and economic deprivation of individuals and their family, and may have important implications also for their overall wellbeing. For instance, it is well-known that divorce is associated with strong reductions in household income, especially for women who end up living with and caring for the children (Aassve *et al*., 2007). Single-parent households are, in fact, one of the types with the highest risks of poverty and social exclusion, in Italy and elsewhere (Bozzon *et al*., 2015; Pirani *et al*., 2021). More generally, the living arrangement has been found to be associated with individuals' wellbeing (Vignoli *et al*., 2014). Specifically, a recent study (Waldvogel and Ehlert, 2016) found that fathers who live apart from their minor children, for instance after a separation or divorce, generally report lower psychological wellbeing than men living with children in the same household. The absence of one of the parents, parental stress, and loss of economic resources can produce negative consequences also for children, for instance in terms of educational outcomes (Guetto and Panichella, 2019).

All in all, considering current increasing trends in union dissolutions even in presence of young child(ren), and given that these transformations are progressively spreading to more socioeconomically disadvantaged social groups (Pirani, 2019), it is of paramount importance to understand the characteristics of these families. This is truer in the light of the COVID-19 pandemic outbreak. The COVID-19 pandemic can be thought as a multiplier of previously existing stress factors for parents who experienced union dissolution. As far as the custodial parents, usually mothers, lockdowns and other government restrictions, especially the closure of childcare services and schools, increased their already high burden of housework and childcare. More in general, the suspension of many non-essential working activities may have caused additional economic stress to individuals who already faced the direct and indirect costs of separation. Last but not least, the pandemic and its responses may have increased the above-mentioned risks of psychological distress and social exclusion, also due to reduced contacts with the non-cohabiting children during the lockdown periods. To this aim, in what follows we provide some preliminary results of an *ad hoc* study on the consequences of the pandemic for single parents with and without cohabiting children.

3. Data

In this study, we used a web-based survey to obtain primary data on Italian single parents with cohabiting and non-cohabiting children, considering the consequences of the pandemic on their lives (*SParWell*). We collected information on respondents' socioeconomic situation, sociodemographic characteristics, and several dimensions of wellbeing. Whenever possible, we adopted the same wordings of questions already asked by ISTAT within the nationally-representative survey Aspects of Daily Life (ADL) carried out in 2019 (see next section). Specifically, we created a questionnaire addressed to the members of an organization named GenGle (https://gengle.it/), which aims to connect single parents' families all over Italy and foster the development of support networks among them. The questionnaire was disseminated through the organization's social media channels and their newsletter, reaching thousands of individuals. The data collection phase lasted one month (April 2021) and resulted in 871 responses. Among these, we considered a target of respondents with one or more children conceived with the same partner[1], who experienced the breakup of the relationship (either marriage or cohabitation) with the other parent before 2020, namely

[1] We acknowledge that individuals with children originated in multiple unions (or multiple partners) may represent a significant group in this population segment. Nevertheless, they would have added further complexity in an already heterogeneous scenario, especially when considering parent-child relationships.

before the beginning of the pandemic. We ended up with an analytical sample of 715 individuals, whose characteristics are summarized in Table 2.

Table 2 – *Summary statistics of the SParWell sample (percentages)*

		Sample composition (%)
Gender	male	21.1
	female	78.9
Age group	<40	12.0
	40-45	30.4
	46-50	33.0
	>50	24.6
Area of residence	north	57.5
	center	34.4
	south/islands	7.8
	abroad	0.3
Educational level	lower secondary (ISCED 0-2)	6.3
	upper secondary (ISCED 3-4)	43.4
	tertiary (ISCED 5+)	50.3

Source: Author's elaborations.

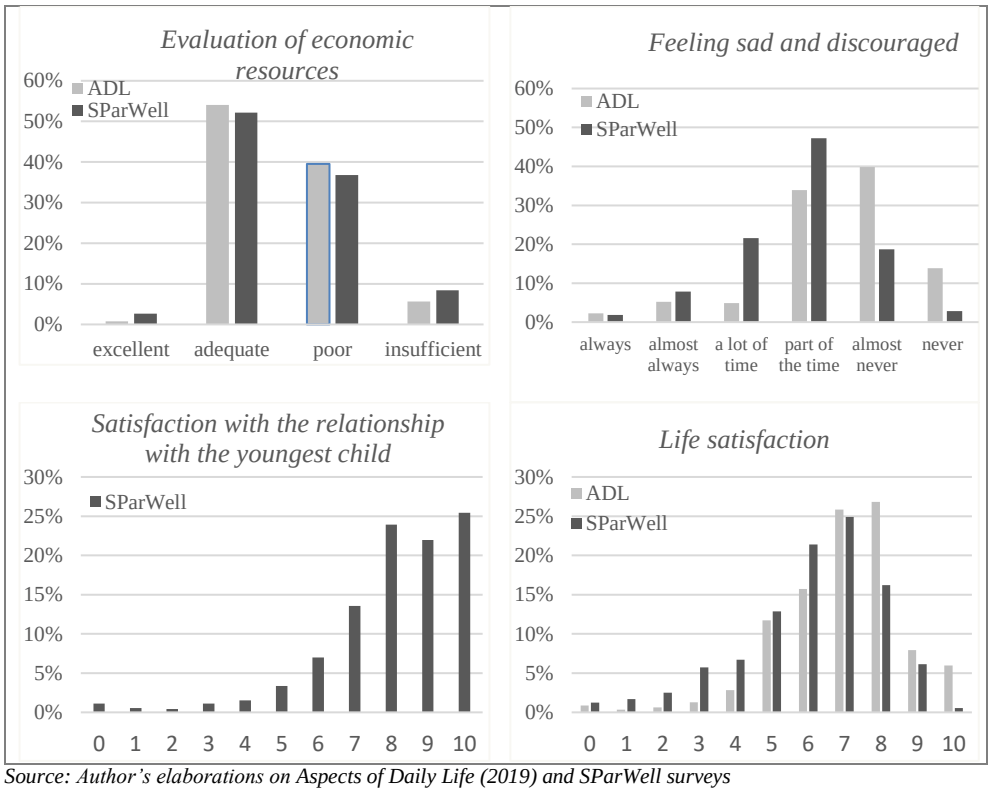
Beyond the usual advantages provided by web surveys, e.g. low costs and high efficiency, there are two specific reasons why we employed this method of data collection. First, in a period of social distancing, face-to-face interviews were not an option. Second, and more general, obtaining a random sample of our target population is not easy, especially as far as the sub-group of parents with non-cohabiting children. On the other hand, we realize that web surveys are exposed to important issues of sampling bias. However, apart from the strong over-representation of highly educated individuals (half of the respondents achieved tertiary education), which is typical of this method of data collection (Duffy *et al.* 2005), the sample characteristics in terms of sex, age, and area of residence are in line with the distributions in the target population at the national level. In fact, 79% of the respondents are women, and the majority reside in the North of Italy (58%). The sample is rather young, 12% of respondents are aged less than 40, 30% between 40 and 45, 33% between 46 and 50, and 25% more than 50.

4. Results

In this section, we present the main descriptive results of the *SParWell* survey, concerning single parents' wellbeing across several dimensions (Figure 2). The answers provided by our respondents are compared with those of the same target population collected within the ADL survey carried out by ISTAT in 2019. The latter is a useful

benchmark, as it is a nationally-representative survey carried out just before the COVID-19 pandemic[2].

Figure 2 – *Single parents' wellbeing, percentage distributions*



Source: Author's elaborations on Aspects of Daily Life (2019) and SParWell surveys

Starting from respondents' evaluations of their own economic resources, almost half of the *SParWell* sample judges them as poor or insufficient. This is not surprising, as we mentioned how the economic conditions of single-parent households are usually worse than those reported in other living arrangements. Moreover, the distribution we obtained is very much in line with that of the ADL survey for this type of household. The situation is markedly different when it comes to non-economic dimensions of wellbeing. In fact, almost half of our respondents felt sad and discouraged for at least part of the time during the four weeks preceding the survey (approximately March/April 2021), with an additional 32% who declared to have felt sad and discouraged for most of the time. The

[2] It should be noted, however, that in the ADL survey only single parents *with cohabiting children* are considered, due to a lack of information to identify non-cohabiting children.

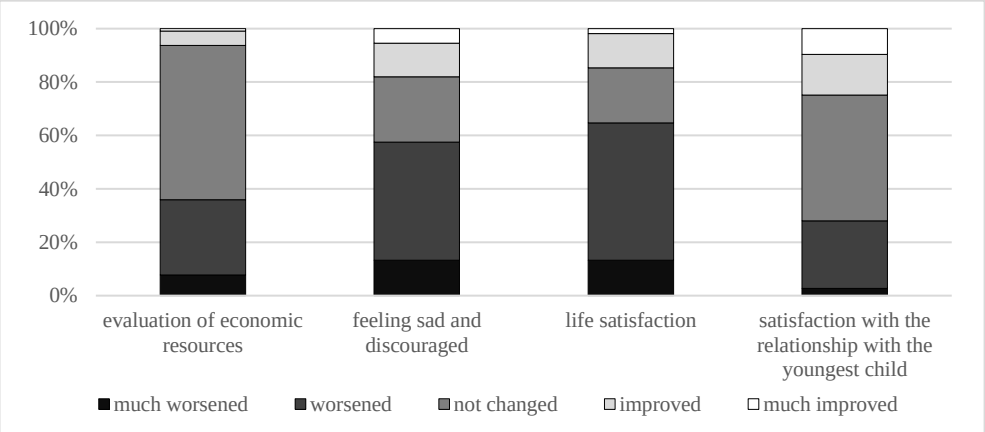
same figures are 34% and 12% in the ADL survey. Similar differences between *SParWell* and ADL data are found for overall life satisfaction. The distribution in our data is still left-skewed (mean and median equal to 6), but the share of respondents with high levels (i.e., 8 or more) of life satisfaction is much lower compared to the ADL survey. A last dimension of wellbeing considered is the satisfaction with the relationship with the youngest (or only) child, which was not asked in the ADL survey. In this case, we detect higher levels of satisfaction (mean and median equal to 8).

Although indirect, the comparison between ours and ADL data suggests that whereas the economic situation of single parents remains difficult, it did not deteriorate substantially due to the pandemic and its responses. On the other hand, non-economic and more subjective measures of wellbeing worsened a lot. Our next results shown in Figure 3 provide more direct evidence on this point. We asked the respondents to compare their current situation concerning the same dimensions of wellbeing analyzed above with their situation before the pandemic outbreak (approximately January 2020). As far as the economic resources, around 60% of our respondents declared that their situation did not change. For approximately one-third of the sample the situation worsened, but only to some extent. On the contrary, the emotional situation and overall life satisfaction deteriorated for approximately 60% and 65% of the sample, respectively. More positive changes concerned, instead, respondents' perception of the quality of the relationship with their youngest child: Whereas for a quarter of our sample the relationship improved after the pandemic, for another quarter it deteriorated. Thus, lockdowns and restrictions to mobility seem not to have impacted, on average, on parent-child relationship in our target population, at least as far as the youngest child is concerned.

In addition to the results shown in Figures 2 and 3, we explored possible heterogeneity in wellbeing, before and after the pandemic, by gender and partnership status. All the reported bivariate associations between gender and the different measures of wellbeing are statistically significant at least at the 5% level. Our data suggest that single-parent women have less wellbeing than single-parent men, especially when measured through subjective, non-economic measures – e.g., only 19% of women declared to have never or almost never felt sad or discouraged in the four weeks before the survey, against 32% among their male counterparts. This difference can be attributed to a stronger negative impact of the pandemic among female single parents. In fact, the emotional situation worsened for 60% of sampled women, against 48% of men. Considering that 96% of single-parent women in our sample cohabit with their youngest child, against only 4% of single-parent men, these differences may be reasonably attributed to the impact of the pandemic on women's unpaid work burden. In support of this interpretation, women reported a substantial worsening of their relationship with their youngest child. Whereas only for a tiny minority of men and women the relationship worsened *a lot*, it did worsen for 29% of women, against only 13% of men.

Our sample is composed of single parents who experienced union dissolution before January 2020; however, they might have found a new cohabiting partner in the meanwhile. This occurred to 43 respondents, the 6% of the sample. Respondents who currently cohabit with a new partner are substantially more satisfied with their life than their unpartnered counterparts, the average scores being 7.1 and 6, respectively. It is plausible that having found a new partner just before or during the pandemic might have reduced its negative consequences in terms of increasing unpaid work burden and feelings of social isolation. Whereas 66% of single parents who did not find a new partner report a worsening of their overall life satisfaction compared to January 2020, the same figure for newly partnered ones is 51%.

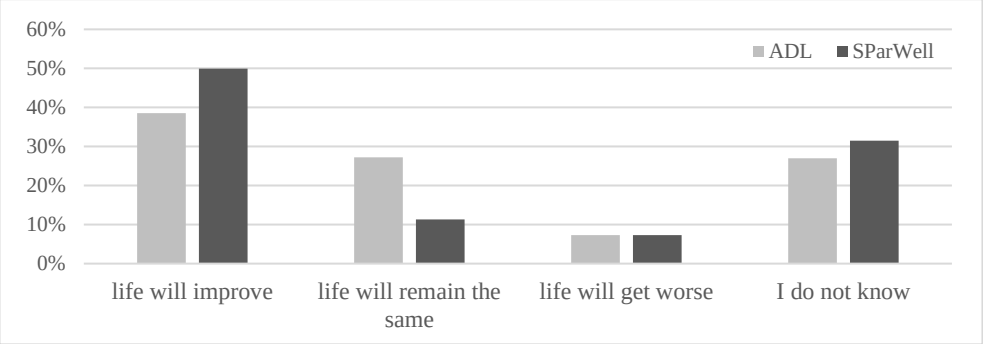
Figure 3 – *Changes in single parents' wellbeing after the pandemic, percentage distribution*



Source: Author's elaborations on SParWell survey

Finally, we asked the respondents whether they think their life will improve, remain the same, or get worse in the next three years. Results are presented in Figure 4 and compared with the ones obtained through the 2019 ADL survey. The share of single parents with an optimistic outlook is much higher in our sample. This suggests that the worsening of wellbeing that we previously outlined can be considered as a direct effect of the COVID-19 crisis. Thus, respondents think their personal situation will improve as soon as the emergency will be resolved.

Figure 4 – *Single parents' expectations for the 3 next years, percentage distributions*



Source: Author's elaborations on Aspects of Daily Life (2019) and SParWell surveys

5. Conclusions

The diffusion of non-marital unions and couple instability in Italy in the last two decades has produced an increase in the share of single parents. We discussed the risks of economic deprivation and social exclusion experienced by this type of family, which make them particularly vulnerable to the negative consequences of the recent COVID-19 pandemic. In this respect, we presented the main results from a survey we carried out in April 2021, *SParWell*, on a sample of single parents with cohabiting and non-cohabiting children. The survey has been carried out with the support of an Italian association (*GenGle*) aiming to connect single parents' families all over the country and to foster support networks among them.

The main results suggest that single parents did not suffer excessively from an economic point of view: This may be due to the already very difficult economic conditions of these families, and partly to the positive selection of our sample. However, our respondents declared a strong decline in psychological wellbeing. This may be connected with increasing difficulties in work-family reconciliation, as the situation especially deteriorated for single mothers, whereas the negative effects of the pandemic seem weaker among single parents who found a new partner during the pandemic. In addition, whereas the relationship with the youngest child did not substantially worsen among our respondents, it did deteriorate among women, which is suggestive as virtually all of them are cohabiting with their children, whereas their male counterparts are not.

Notwithstanding data suggest an important worsening of the wellbeing of single parents due to the pandemic, our respondents remain optimistic toward the future. The pandemic has exacerbated some of the distressing factors of Italian single parents. Some emergency measures have been introduced to support families' income and facilitate work-family life conciliation. However, policies designed on "traditional" families (i.e.

couples with children) may not be suitable to sustain the economic and psychological wellbeing of single-parent families. Their specific social risks – e.g., single income, reduced support from the partner for childbearing, lower-quality parent-child relationships, especially for the non-custodial parent, – require more incisive and targeted policies for the future, even beyond the pandemic.

Acknowledgments

The authors acknowledge the financial support provided by the Italian Ministry of University and Research (MIUR), 2017 MIUR-PRIN Grant Prot. N. 2017W5B55Y ("The Great Demographic Recession": P.I.: Daniele Vignoli).

References

AASSVE A., BETTI G., MAZZUCCO S., MENCARINI L. 2007. Marital disruption and economic well-being: A comparative analysis, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 170, No. 3, pp. 781-799.

BILLARI F.C., TOMASSINI C. 2021. L'eccezionalismo demografico dell'Italia. In BILLARI F.C. and TOMASSINI C. (Eds.) *Rapporto sulla Popolazione, L'Italia e le sfide della demografia*, Bologna: Il Mulino, pp. 38-41.

BOZZON R., GUETTO R., SCHERER S. 2015. *Strutture familiari e rischi di povertà in Europa*. Trento: Quaderni del Dipartimento di Sociologia e Ricerca Sociale.

CASTIGLIONI M., DALLA ZUANNA G. 2009. Marital and reproductive behavior in Italy after 1995: Bridging the gap with Western Europe? *European Journal of Population/Revue européenne de Démographie*, Vol. 25, No. 1, pp. 1-26.

DE ROSE A., VIGNOLI D. 2011. Families all'italiana: 150 years of history, *Rivista Italiana di Demografia, Economia e Statistica*, Vol. 65, No. 2, pp. 121-144.

DUFFY B., SMITH K., TERHANIAN G., BREMER J. 2005. Comparing data from online and face-to-face surveys, *International Journal of Market Research*, Vol. 47, No. 6, pp. 615-639.

EUROFOUND 2019. *Household composition and wellbeing*. Luxembourg: Publications Office of the European Union.

GUETTO R., PANICHELLA N. 2019. Family arrangements and children's educational outcomes, *Demographic Research*, Vol. 40, pp. 1015-1046.

MCLANAHAN S. 2004. Diverging destinies: How children are faring under the second demographic transition, *Demography*, Vol. 41, No. 4, pp. 607-627.

PIRANI E. 2019. L'istruzione fa la differenza? Aspetti demografici. In DE SANTIS G., PIRANI E. and PORCU M. (Eds.) *Rapporto sulla Popolazione. L'istruzione in Italia*, Bologna: Il Mulino, pp. 83-104.

PIRANI E., GUETTO R., RINESI F. 2021. Le Famiglie. In BILLARI F. and TOMASSINI C. (Eds) *Rapporto sulla Popolazione, L'Italia e le sfide della demografia*, Il Mulino, Bologna, pp. 55-82.

POWER K. 2020. The COVID-19 pandemic has increased the care burden of women and families, *Sustainability: Science, Practice and Policy*, Vol.16, No.1, pp.67-73.

PROFETA P. 2020. Gender equality and public policy during COVID-19, *CESifo Economic Studies*, Vol. 66, No. 4, pp. 365-375.

VIGNOLI D., PIRANI E., SALVINI S. 2014. Family constellations and life satisfaction in Europe, *Social Indicators Research*, Vol. 117, No. 3, pp. 967-986.

WALDVOGEL P., EHLERT U. 2016. Contemporary fatherhood and its consequences for paternal psychological wellbeing–A cross-sectional study of fathers in Central Europe. *Frontiers in Public Health*, Vol. *4*, No. 199, pp. 1-14.

SUMMARY

The wellbeing of single parents in Italy before and after the COVID-19 pandemic

In the last two decades, the Italian society has been experiencing great transformations in its family demography, with an impressive spreading of out-of-wedlock childbearing, non-marital unions, and marital separations. Among the consequences of these changes, an increasing share of children experience the dissolution of their parents' union and are raised in a non-intact family, i.e. a household where either the mother or the father is absent. After a brief description of the recent evolution of Italian families and their characteristics, in this paper we focus on single parents, considering both those living with their children and without them. Our study exploits data from an *ad hoc* survey carried out in April 2021, on a sample of single parents with cohabiting and non-cohabiting children, to evaluating the consequences of the recent COVID-19 pandemic outbreak on the economic and psychological wellbeing of single parents in Italy. Our results suggest that single parents' (low) satisfaction for their economic resources did not change dramatically, also due to their pre-existing economic difficulties, but a substantial decline in psychological wellbeing has been detected, especially among women. Nevertheless, respondents remain optimistic in their expectations for the future.

Raffaele GUETTO, University of Florence, raffaele.guetto@unifi.it
Elena PIRANI, University of Florence, elena.pirani@unifi.it
Patrizio LODETTI, University of Florence, patrizio.lodetti@unifi.it

A WELL-BEING MEASUREMENT APPROACH FOR RANKING ITALIAN MUNICIPALITIES[1]

Livia Fioroni, Valeria Quondamstefano

1. Introduction

In recent years, the measurement of multidimensional phenomena has become increasingly important in the scientific community. Socio-economic phenomena cannot be measured by a single descriptive indicator but should be represented with multiple dimensions. Phenomena such as development, poverty, well-being, etc. require multiple dimensions in order to be measured, the "combination" of several dimensions, to be considered together as components of the phenomenon (Mazziotta and Pareto, 2013). This combination can be achieved by applying methodologies known as composite indicators (Salzman, 2003; Mazziotta and Pareto, 2011; Diamantopoulos *et al*., 2008). The publication, in September 2009, of the Commission's report on the measurement of economic performance and social progress (Stiglitz Commission) was crucial to develop several studies on 'Beyond GDP' scenarios. The objective of the Commission was to identify the limits of GDP as an indicator of economic performance and social progress, consider what additional information might be needed for the production of more relevant social progress indicators, assess the feasibility of alternative measurement tools, and discuss how to present statistical information in an appropriate way. In Italy, the first report on "Equitable and Sustainable Well-being" (BES) of the Committee composed of Istat (Italian Institute of Statistics) and CNEL (National Council for Economy and Labour) was published in March 2013. It consists of a dashboard of 134 individual indicators distributed in 12 domains. In the three BES reports, published in December 2015, 2016, 2017, composite indicators have been calculated at regional level and over time for the 9 result domains, creating a unique precedent in official statistics at international level. The Italian Parliament approved in 2016 the reform of the Budget Law, in which the BES indicators will be included in the Economic and Financial Document (DEF). The new legislation also requires that by 15 February each year, the Parliament receives a report from the Minister of the

[1] The paper is the result of the authors' work; in particular, sections 1, 2 and 3 are written by Livia Fioroni, sections 4, 5, 6 and 7 are written by Valeria Quondamstefano.

Economy on the evolution of the BES indicators The project, from national, is becoming local and already several local authorities, despite not having legislative obligations, are studying the indicators of well-being in their territory.

With these assumptions, it seems necessary to calculate well-being measures for all Italian municipalities so that administrators and citizens can use them to understand and decide on better policies. Since current statistical surveys do not provide socio-economic indicators disaggregated to the level of municipalities (the Census is the only source, every ten years), it is necessary to use administrative sources, hopefully collected in information systems.

This paper aims to study a measure for quantifying and monitoring the well-being of the population in Italian municipalities. As known, this phenomenon can't be represented exclusively by economic components but also by dimensions that represent domains having demographic, social and ecological nature. This work considers well-being from a multidimensional point of view and wants to measure it for 7,959 Italian municipalities in order to highlight differences and similarities, also in time series, by using 6 elementary indicators. The reference period is from 2014 to 2017. The methodology is based on composite indicators since we wanted to make the complex phenomenon of well-being more readable. In particular, the Adjusted Mazziotta-Pareto Index (AMPI) method was used.

2. Data source

The integration between surveys and administrative data has become increasingly important over the years. In order to enhance the use of administrative sources, we use data from two different information systems: "A Misura di Comune" and "ARCH.I.ME.DE". "A Misura di Comune" (http://amisuradicomune.Istat.it/aMisuraDiComune /) is a multi-source information system published by Istat in August 2018 and continuously updated. The system reports an articulated set of indicators useful for planning, programming and management of Local Authorities, making available data that represent the economic, demographic, social and environmental conditions of the Italian territory. The project "ARCH.I.ME.DE" (Integrated Archive of Economic and Demographic Micro Data) developed in 2013 by Istat - together with Sistan – has the scope to improve the range of information through the production of longitudinal and cross-sectional elementary data collections. The project collects elementary data, obtained from the integration of administrative archives, useful for territorial and sectorial planning and for the evaluation of public policies at regional and local level, with an aim to reducing costs and statistical burden.

3. Individual Indicators

Examining the set of indicators present in the portal "A misura di Comune", it was decided to select six of them, each representative of a specific thematic dimension, describing socio-economic conditions of population in Italian municipalities for the 2014-2017 years. The six dimensions reveal individual and territorial characteristics that link material conditions (labour, economic well-being, economy on the territory, infrastructures and mobility) to quality of life (education, environment). In this case study, indicators with high data quality in terms of clarity, comparability, completeness and accuracy were chosen in a deterministic way (Daas *et al*. 2011). The description of elementary indicators and respective domain is shown below:

(A) *Not (engaged) in Education, Employment or Training (NEET)*. Domain "Education". Person in population municipal register of 15-29 years not engaged in education, employment or training for 100 persons in population municipal register of 15-29 years;

(B) *Percentage of regular employed of 20-64 years on the population of 20-64 years*. Domain "Labour". Person in population municipal register of 20-64 years with regular employment in October per 100 persons in population municipal register aged 20-64 years;

(C) *Income gaps before tax*. Domain "Economic well-being". Ratio between the total income equivalent owned by 20% of the population in municipal register with the highest income and the one owned by 20% of population in municipal register with the lowed income;

(D) *Cars on the road with emissions standard lower than Euro 4 class*. Domain "Environment". Number of cars in the Euro 0-3 class circulating for 1000 persons in population municipal register;

(E) *Entrepreneurship rate*. Domain "Economy on the territory". Number of companies for 1000 person in municipal population register;

(F) *Attractiveness Index.* Domain "Infrastructures and mobility". Ratio between the flows of individuals who work, or study inbound with respect to the total number of active in the municipality of residence.

It should be noted that administrative data not always perfectly match with data survey as there are theoretical differences. For example: the indicator of regular employment use in this study, differs from the *employment rate* measured by Labour Force Survey. This study considers only people registered in the municipalities age between 20 and 64 years old with regular employment in October. Instead, Labour Force Survey considers all employees that worked at least 1 hour on the reference week therefore, also irregular workers and people not registered at the municipality are included.

4. Composite Indicator

In order to synthesize the individual indicators into a single measure, we use the official composite indicator adopted by Istat for the BES project: The Adjusted Mazziotta-Pareto Index (AMPI) because the influence analysis demonstrates the validity compared to other methods in terms of robustness. It is a partially non-compensatory composite indicator based on a standardization of the individual indicators, at the reference time, which makes the indicators independent of the unit of measurement (De Muro *et al.*, 2011). All individual indicators are assigned equal weights and time comparisons are allowed (Mazziotta and Pareto, 2016). In fact, a downscaling of individual indicators in the range (70; 130) according to two 'goalposts', i.e. a minimum and a maximum value representing a minimum and a maximum value that represent the possible range of each variable for all time periods and all units. For the methodology and mathematical properties of AMPI see Mazziotta and Pareto (2016).

5. Descriptive data analysis

This chapter describes the exploratory analyses carried out on the matrix composed of 7,959 municipalities for the 6 domain indicators chosen from year 2014 to year 2017. In the various years, Italian municipalities have undergone territorial variations (mergers, aggregations, etc.). To overcome this problem we have considered the municipalities of the year 2017 and brought back to this geography the municipal territories of the years 2014, 2015 and 2016. With regard to the elementary indicators, where the data was missing, we proceeded to an imputation with the value of the following year for 2014, with the value of the previous year for 2017 and for an average of the immediately contiguous years for the years 2015 and 2016. Table 1 shows the results of the correlation between the 6 indicators chosen in the year 2017. The trend is the same in all years, with the highest correlation (0.69) between “Percentage of regular employed of 20-64 years on the population of 20-64 years” (B) and “Income gaps before tax” (C) in 2014 and 2015. The highest negative correlation (-0.74) is between “Income gaps before tax” (C) and “Cars on the road with emission standards lower than the Euro 4 class” (D) in 2014, 2015 and 2017. In 2016 this negative correlation, which is however the highest between the indicators in this year, is -0.73. This means that the ownership of new cars (i.e., with emission standards higher than Euro 4) depends on income and therefore on the possibility of purchasing.

Table 1 – *Correlation between selected indicators, year 2017.*

Year 2017	A	B	C	D	E	F
A	1.00					
B	-0.70	1.00				
C	-0.42	0.68	1.00			
D	0.48	-0.71	-0.74	1.00		
E	-0.34	0.37	0.34	-0.35	1.00	
F	-0.19	0.29	0.37	-0.37	0.53	1.00

Having chosen a formative measurement model for the analysis, the level of correlation between individual indicators is not important. In fact, for this approach, polarities and correlations are independent and individual indicators can have positive, negative or no correlations (Maggino, 2008). The latent variable is estimated by taking a weighted average (or other function) of the indicators that make up the concept (Shwartz *et al*., 2015).

6. Analysis of the Results

The choice of a composite index was fundamental for the treatment of data. "A composite index is a mathematical combination (or aggregation as it is termed) of a set of individual indicators (or variables) that represent the different components of a multidimensional phenomenon to be measured (e.g., development, well-being or quality of life). Therefore, the composite indices are used for measuring concepts that cannot be captured by a single indicator" (Mazziotta and Pareto, 2018).

For our analysis we needed a composite indicator valid in both time and space, therefore we chose the Adjusted Mazziotta Pareto Index (AMPI).

The aim of the work is to compare the levels of well-being in the 7,959 Italian municipalities over the four years examined, to highlight the favorable points or the critical points so as to be able to provide a tool for improving the situation.

Table 2 provides the ranking of the 10 best Italian municipalities in terms of well-being according to the elementary indicators selected for the year 2017, while Table 3 shows the ranking of the 10 worst. The trend is the same in 2014, 2015 and 2016. It confirms what is shown in the map: the best 10 municipalities are located in the North-East and North-West areas, while the worst 10 range over 3 geographical areas (North-West, Centre and South). The exception is the municipality of Santo Stefano di Sessanio in 2016 (municipality in southern Italy in sixth place in the ranking of municipalities with the highest level of well-being). Santo Stefano di Sessanio was

always among the top 100 Italian municipalities in the four years examined. This anomaly can be explained by looking more directly at this municipality in the province of L'Aquila. It has been elected one of the most beautiful villages in Italy and this has given it prominence from a national tourism point of view. In fact, with a population of about 110 inhabitants it has 18 accommodation facilities, 16 bars, shops and craft workshops and 7 restaurants. For these reasons, the basic indicator "Neither in Employment or in Education or Training" is close to zero, while the basic indicators "Attractiveness index" and "Entrepreneurship rate" are very high. It is much easier to explain the presence of municipalities in the North-West of Italy among the 10 municipalities with the worst level of well-being. These are border municipalities where the inhabitants move for work reasons to neighboring countries (especially Switzerland and France). In these cases, the levels of 'Neither in Employment or in Education or Training' and 'Percentage of regular employed of 20-64 years on the population of 20-64 years' are very high.

Table 2 – *The first 10 Italian municipalities sorted by well-being, year 2017.*

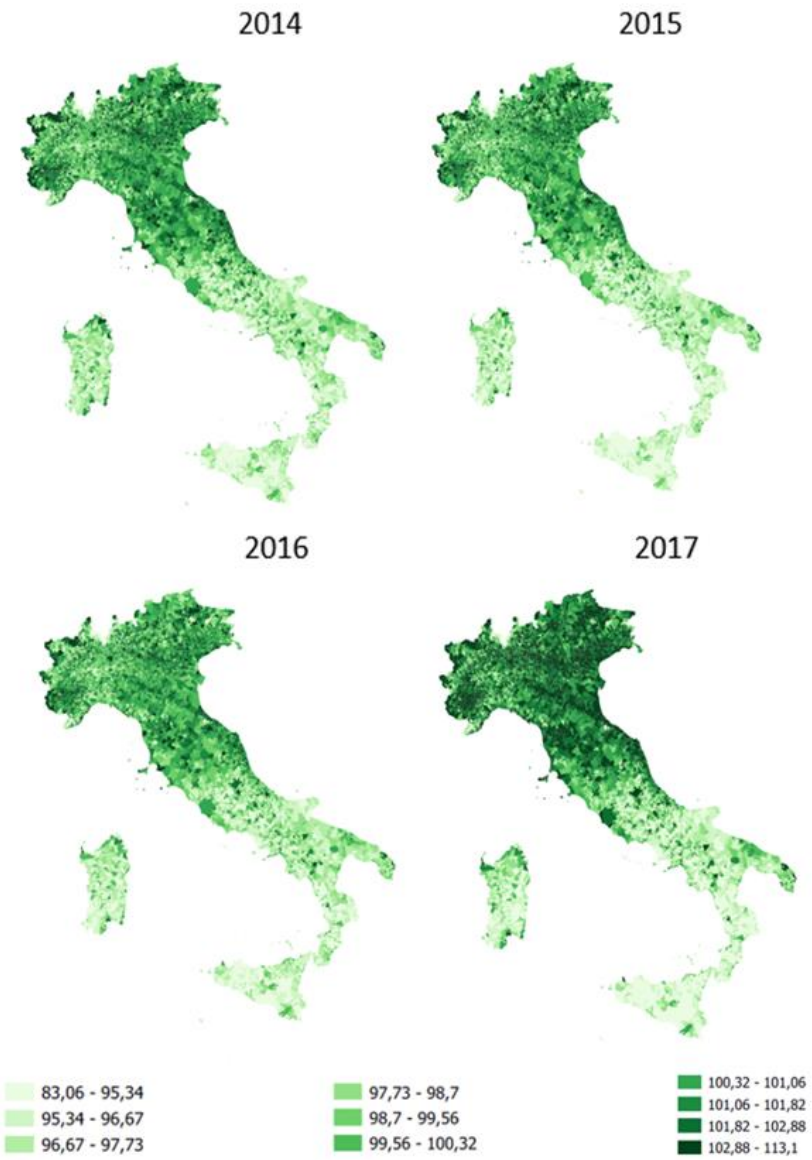
Municipality	Province	Region	Zone	AMPI	Ranking
Marmora	Cuneo	Piemonte	Nord Ovest	113.10	1
Alagna Valsesia	Vercelli	Piemonte	Nord Ovest	112.50	2
Argentera	Cuneo	Piemonte	Nord Ovest	111.85	3
Affi	Verona	Veneto	Nord Est	111.46	4
Scopello	Vercelli	Piemonte	Nord Ovest	111.31	5
Vernazza	La Spezia	Liguria	Nord Ovest	111.14	6
Limone sul Garda	Brescia	Lombardia	Nord Ovest	110.92	7
Castelmagno	Cuneo	Piemonte	Nord Ovest	110.65	8
Ceresole Reale	Torino	Piemonte	Nord Ovest	110.51	9
Gressoney La Trinité	Aosta	Valle d'Aosta	Nord Ovest	110.13	10

Table 3 – *The last 10 Italian municipalities sorted by well-being, year 2017.*

Municipality	Province	Region	Zone	AMPI	Ranking
Cavargna	Como	Lombardia	Nord Ovest	84.52	7,959
Val Rezzo	Como	Lombardia	Nord Ovest	84.67	7,958
Casalattico	Frosinone	Lazio	Centro	86.57	7,957
Giffone	Reggio di Calabria	Calabria	Sud	86.57	7,956
San Benedetto in Perillis	L'Aquila	Abruzzo	Sud	87.22	7,955
Gurro	Verbano Cusio Ossola	Piemonte	Nord Ovest	87.42	7,954
Verbicaro	Cosenza	Calabria	Sud	88.47	7,953
Nardodipace	Vibo Valentia	Calabria	Sud	88.60	7,952
Sciara	Palermo	Sicilia	Isole	89.04	7,951
Limina	Messina	Sicilia	Isole	89.33	7,950

Figure 1 shows the mapping of Italy for the years 2014, 2015, 2016 and 2017 according to the level of well-being.

Figure 1 – *Italian municipalities by value of well-being calculated by AMPI, years 2014, 2015, 2016, 2017.*



The municipalities with the highest level of well-being are shown in dark green, while those with the lowest level of well-being are shown in light green. The proposed scale is given by the deciles measured in the four years. Comparing the situation over the four years, a fairly similar trend can be seen. Municipalities with a high level of well-being are concentrated in the North of Italy, along the Via Emilia, the Adriatic coast of the Marche and in upper Tuscany, those with an average level of well-being in the rest of central Italy and those with a low level of wellbeing in the South. The maps show an increase in the level of well-being from 2014 to 2017 at national level, but also a growth in the gap between municipalities with high and low levels of well-being.

The analysis of the composite indicator for all municipalities should be a starting point both for micro studies, but also for macro studies, such as measuring the correlation with other variables/indicators available at this territorial level. Considering the correlation of the composite indicator with the size of the municipality's population, it can be seen that the size of the municipality is not as decisive as geographical location. There is no factor linked to the size of the municipality that can determine the socio-economic condition, and vice versa: these two information contributions do not influence each other. This is certainly a strength of the composite index that can explain a multidimensional phenomenon that is independent of an important variable, especially in Italy, such as the size of the municipality.

Figure 2 shows the best node for all four years under review (corresponding to the best node in 2017). In particular, it is the node corresponding to the 70 municipalities belonging to the provinces of Trento, Verona, Padua, Forlì-Cesena and Ravenna (belonging to the North-Eastern Zone, and to the Trentino Alto Adige, Veneto and Emilia Romagna Regions) with a population between 10,000 and 500,000 inhabitants (excluding the two population classes 40,000-50,000 and 65,000-80,000).

Figure 2 – *The best node of the regression tree.*

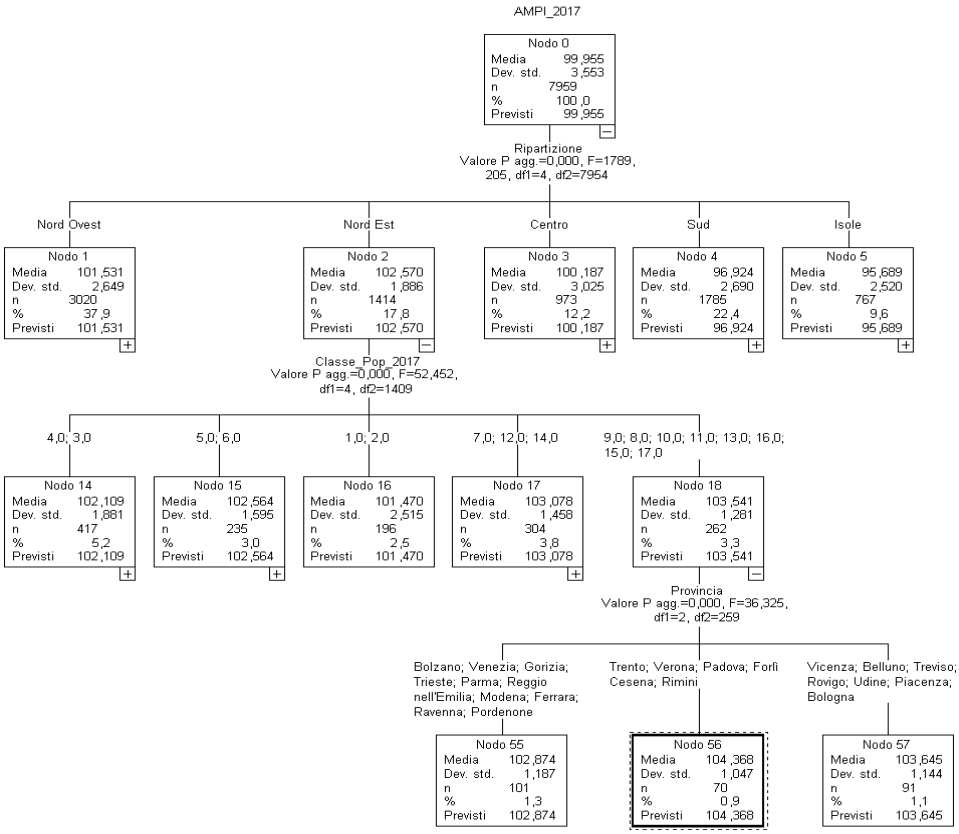
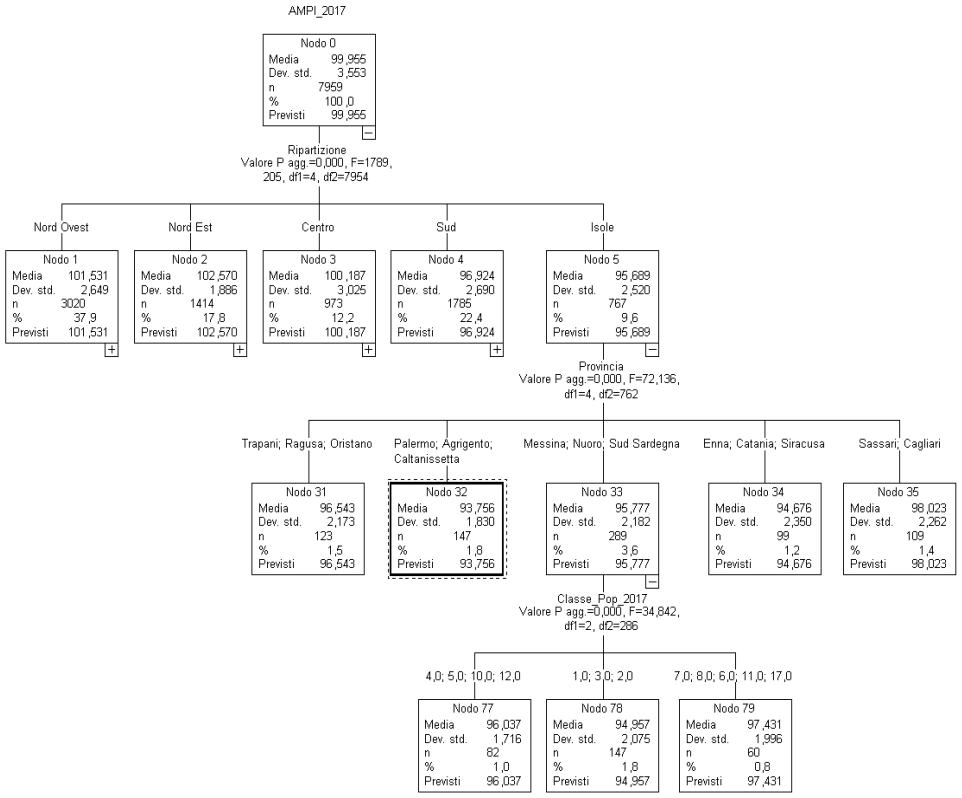


Figure 3 shows the worst node for all four years under review (corresponding to the worst node in 2017). In particular, it is the node corresponding to the 147 municipalities belonging to the provinces of Palermo, Agrigento and Caltanissetta (belonging to the South Zone, and to the Sicily Region) whatever population size.

Figure 3 – *The worst node of the regression tree.*



The method of sorting the municipalities by AMPI is interesting and can provide information on the evidence of the phenomenon. However, a more systematic approach is needed that can classify municipalities taking into account the well-being composite indicator as a function of some covariates. In this perspective, a good classification method is the regression tree, called CHAID (Chi-squared Automatic Interaction Detector). The dependent variable is the composite indicator of the well-being of the Italian municipalities; the independent variables are the three territorial levels (Zone, Region and Province) and the population size. The goal is to classify the well-being of the municipalities according to the localization on the territory and the population size (divided into 18 size classes).

7. Conclusion and next steps

The proposal of this work can be considered innovative because it considers the development over time of a composite indicator calculated on experimental statistics from administrative data at the highest possible level of disaggregation (municipal level). Moreover, this analysis has a double objective: in fact, these values can be very useful for the evaluation of the intervention's policies by local administrators and for the assessment of the administrators themselves by the citizens. This means that one of the most important phases of the research is the best practice for publishing these results so that everyone can have easy access in order to better understand the socio-economic context and decide independently through data recognized as impartial by the Community. The composite index calculated on all Italian municipalities draws a well-known geography of social and economic conditions. In fact, the peninsula seems to be divided into four parts with the conditions getting worse going south. The North-East seems to be a little better off than the North-West and the Centre-North better off than the Centre-South. The South, at the bottom of the list, lags far behind the Centre. The trend is similar in the years 2014 to 2017 (with an increase in the level of AMPI), but the gap between the best and worst municipalities is increased.

References

DAAS P., OSSEN S. 2011. Report on methods preferred for the quality indicators of administrative data sources. *Second deliverable of work package 4* of the BLUE Enterprise and Trade Statistics project.

DIAMANTOPOULOS A., RIEFLER P., etc. 2008. *Advancing formative measurement models*. Journal of Business Research 61, pp.1203–1218.

DE MURO P., MAZZIOTTA M., etc. 2011. Composite Indices of Development and Poverty: An Application to MDGs. *Social Indicators Research*, Vol. 104, pp.1-18.

MAGGINO F. 2008. La misurazione dei fenomeni sociali attraverso indicatori statistici, Aspetti metodologici. *Working Papers*, Università degli Studi di Firenze.

MAZZIOTTA M., PARETO A. 2018. Measuring well-being over time: The adjusted Mazziotta–Pareto index versus other non-compensatory indices. *Social Indicators Research*, Vol.136, pp. 967-976.

MAZZIOTTA M., PARETO A. 2016. On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena, *Social Indicators Research*, Vol. 127, No 3, pp.983-1003.

MAZZIOTTA M., PARETO A. 2013. Methods for Constructing Composite Indices: One for all or all for one, *Rivista Italiana di Economia Demografia e Statistica*, Vol. LXVII, No. 2, pp.67-80.

MAZZIOTTA M., PARETO A. 2011. Un indice sintetico non compensativo per la misura della dotazione infrastrutturale: un'applicazione in ambito sanitario, *Rivista di Statistica Ufficiale*, Vol. 1, pp.63-79.

SALZMAN J. 2003. *Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being*. Ottawa: Center for the Study of Living Standards.

SHWARTZ M., RESTUCCIA J.D., ROSEN A.K. 2015. Composite measures of health care provider performance: A description of approaches, *The Milbank Quarterly*, Vol. 93, pp. 788-825.

SUMMARY

A well-being measurement approach for ranking Italian municipalities

This paper analyses a measure capable of quantifying and monitoring the well-being of the population in Italian municipalities. This phenomenon can't be represented exclusively by economic components but also by dimensions that represent domains having demographic, social and ecological nature. This work considers well-being from a multidimensional point of view and wants to measure it for 7,959 Italian municipalities in time series, from 2014 to 2017, by using 6 elementary indicators. The methodology is based on composite indicators. In particular, the Adjusted Mazziotta-Pareto Index (AMPI) method was used. The composite index calculated on all Italian municipalities draws a well-known geography of social and economic conditions. In fact, Italy seems to be divided into parts with the conditions getting worse going south. The North-East seems to be a little better off than the North-West and the Centre-North better off than the Centre-South. The South, at the bottom of the list, lags far behind the Centre. The trend is similar in the years 2014 to 2017 (with an increase in the level of AMPI), but the gap between the best and worst municipalities is increased. These values can be very useful for the evaluation of the intervention's policies by local administrators and for the assessment of the administrators themselves by the citizens. In fact, one of the best practices for publishing these results is the follow: everyone can have access in order to better understand the socio-economic context and decide independently through data recognized as impartial by the Community.

Livia FIORONI, Istat, fioroni@istat.it
Valeria QUONDAMSTEFANO, Istat, quondamstefano@istat.it

SUSTAINABILITY AND NON-FINANCIAL DISCLOSURE: A POSET'S APPROACH

Margaret Antonicelli, Filomena Maggino, Michele Rubino

1. Introduction

In recent decades, sustainability has been at the centre of both academic and public debate with very strong repercussions, especially in the corporate sphere. The need to seek a balance between economic, environmental, and social dynamics in favour of concrete and lasting development is the basis of the reference literature. Furthermore, together with the purely economic and business aspects, it assumes the concept of well-being declined in various aspects. According to the American Management Association (2007), companies are increasingly inclined to adopt sustainability guidelines, focus on human well-being and respect for the environment, develop values and implement relevant policies and practices. The issue of corporate sustainability has grown significantly in the last 4 years also due to the obligation imposed by Directive 2014/95 / EU for large companies to disclose non-financial information. Despite the importance of this issue, the metrics to quantify and evaluate the phenomenon are many and unfortunately in most cases they are not suitable for addressing the topic in a timely, correct, and truthful manner.

The various measurement attempts and the current lack of clarity that the aspects to be detected and assessed lead to a preliminary assessment of the company profiles in which to evaluate the fundamental aspects relating to the individual sections relating to sustainability present in the NFS: environmental, social, employees, respect for human rights, diversity, and the fight against corruption.

After having deepened and analysed the reference literature and the different areas in which this phenomenon is declined, this work aims to outline the company's peculiarities in the light of the assessment of sustainability.

The relevance of the subject has a counterpart of a methodological nature as the assessment of well-being involves the use of complex systems. This apparently second-order peculiarity has a significant impact on the evaluation process, because it questions the methodological paradigm with which this type of problem is usually faced, on the playing field of socioeconomics. As a result, they open new areas of research, both theoretical and applied. The dimensions of corporate sustainability were analysed through six systems of indicators. We reduce their complexity using

synthesis obtained with the Partially ordered set. Results highlight the differences among the dimensions of sustainability. Considering what was published by Consob, in compliance with Legislative Decree no.254 of December 2016, the study analysed the 204 Italian companies that have compulsorily published the non-financial disclosures, as having at least 500 employees or at least 40 million in revenues.

The results obtained show that the methodology used, and the indicator achieved were able to better grasp the multiple characteristics of the phenomenon, without ever underestimating the individual companies. Furthermore, it is evident that the methodology used has been able to better grasp the multiple characteristics of the phenomenon, without ever underestimating the company peculiarities relating to the topic analysed.

2. Theoretical background

Over the years, the issue of sustainability has assumed an increasingly central role in the academic debate (Baumgartner, 2014; Hopkins, 2017) and is defined in different ways according to the different perspectives of investigation. From a business point of view, sustainability represents a real business management philosophy that requires the conciliation and harmonization of different categories of economic, competitive and social objectives. Sustainability is the result of an action that generates synergies from mutual interconnections, according to the model of balance between the different company dimensions. The idea is that an organization must extend its focus beyond making profits, considering the impact of its operations on the community, society and the environment following the triple fund perspective: profit, people, and planet (Carroll, 1979).

Sustainability is closely related to corporate social responsibility (CSR) (Clark, 1916; Freeman, 1994, Carroll and Shabana, 2010) as it is the goal that the company must achieve. Starting from the stakeholder theory, the issue of CSR and sustainability has been extensively explored in the literature which has highlighted how companies can reap benefits in terms of performance. First, it is the pressures exerted by stakeholders that push companies to engage in social or environmental investments. Therefore, many companies can better manage the multiple forms of relationships they have with various types of stakeholders, such as environmental and social groups, employees, and customers (Freeman and Evan, 1990; Sharma and Vredenburg, 1998). Furthermore, socially responsible firms are more attractive to investors as they more easily preserve their value in times of crisis (Schnietz and Epstein, 2005; Barnett and Salomon, 2006). Thirdly, the positive impact of CSR can derive from the fact that the investments and projects that companies implement to be more socially responsible generate long-term positive effects (Shrivastava, 1995;

Russo, Fouts, 1997; Christmann, 2000). Finally, the adoption of sustainability policies helps companies to obtain legitimacy from the various stakeholders.

Sustainability is closely associated with communication tools used by companies to disclose information of a non-financial nature. Over the years there have been various reporting tools such as the social, environmental, and sustainability reports. At the same time, many professional organizations, such as Global Reporting Initiative (GRI), have developed a set of globally accepted sustainability reporting guidelines. The Directive 2014/95/ EU with which the European legislator imposed the obligation for large companies to disclose specific non-financial information is included in the context of the various tools for disclosing non-financial information by companies (Rubino, 2020).

3. Methodology

In the measurement and evaluation of socio-economic phenomena, one of the main critical points is the identification of the most appropriate statistical methods, ensuring that the analysis respects the nature of the phenomena, both from a conceptual and methodological point of view. Therefore, we chose the synthesis method that respects the nature of this phenomena, a non-aggregative method was used, the Partially Order Set, Poset (Maggino, 2017). This approach is based on the attempt to aggregate elementary indicators into synthetic indices that allow measurement of the latent characteristic of interest and permit the construction of rankings among statistical units.

The aggregation almost always takes the form of a weighted average of indicators constructed using models with latent variables or, more simply, by identifying the weights of the weighted average through heuristic procedures or with the help of experts, assuming in fact the latent unidimensionality of the phenomenon of interest and aggregation as a method of information synthesis (Fattore, 2013).

Very often, however, the problem of constructing synthetic indices collides with the technical impossibility of applying the usual aggregative synthesis procedures. Attempting to circumvent the problem by scaling algorithms does not necessarily improve the situation; on the contrary, there is a risk of adding another element of ambiguity to the results. The reality is that it is necessary to choose between two different approaches: either to look for even more sophisticated tools to transform ordinal data into numerical variables, or to tackle the issue of evaluation on other conceptual and methodological bases that go beyond the aggregative paradigm and allow to respect the ordinal nature of some data. The two fundamental steps in this second perspective are conceptual and formal: the first consists in posing evaluation as a problem of multidimensional comparison between the "well-being" profiles of

the individual company based on an advanced level of propensity for digitization and agility, the second in identifying the mathematical tools capable of making this comparison process operational (Fattore, Maggino, Greseling, 2011a).

Ordinal data are currently available, as in the case of this study, the synthesis of multidimensional systems of ordinal data using non-aggregated methods allows the construction of measures without aggregation of the scores of the basic indicators (Alaimo, 2020). As confirmed by the literature scientifically, Poset is a reference within this approach (Maggino *et al*., 2021) since it correctly leads to the synthesis of ordinal indicators.

The Poset provides concepts and tools that adapt in a very natural way to the needs of synthesis. This approach is based on profiles that represent the combinations of the scores of each statistical unit in the basic indicators considered. Therefore, Poset respects the nature of the data, and the construction of the synthetic indicator does not require any operation on the basic indicators (normalization, aggregation). Poset are the mathematical structure, $P = (X, \trianglelefteq)$, composed by a set X endowed with a partial order relation $\trianglelefteq$, i.e. a binary relation satisfying the properties of reflexivity, antisymmetry and transitivity (Davey and Priestley 2002; Neggers and Kim 1998; Schröder 2002):

1. $x \trianglelefteq x$ for all $x \in X$ (reflexivity);
2. if $x \trianglelefteq y$ and $y \trianglelefteq x$ then $x = y$, $x, y \in X$ (antisymmetry);
3. if $x \trianglelefteq y$ and $y \trianglelefteq z$, then $x \trianglelefteq z$, $x, y,z \in X$ (transitivity).

Let $k = |X|$ be the cardinality of X, the incidence matrix is a $k \times k$ boolean matrix Z summarizing the comparability relation $\trianglelefteq$, whose element for all $xi, xj \in X$

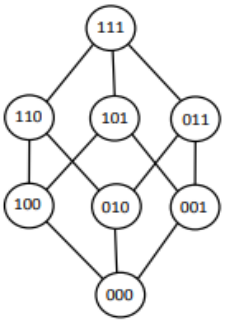
$$z_{ij} = \begin{cases} 1 \; if \; xi \trianglelefteq xj \\ 0 \; otherwise \end{cases}$$

for all $xi, xj \in X$. Consider two elements $xi , xj \in X$, the element xj covers the element xi , $xi \prec xj$, if xj dominates xi , $xi \trianglelefteq xj$, and there is no other element $z \in X, z \neq xi , xj$ such that $xi \trianglelefteq z \trianglelefteq xj$. A directed acyclic graph describes the cover relation $\prec$.

The Hasse diagram is the graphical representation of this graph, with the top-down orientation replacing the arrows in representing the edge direction.

The interpretation of the diagram is immediate: each node corresponds to a profile and two profiles are ordered (comparable) if and only if they are connected by a descending path (or, equivalently, ascending) (Fig.1). The pairs of profiles that are not connected by a descending/ascending path are called incomparable.

Figure 1 - *Example of a Hasse diagram of a Poset defined on three binary variables.*



Two nodes connected by a path are comparable by transitivity. An extension of P = (X,⊴) is a poset Pe = (X,⊴e) on the same set X but equipped with a partial order relation ⊴e extending the relation ⊴, therefore all the pairs of elements comparable in ⊴ are comparable in ⊴e and some pairs comparable in ⊴e are not comparable in ⊴. A linear extension of P is an extension of P where all the elements of the set X are comparable, therefore it is a complete (or linear) order obtained extending the initial poset. A poset usually has more than one linear extension. Let ΩP be the set of all the linear extensions of P. The mutual ranking probability (MRP) matrix of P is a k × k matrix MP = (m_{ij}), where m_{ij} is the fraction of linear extensions in ΩP such that the element x_i is dominated by the element x_j. We use posets to define the structure of comparability among units of multi-indicators systems (Fattore, 2013). Once the structure is defined, we can analyze it through mathematical tools.

4. Data

Considering all the Italian companies that in 2020 published the non-financial statement, in this work it was decided to use the sample made up of 204 Italian companies that in 20 20 compulsorily published the non-financial communication as having at least 500 employees or at least 40 million in revenues. As previously mentioned, the identification of the companies obliged to draw up this type of declaration is well defined and determined by Legislative Decree 254 of 2016. Furthermore, always in compliance with the legislative decree, the elements relating to each of the 6 mandatory areas of analysis of this declaration. Subsequently, after an initial content analysis on each individual declaration made in a computerized manner, a second (control and correction) was carried out manually.

The results obtained through the analysis of the content subsequently made it possible to transform the results into ordinal variables, with the attribution of the

following values: 0 in case of absent article, 1 in case of article treated only textually, 2 in case of article treated numerically, 3 in case of article treated both in textual and numerical way. 44 items were identified, divided as follows:

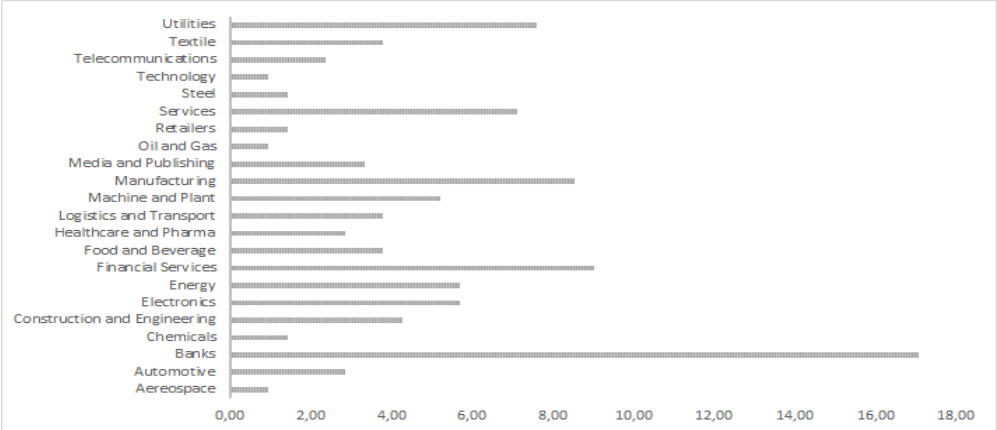
- environmental section: environmental policy, materials, water, energy, hazardous waste, external environmental certifications, compliance with environmental regulations, management of environmental risks;
- social section: involvement of local communities, evaluation of suppliers according to social criteria, negative social impacts on the supply chain and marketing policies, customer privacy, compliance with laws and regulations on social matters, evaluation of the impacts on health and safety by categories of products and services, political contributions, share of senior managers hired by the local community;
- section relating to employees: hiring and turnover, expected benefits for employees, adoption of parental leave policies, minimum notice period, occupational health and safety management system, identification of dangers and associated risks, promotion of employee health, average annual hours of training per employee, employee skills refresher and transition assistance programs, percentage of employees receiving periodic performance and career development reviews, career advancement prospects, retirement plans and defined benefits and other pension plans, ratios between a new employee's standard wage by gender and local minimum wage;
- section on diversity: gender diversity in governing bodies and among employees, ratio of basic salary and remuneration of women to men, composition of the board of directors and its committees by gender, selection criteria for members of the board of directors
- human rights section: incidents of discrimination and corrective measures taken, child labour cases, security personnel trained in human rights policies or procedures, investment agreements and significant contracts that include human rights clauses, management of internal rights risks human;
- section on the fight against corruption: operations assessed for the risks associated with corruption, communication and training on anti-corruption policies and procedures, confirmed episodes of corruption and actions taken, adoption of organizational models.

5. Result and discussion

Regarding some parameters of utmost interest in a statistical business approach, it's important to shows the main characteristics of our analysed sample. Before proceeding with the Poset analysis, it is essential to dwell on some purely corporate parameters of extreme interest for the purposes of the analysis. As mentioned, several times in the corporate literature, the sector is a fundamental parameter, especially in NFDs.

As can be seen from Figure 2, the most represented sector is the banking sector (17%), followed by those relating to financial services (9%), manufacturing (8.6%) and utilities (7.6%). This aspect is extremely important for the purposes of the analysis as it highlights how the banking and finance sectors, in addition to being the largest in terms of employees and revenues, are those whose transparency and sustainability declined in economic but above all social and ethical terms, it appears to be fundamental for entrepreneurial activities.

Figure 2 - *Distribution companies by sector, 2020*



Another extremely interesting variable is the region where the company is located. As you can see from Figure 3, the northern regions are those with the greatest presence of large companies in terms of revenues and employees and, therefore, obliged to draw up the non-financial disclosure.

This is a symptom of how even non-financial reporting, together with purely economic and entrepreneurial issues, have a strong territorial connotation in common, which confirms the presence of large and important companies in northern Italy.

Figure 3 - *Distribution companies for Italian regions, 2020.*

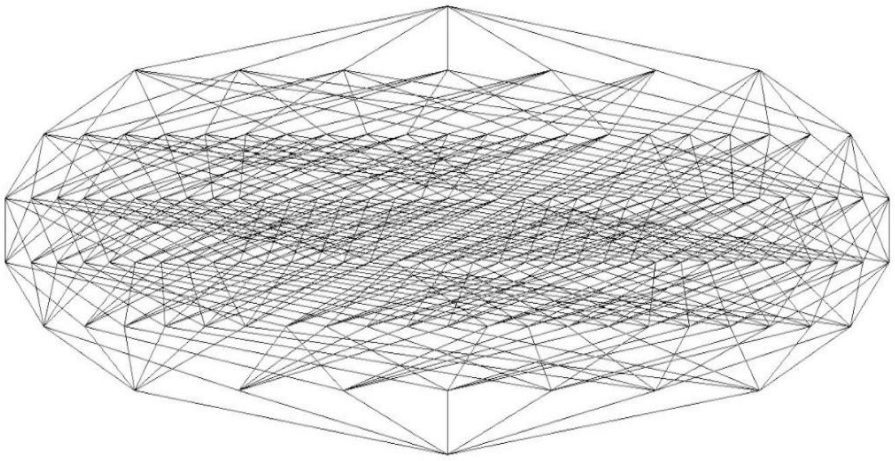
Consequently, it is essential that all the companies analysed are partially ordered differently in the various environmental, social, personnel, diversity, human rights, and fight against corruption dimensions. After defining the characteristics of the companies included in the sample in terms of parameters considered interesting, it is important to define the incidence matrix.

Then, we proceed to the construction of the Hasse diagram. This approach provides, given the values measured for each company and for each parameter analysed, to identify all the possible profiles that can be outlined as the sustainability indicator defined in our study.

By combining the six areas of analysis, in creating the sustainability index defined through the non-financial disclosure, 13,479 possible profiles have been identified. Each individual profile is the result of the combination of the value that the individual company has obtained for each of the six index areas analysed. In detail, items were analysed for each individual element of the sustainability index, as imposed by Legislative Decree 254 of 2016.

As can be seen from Figure 4, relating to the Hasse diagram for the sustainability index, the upper part shows the high and medium-high profiles, while the worst and medium-low profiles are shown below.

Figure 4 - *Hasse Diagram, Sustainable Index.*



In this type of statistical approach, the number of possible profiles that can be generated by combining the responses to each of the elements that make up the indicator does not coincide in most cases with the actual number of profiles that appear in the empirical analysis.

The presence of a high number of profiles identified for the sections relating to environmental, social and employee aspects shows a significant and not positive heterogeneity between companies in transparency and management towards stakeholders and policy makers.

A much less heterogeneous situation is found for the issue of gender equality addressed through the section on diversity. Regardless of the sector in which the company operates, unfortunately, the differentials in terms of female presence in decision-making bodies are still very high even if companies have clearly expressed their desire to reduce this gap through ad hoc actions. The wage gap is smaller where much has already been done but much still needs to be implemented.

The situation also seems optimal about the discussion and approach to the fight against corruption. Companies reach optimal levels of management of this aspect also because they are supported by various supporting laws. the characterizing aspect of this section is the strong link with the sector in which the company operates. For the banking and financial sector, the aspects linked to the fight against corruption are deeply felt and the management of the related risk is very articulated and well structured.

Apart from the difference between the possible profiles that we could have had from a theoretical point of view (given by the linear combination of the elements)

and the real profiles identified in our study, the apparent difference in the results between the six areas depends on an important difference in the parameters analysed and, as we will see later, on a situation of important incompatibility (Maggino, 2017) between companies in relation to the different area analysed.

Another essential element of the analysis is the identification of the thresholds. This phase of the analysis inevitably has a subjective component which, contrary to popular belief, is not necessarily a negative component. As this is a new and innovative approach, there is a lack of solid scientific literature on universal criteria for threshold identification. In consideration of this, we decided to adopt the approach defined by Arcagni et al. (2019). Analysing the poset, we identify for each of them a subset πl of profiles that are incomparable to each other (lower threshold); all profiles in πl or below an element of it are classified as dissatisfied. At the same time, we define another subset πc (upper threshold) such that the profiles in c or above one of its elements are identified as fully satisfied. The identification of medium-high and medium-low thresholds derives mainly from the fact that companies (especially due to the diversity of the sector in which they operate and therefore a very different approach to the six areas analysed) very often fail to reach high levels in the treatment of the various sections of the NFD.

6. Conclusion

Sustainability, and in particular the aspects relating to the business world, declined in all its aspects, is currently a very important issue in the life of a company both in economic terms and in terms of attractiveness and transparency towards stakeholders, customers, and policies. makers. However, despite the different legislative guidelines that regulate these aspects, especially those at the basis of the NFD, the subjectivity of management granted to companies still brings a lot of heterogeneity in the treatment and management of these aspects.

This study wanted, in this first phase using Poset, to highlight this extreme heterogeneity even among very similar companies in terms of activity and sector.

The next step will be the construction of a composite indicator which is also useful for assessing the impact of these essentially subjective and social aspects with company economic and budget parameters.

References

ARCAGNI A., BARBIANO DI BELGIOSO E., FATTORE M., RIMOLDI L. 2019. Multidimensional analysis of deprivation and fragility patterns of migrants in

Lombardy, using partially ordered sets and self-organizing maps, *Social Indicators Research*, Vol. 141, No. 2, pp. 551–579.

BAUMGARTNER, R.J. 2014. Managing corporate sustainability and CSR: A conceptual framework combining values, strategies and instruments contributing to sustainable development, *Corporate Social Responsibility and Environmental Management*, Vol. 21, No. 5, pp. 258-271.

BARNETT M.L., SALOMON R.M. 2006. Beyond dichotomy: The curvilinear relationship between social responsibility and financial performance, *Strategic Management Journal*, Vol. 27, No. 11, pp. 101-122.

CARROLL A.B. 1979. A three-dimensional conceptual model of corporate performance, *Academy of Management Review* 4, No. 4, pp. 497-505.

CARROLL, A.B. and SHABANA, K.M. 2010. The business case for corporate social responsibility: A review of concepts, research and practice, *International journal of management reviews*, Vol. 12, No. 1, pp. 85-105.

CLARK, J. M. 1916. The changing basis of economic responsibility, *Journal of political economy*, Vol. 24, No. 3, pp. 209-229.

CHRISTMANN P. 2000. Effects of 'best practices' of environmental management on cost advantage: The role of complementary assets, *Academy of Management Journal*, Vol. 43, No. 4, pp. 663-680.

FATTORE M. 2013. La valutazione statistica del benessere: ordinamenti parziali e nuovi paradigmi metodologici, *Statistica & Società/Anno 2, N. 2/Lavoro, Economia, Finanza.*

FATTORE M., MAGGINO F., GRESELIN F. 2011a. Socioeconomic evaluation with ordinal variables: integrating counting and poset approaches, *Statistica & Applicazioni, Partial Orders in Applied Sciences*, Special Issue 2011, 31- 42.

FREEMAN, R.E. 1994. The politics of stakeholder theory: Some future directions, *Business ethics quarterly*, Vol. 4, No. 4, pp.409-421.

FREEMAN R.E. and EVAN W.M. 1990. Corporate governance: A stakeholder interpretation, *Journal of behavioral economics*, Vol. 19, No. 4, pp. 337-359.

MAGGINO F., 2017. Developing indicators and managing the complexity, In: Maggino F, (Ed) *Complexity in society: from indicators construction to their synthesis.* Cham: Springer; pp. 87–114.

MAGGINO F., BRUGGEMANN R., ALAIMO L. S. 2021. Indicators in the framework of partial order, In Brüggemann R, Maggino F, Suter C, Beycan T, (Eds) *Measuring and understanding complex phenomena: indicators and their analysis in different scientific fields*. Cham: Springer; pp. 17–29.

HOPKINS, M. 2017. *CSR and Sustainability. From the Margins to the Mainstream: A Textbook*. London: Routledge.

RUBINO, M. 2020. *La disclosure sui rischi nell'informativa non finanziaria: Evidenze empiriche nel contesto europeo*. Milano: Egea.

RUSSO M.V. and FOUTS P.A. 1997. A resource-based perspective on corporate environmental performance and profitability, *Academy of Management Journal*, Vol. 40, No. 3, pp. 534-559.

SCHNIETZ K.E. and EPSTEIN M.J. 2005. Exploring the financial value of a reputation for corporate social responsibility during a crisis, *Corporate Reputation Review*, Vol. 7, No. 4, pp. 327-345.

SHARMA S. and VREDENBURG H. 1998. Proactive corporate environmental strategy and the development of competitively valuable organizational capabilities, *Strategic Management Journal*, Vol. 19, No. 8, pp. 729-753.

SHRIVASTAVA P. 1995. The role of corporations in achieving ecological sustainability, *Academy of Management Review*, Vol. 20, No. 4, pp. 936-960.

SUMMARY

Sustainability and non-financial disclosure: a Poset's approach

In recent decades, sustainability has become one of the main aspects of Non-Financial Disclosure (NFD), a type of mandatory reporting for large companies in terms of turnover and number of employees, reporting on social and environmental aspects. There are currently many metrics to evaluate this aspect, but in most cases, they are not suitable for addressing a phenomenon of such great importance in a timely and truthful manner. The relevance of the subject has a counterpart of a methodological nature since the assessment of well-being involves the use of complex systems of ordinal variables. This impacts on the evaluation process, because it calls into question the methodological paradigm with which this type of problem is usually faced. The statistical analysis carried out is based on two extremely important phases: both manual and computerized content analysis and the Partial Order Set.

The sample defined by Consob, in compliance with Legislative Decree no. 254 of December 30th 2016, is made up of 204 Italian companies obliged to register non-financial communication in 2020. The results obtained show that the methodology used has been able to better grasp the multiple characteristics of the phenomenon, without underestimating the individual companies in terms of different discussion of aspects related to sustainability.

Margaret ANTONICELLI, Università La Sapienza,
margaret.antonicelli@uniroma1.it
Filomena MAGGINO, Università La Sapienza, filomena.maggino@uniroma1.it
Michele RUBINO, Università Lum Giuseppe Degennaro, rubino@lum.it

USE OF OPEN STREET MAP FOR ACCIDENT INVESTIGATION ON THE ROAD AND MOTORWAYS NETWORKS

Marco Broccoli, Silvia Bruzzone

1. Road Safety performance Indicators and Big data use

Road Safety Performance Indicators (RSPI) give a multidimensional approach for accidents investigation concerning roads, vehicles and persons involved. Combining the use of statistical surveys, administrative geographical information systems (GIS) and Big Data (BD) sources, the result gives new elements on planning infrastructure solutions, applying policies to reduce deaths and serious injuries, reducing social costs on collectivity and estimating efficiency and effectiveness of safety initiatives. Preventing road trauma on public roads is a core responsibility for government, its agencies and stakeholder. It requires a common and shared responsibility.

Nowadays there is a clear information bias as regards the appropriate reference denominators to be placed as basis in construction of statistical indicators linked to road accidents. Resident population is used as a common proxy for exposed at risk in a specific geographical area, but not always an appropriate solution, especially in the light of the seasonal nature of road accidents and concentration, in some periods of the year, in specific locations. Vehicle fleet (Automobile Club of Italy - ACI) can be another administrative source that gives a more accurate information, but the characteristic of the phenomenon implies a deductible distortion on measures due to the mobility of road users. The length of the road network (from Open Street Map) gives for sure a consistent first set of information concerning the different territories.

This information is not available from official statistics at national level, although there are archives and detailed road graphs for each municipality, province and region, a harmonized and systematic national road registry has not been established. The first output of the project is the focus on the exploitation of existing administrative sources, the scouting of new sources and the analysis of integrated and auxiliary data.

The aim of the project is to calculate road accidents rates, mortality and harmfulness indexes, comparing these measures to the correspondent length in meters of carriageway by road direction from the Open Street Map. This approach has been used for the first time, by the authors, to build road accidents indicators and adopted by the “Experimental statistics” at Istat (Broccoli and Bruzzone, 2021).

Although Open Street Map, in fact, is cited as a source of data in many scientific articles, the main purpose was for GIS-Based Traffic Simulation or other aims (Zilske and Neumann, 2011). While the product represents a first result, the final purpose of the project is to expand statistical information with the supply of traffic flows (vehicles / km) on the national road network. This would make it possible to calculate the probability of being involved in the accident, taking into account the different exposure to risk of accidents.

2. The use of GIS systems for a graphical representation

GIS is a geographic system designed to capture, store, manipulate, analyses, manage and present spatial or geographic data. GIS applications are tools that allow users to analyse spatial information, edit data in maps and present the results of all these operations. In order to relate information from different sources, GIS uses a spatial location as the key index variable (key reference by position). Just as a relational database containing text or numbers can relate many different tables using common key index variables, GIS can relate otherwise unrelated information by using location as the key index variable. This key characteristic of GIS has begun an alternative frontier on producing statistical information. Any variable that can be located spatially using an x, y, and z coordinates, representing longitude, latitude, and elevation, respectively. These GIS coordinates may represent other quantified systems of territories (polygons), road networks (lines) and point of traffic (points).

"Join attributes by location" is the algorithm that takes an input vector layer and creates a new vector layer that is an extended version of the input one, with additional characteristics in its attribute table. The additional attributes and their values are taken from a second vector layer. A spatial criterion is applied to select the values, from the second layer, added to each feature from the first layer in the resulting one.

2.1. Istat Census Map localities

The Istat Census Map localities used with the GIS system are provided for the following administrative units: Regions; Provinces; Municipalities (yearly updated); Localities (only 2011 Census). An upgrade of Census Map localities shapes referred to 2011 to municipalities 2016 has been done. It has been built a link table with the aggregation of all 8090 local administrative units' territory, at 2011, to the 7998 municipalities at 2017, included in the Italian territory.

The choice of the localities shapes is due to the harmonization need of the roads graph to the "road type classification" used by the road accidents survey. The localities classification includes four voices:

1. Urban areas
2. Small inhabited areas
3. Productive areas
4. Wide spread houses

As regards the definitions, the Istat[1] descriptions are reported below.

Localities - A more or less wide area, usually known by name, on which one or more houses are grouped or scattered; there are two types of localities: inhabited localities and productive locations. The borders of the inhabited localities (center and inhabited area) are the external limit of the buildings placed at the edges of a grouping of at least fifteen buildings. The confines of the localities include gardens and other accessories areas of the considered buildings, non-built neighboring areas not included (such as fields with or without crops). Buildings located at a distance of more than 70 meters, within built-up areas and 40 meters for small-inhabited areas excluded. If the buildings included in the new locality are adjacent or in proximity (to a max distance of 140 meters for urban areas and 60 meters for small inhabited areas) of transport infrastructures or hydrographic limits, the location border will be extended to the middle of these elements.

Urban areas - Aggregate of contiguous or near houses with roads, squares and similar, or however areas characterized by services or public activities (school, public office, pharmacy, shop or similar), detecting a social life and, generally, also a place of collection for the inhabitants of the neighboring areas. The places of tourists meetings, houses, hotels and similar used for the vacation, inhabited seasonally, are considered as temporary inhabited centers too.

Small-inhabited areas is an area without the place of collection, characteristic of the urban area. It is based on a group of at least fifteen contiguous and near buildings, with at least fifteen families, with roads, paths, squares, farmyards, small gardens and similar, as long as the distance between the buildings does not exceed thirty meters and it is lower than the distance between the center and the nearest of the houses clearly scattered.

Productive areas - Extra-urban area not included in the centers or residential areas with more than 10 local units, or with a total number of employees' upper than 200. The local units are contiguous or close, with roads, squares or similar, or

[1] Istat - Basi territoriali e variabili censuarie https://www.istat.it/it/archivio/104317 ; Istat - Descrizione dei dati geografici e delle variabili censuarie delle Basi territoriali per i censimenti: anni 1991, 2001, 2011 https://www.istat.it/it/files//2013/11/Descrizione-dati-Pubblicazione-2016.03.09.pdf

anyway in a continuous line, not exceeding 200 meters; the minimum area must be 5 hectares.

Wide spread houses - Houses scattered in the municipal territory at a distance not enough to constitute a built-up area.

2.2. Open Street Map and road arch type

Open Street Map (OSM)[2] is a collaborative project aimed on creating free content maps of the world. The project aims at a collection world of geographical data, with the main purpose of creating maps and cartography. The key feature of the geographic data present in OSM is having a free license, the Open Database License. It is therefore possible use them freely for any purpose with the only constraint of mentioning the source. Everyone can contribute by populating or correcting data. The maps are created using the data recorded by portable GPS devices, aerial photographs and other free sources. Most of the Android and iOS GPS navigation software on portable devices are powered by OSM as WisePilot, Maps.me, NavFree, Scout etc.

The Open Street Map vector layers, used in this work, daily updated and free downloadable data, are the following:

- Road graph;
- Point of traffic (POT);
- Added shape: Buildings; Use of the land; Natural; Places; POWF (Point of Worship); POIS (Point of interest); Railways; Transport; Water; Water ways.

Although OSM it is an Open Source tool based on information from a community, the product provides data to be considered reliable and consistent, so much that the major part of GPS Android and iOS navigation software on portable devices are powered by OSM, for example WisePilot, Maps.me, NavFree, Scout, etc. Table 1 contains the list of different type of road arch by Open Street Map.

An innovative method of measuring the length in meters of the road graph is given by the information on the number of carriageways of each road arch of OSM. In the future, in order to provide more detailed information, the use of the number of lanes containing each carriageway is being evaluated.

[2] OpenStreetMap provides geographic data on thousands of websites, mobile and hardware devices. OpenStreetMap is built by a community of mappers, who contribute, update and monitor data on roads, cafes, railway stations and much more, all over the world - OSM: http://www.openstreetmap.org/about

Table 1 – *Open Street Map road arch classification (a)*

Road type	Road type Description
Secondary link	The link roads (slip roads/ramps) leading to/from or from/to a secondary road or lower class highway.
Tertiary	Roads of local rank. They connect smaller municipalities together. In urban areas, they are side roads to primary and secondary roads with a medium flow of traffic.
Tertiary link	The link roads (slip roads/ramps) leading to/from or from/to a tertiary road or lower class highway.
Unclassified	Classification for some extra-urban road.
Residential	Roads in a residential area, which serve as an access to housing, without function of connecting settlements.
Living Street	Residential road where pedestrians have legal priority over cars, speeds are kept very low.
Pedestrian	Pedestrian areas (roads or squares in urban areas), accessible mainly or exclusively to pedestrians.
Service	Access roads or internal service areas, beaches, camping, industrial areas, shopping centers, parking places etc.
Track	Roads for mostly agricultural or forestry uses
Cycle way	Cycle paths on dedicated carriageway, mainly or exclusively for cycling tourism.
Footway	Paths mainly/exclusively for pedestrians. This includes walking urban tracks, paths in a public park and footpaths
Path	Paths not structured for a public use
Steps	Stairs in steps, exclusively accessible by pedestrians
Unknown	Not classified

(a) Road Archs by Open Street Map – update 16/2/2017.

2.3. Classification of OSM road arches and road accidents localization

The arch road types selected to calculate the indicators analysed are referred to the motorized vehicles flow: motorway, trunk, primary, secondary, tertiary, unclassified, residential, living street, motorway link, trunk link, primary link, secondary link, tertiary link, service, unknown. Pedestrian, track, track_grade, bridleway, cycle way, footway, path, steps are not object of the survey definition.

To build road accidents indicators, with denominator represented by the arches length in Open Street Map, we built a "bridge matrix" between road categories, classified by functional road type, used by OSM, and the categories linked to the

roads holder, used by Istat road accidents survey. This methodology can be defined as a systematic classification technique.

The systematic classification of road arches (Open Street Map), classified in the categories used by Istat, has been modified, in respect of a first release (Broccoli and Bruzzone 2021). A new analytical classification has been adopted, using a more refined technique of attributing single road arches, about three and a half million in total (OSM), to the Istat classification groups (table 2). The operational criterion applied provides the roads classification, through the textual analysis of the Name and Reference attributes, according to the different classes of road arch and spatial attribution of the location type.

Table 2 – *Bridge coding table between roads arches classification by OSM, localities and road type (b)*

Road Arches classification by Open Street Map	Localities at Census 2011			
	Urban areas + Small		Productive areas + Wide	
	Road Localisation by Road accidents survey			
	Motorways	Urban Roads	Motorway	Rural Roads
Motorway	x		x	
Trunk	x		x	
Primary		x		x
Secondary		x		x
Tertiary		x		x
Unclassified		x		x
Residential		x		x
Living Street		x		x
Motorway Link	x		x	
Trunk Link	x		x	
Primary Link		x		x
Secondary Link		x		x
Tertiary Link		x		x
Service		x		x
Unknown		x		x

(a) Istat computing.

In order to use 2017 data, improved by the new technique and to guarantee the comparability to 2016 data, already disseminated, the new 2017 indicators were also completed by a recalculation of the provincial rankings, adopting the 2016 systematic classification (Examples: Main Roads maps Chart 1-4).

Chart 1 – *Motorways. Year 2017 (Istat: analytical classification).*

Chart 2 – *Main Roads. Year 2017 (Istat: analytical classification)*

Chart 3 – *Regional Rural Roads. Year 2017 (Istat: analytical classification)*

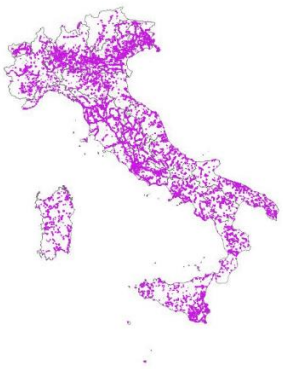


Chart 4 – *Rural Roads. Year 2017 (Istat: analytical classification)*

3. The indicators on road accidents

The proposal to calculate new road accidents indicators was born to provide, new and more suitable measures, for the risk and probability of being involved in a road accident. Even if Resident Population and Vehicles fleet, traditional sizes used, are both denominators considered, at present, *proxy* for the risk exposure to the road accidents, it is clear that they present some critical points. As regards the data updating (numerators of indicators) used to build indicators on the road accidents risk, they refers to Istat "road accidents survey", concerning all road accidents

resulting in deaths or injuries in 2017, involving at least a vehicle circulating on the national road network and documented by a Police authority[3].

As cited before, the Resident Population[4], not always is an appropriate solution to build road accident indicators, because the seasonality of accidents and the concentration in some specific places during the year. The Vehicles fleet by registration province (Automobile Club of Italy – Public Register of Vehicles PRA)[5] provides more suitable information than the population, but it does not consider the mobility of users through the country.

The length in meters of carriageway by road direction of road arch (by OSM)[6] provides, instead, a set of more coherent information referred to different areas, because independent by the seasonality and the mobility users influence.

For this reason, Istat decided to investigate the new frontier of big data and to use the new data source to build road accidents indicators and to start the consolidation process to adopt officially the calculation.

The information on the roads length, provided by OSM, is not available at national level from an institutional source, since, although archives and detailed road graphs by single municipality, province or region exist, a national and harmonised road register has been not created yet.

3.1. Road accident indicators: weighting with Traffic Points (PoT) from Open Street Map

The length in meters of carriageway per direction of the road arch (from Open Street Map), used to calculate the indicators, provides a first coherent set of information relating to the various territories. Although the product constitutes a first usable result, the project's most ambitious goal is to estimate the real traffic flows (vehicles/Km) on the national road network, as widely documented in literature, suitable indicator (Farchi et al 2006; Hakkert et all 2002). This would make it possible to calculate the probability of being involved in a road accident and therefore of real exposure to the risk of accident.

To fill this information gap, other additional information has been obtained from the rich source Open Street Map, in particular, data on the traffic points detected on the road arches (PoT Point of Traffic). A monthly information, downloadable from OSM, is available and the data refers to points over an arch, in which is detected an intensity of traffic. The proposed new road accident indicators, "weighted" with the

[3] Survey on road accidents resulting in death or injuries http://siqual.istat.it/SIQual/lang.do?language=UK
[4] Italian Resident Population - 31/12/2017.
[5] ACI Vehicle fleet (Automobile Club of Italy) all motorized vehicles except the trailers - 31/12/2017.
[6] GIS (Census Map Open Street Map road graph 1/1/2017) by length in meters per carriageway.

information on traffic intensity, was built considering, as a central element, the kilometers of carriageway with the presence of a traffic point on the arch.

The authors calculated the length of roads, considering the extension, in meters, of carriageway with arches on which including the presence of traffic points. The relative frequency (f) of the road length with the presence of traffic points on an arch (by province and type of road) and the complementary frequency (1-f) of arches without point of traffic has been calculated too.

The number of accidents, vehicles involved, deaths and injuries per 100 kilometers of carriageway in the province, excluding the effect on accidents caused by the presence of traffic points on the roads has been processed. The calculation was done by multiplying the value of the indicators for road length and the frequency (1-f) referred to arches without traffic points on the roads of the province.

With reference to the notation used for the calculation, the individual items are defined as follows:

LA Tot $_{i,j,k}$ = Total Length of Arches in meters in the province (i), by functional class of the Open Street Map Arch (j) and Type of locality (k);

LA PoT $_{i,j,k}$ = Length of Arches in meters with presence of Points of Traffic (PoT) in the province (i), by functional class of the Open Street Map Arch (j) and Type of locality (k);

LA Tot $_{i,k,l}$ = Total Length of Arches in meters in the province (i), by Type of locality (k) and Type of road by Istat classification (l);

LA PoT $_{i,k,l}$ = Total Length of Arches in meters in the province with presence Points of Traffic (PoT) in the province (i), by Type of locality (k) and Type of road by Istat classification (l);

The total length of the national roads, referred to arches length, with the presence of Points of Traffic (PoT) or in total, is given by the expressions below. Total Length of Arches with Points of Traffic (PoT):

$$\text{LA PoT}_{\text{tot}} = \sum_{i=1}^{111}\sum_{j=1}^{27}\sum_{k=1}^{4} \text{LA PoT}_{\text{i,j,k}} \quad or \quad \sum_{i=1}^{111}\sum_{k=1}^{4}\sum_{l=0}^{9} \text{LA PoT}_{\text{i,k,l}} \tag{1}$$

$$\text{LA Total}_{\text{tot}} = \sum_{i=1}^{111}\sum_{j=1}^{27}\sum_{k=1}^{4} \text{LA Totale}_{\text{i,j,k}} \quad or \quad \sum_{i=1}^{111}\sum_{k=1}^{4}\sum_{l=0}^{9} \text{LA Total}_{\text{i,k,l}} \tag{2}$$

The percentage weight of the Arches Length in meters, with the presence of Points of Traffic (PoT), in the Province (i), by functional class of Open Street Map Arch (j) and Locality type (k) out of the total length is given by the following expression:

$$p\ \text{LA PoT i, j, k} = \frac{\text{LA PoT i,j,k}}{\text{LA Tot i,j,k}} * 100 \quad (3)$$

The percentage weight of the Arches Length in meters, with the presence of Points of Traffic (PoT), at national level, out of of the total length, is given, at last, by the following expression:

$$\text{p LA PoT}_{\text{tot}} = \frac{\text{LA PoT}_{\text{tot}}}{\text{LA Total}_{\text{tot}}} \quad (4)$$

where: **i**=1,…, 103, 108, 109, 110, 111 (Istat Code of province); **j**=1,.....,27 (Functional class of the Open Street Map Arch); **k**=1,.....,4 (Type of locality: 1=Urban areas, 2=Small inhabited areas, 3=Productive areas, 4=Wide spread houses); **l**= 0,……,9 (type of road by Istat classification).

In 2017, there are about 220,000 Traffic Points on the road arches of the national road network and almost 68,000 km of carriageway, considering the extension of all the arches that contained at least one traffic point. This extension represents, in Italy, 6.3% out of the total network. The types of roads, according to the Istat classification, which have the highest percentages of traffic points, are motorways (31.7%), national roads in the built-up area (30.7%) and regional roads in the built-up area (30.2 %). With this criterion, in fact, with the same number of accidents and road kilometers extension, for a similar category of road in two different provinces, for example, the presence of PoT modifies the risk of road accidents, with disadvantage of road sections with less traffic flow, thus they result with an higher danger in terms of vehicles / km.

4. Road accidents indicators computing: main results

The proposed indicators refer to the ratio between road accidents, deaths, injuries and vehicles to the road length, considering the additional information on points of traffic too, the resident population and the vehicle fleet. The following results presented concern indicators out of the road length and the synthesis indicators[7].

The road accidents indicators out of the road length provide a measure of the number of accidents, vehicles involved, dead and injured per 100 kilometers of carriageway in the province. As regards the results, some evidences show that a maximum exposure to the risk of accidents and number of vehicles involved, for motorways and urban roads, is stated mainly in main cities. Milan records a maximum for accidents and vehicles rates per 100 km of carriageway, on Motorways (respectively 144.69 and 313.54 for the 2017 road arches classification) and Genoa on Urban roads, for accidents and vehicles per 100 km (152.91 and 256.66). On

[7] For more details: Istat Experimental Statistics https://www.istat.it/en/archivio/257384

Rural roads, in contrast, the medium-sized provinces result most affected, Monza Brianza has the maximum value.

The Road accidents indicators by road length "weighted" by the PoT (Point of Traffic) information measures the number of accidents, vehicles involved, deaths and injuries per 100 kilometers of carriageway in the province, "adjusted " by the effect on accidents of the presence of traffic points on the roads. The ratios are based on the product between the value of the indicators by road length and the complementary percentage (1-p) to the presence of arches with traffic points on the roads of the province. It is interesting to observe as often the same province presents different positions in the ranking of the indicator by road length, compared to the ranking of indicators adjusted by traffic effect. This is the case of the road accidents on the highways in Genoa and Milan, e.g, the position of Genoa, in fact, reveals a disadvantaged position, in respect of Milan, considering the real danger of the roads of the province with traffic. Finally, a group of *composed and synthetic road accidents indicators* completes the set of 2017 indicators. The method used is the arithmetic mean (z-scores) MZ method (Software Ranker[8] by Istat), (Mazziotta and Pareto, 2013). The values of indices have been calculated according to the MZ (arithmetic average Z scores)[9] method for road arches; vehicle fleet and population. This method allows calculating indicators excluding the traffic volumes on the road and to consider the cause of the accident in a different light, linked to the driving behavior and the characteristics of the infrastructures and vehicles, with or without the traffic flow influence. The results shows that the ranking of the Italian provinces, in respect of the new synthetic indicator, is modified if compared with the ranking of the single indices and underlines a new profile of accidents risk at territorial level.

5. Conclusions

Big Data are not generated directly for statistical purposes, for a statistical use, a suitable methodology must be able to link events in order to refer mainly to units of the population of interest for official statistics. Other elements required are: to process data with the aim to collect them coherently with the statistical framework, to give weights to data so as to guarantee representativeness with respect to the target population and to estimate aggregates of interest including quality measures by spatial and time dimensions. The project of the use of Open Street Map for the

[8] Istat Ranker Tool desktop software http://www.istat.it/en/tools/methods-and-it-tools/analysis-tools/ranker; I.Ranker web application https://i.ranker.istat.it

[9] The three synthesis methods evaluated are: MZ - arithmetic mean (z-scores); MR - relative indices method (IR); MPI - Mazziotta-Pareto Index method (De Muro, 2010). The criterion applied after the robustness analysis is MZ.

investigation of road accidents pattern is adherent to the features cited and the activities in progress are going to provide new developments in the future.

References:

BROCCOLI M., BRUZZONE S. 2021. *Use of the open street map to calculate indicators for road accidents on the Italian roads. Updating with 2017 data*, Istat Roma, Experimental Statistics https://www.istat.it/en/archivio/257384

FARCHI S., MOLINO N., GIORGI ROSSI P. et al. 2006. *Defining a common set of indicators to monitor road accidents in the European Union*. BMC Public Health 6, 183.

HAKKERT A.S., BRAIMAISTER L.2002. *The uses of exposure and risk in road safety studies,* SWOV Institute for Road Safety Research, Leidschendam (NL).

HARBISON F.H., MARUHNIC J., RESNICK J. R. 1970. *Quantitative Analyses of. Modernization and Development*, Princeton University Press, New Jersey.

MAZZIOTTA M., PARETO A. 2013. Methods for constructing composite indices: one for all or all for one?, *Rivista Italiana di Economia Demografia e Statistica*, Vol. 62, No. 2.

RETALLACK A.E., OSTENDORF B. 2020. Relationship between Traffic Volume and Accident Frequency at Intersections, *International Journal of Environmental Research and Public Health*, Vol. 17, No. 4.

SUMMARY

Use of open street map for accident investigation on the road and motorways networks

The aim of the paper is to provide new methods and indicators for road accidents analysis. Traditionally, the road accidents, mortality and harmfulness rates use, as standard denominators, the resident population or vehicles fleet. These measures are not always adequate to rank correctly the exposure to risk of accidents; due to neither the resident population nor the registered vehicles represent correctly the traffic units circulating on the roads. The present research proposes a new approach to the traditional measures, using the length in meters of carriageway by road direction from the Open Street Map, adding the information on the Points of Traffic too. Although the product represents an intermediate result and provides a new ranking of the road accidents risks, the final purpose is to expand the statistical information with the supply of traffic flows (vehicles/km) on the national road network, as widely used in literature, but still missing in Italy.

Marco BROCCOLI, Istat - Istituto Nazionale di Statistica broccoli@istat.it
Silvia BRUZZONE, Istat - Istituto Nazionale di Statistica bruzzone@istat.it

IL DIVARIO DIGITALE PRIMA DELLA PANDEMIA DI COVID-19: FAMIGLIE TRA POSSIBILITA' ED ESCLUSIONE DIGITALE

Laura Zannella

1. Premessa

Nel corso di quest'ultimo anno con la pandemia di Covid-19 l'adozione delle tecnologie digitali ha assunto una particolare rilevanza nei diversi ambiti della vita quotidiana, aprendo alla necessità di una maggiore diffusione dell'utilizzo delle tecnologie della comunicazione e dell'informazione (ICT) per le famiglie italiane. La possibilità che i bisogni emersi in questa fase di emergenza sanitaria diano luogo ad un ampliamento della dimensione digitale nella vita delle famiglie è condizionata sia dal possesso degli strumenti e delle competenze necessarie che dalla disponibilità al cambiamento. Sicuramente la presenza in famiglia di almeno un componente tecnologicamente attrezzato, in grado di assumere il ruolo di "mediatore" per l'accesso alla comunicazione, ai servizi, alla socialità che la rete offre, diventa una risorsa per l'intero nucleo familiare per poter meglio svolgere le attività della vita di tutti i giorni: dal fare la spesa al mantenere relazioni sociali amicali e lavorative, alla didattica, alla pratica di hobby. Con questo studio si vuole analizzare il divario digitale presente nel nostro Paese utilizzando come unità di analisi la famiglia e i divari saranno esaminati attraverso due dimensioni: quella legata all'accesso e quella legata alle competenze digitali.

2. Dati e metodi

Questo studio si basa sui microdati dell'indagine comunitaria sull'uso ICT da parte delle famiglie e degli individui, condotta annualmente dall'Istat mediante l'indagine multiscopo "Aspetti della vita quotidiana" su un campione probabilistico di circa 24.000 famiglie e 54.000 individui. Con l'indagine ICT vengono rilevate informazioni sulle caratteristiche socio-demografiche degli individui (sesso, età, livello d'istruzione, occupazione, risorse economiche, territorio di residenza, etc.) e informazioni dettagliate sull'uso di internet (tipo di connessione, frequenza di utilizzo, attività svolte, competenze digitali, etc.). Lo studio, condotto sulla popolazione residente in Italia, vuole analizzare il divario digitale presente nel nostro

Paese utilizzando come unità di analisi la famiglia. Per l'analisi del digital divide sono state analizzate due dimensioni, una legata all'accesso/utilizzo della rete, l'altra legata alle competenze digitali (Hargattai, E. 2002). Si fa quindi riferimento alla presenza in famiglia di almeno un componente che sia un utente regolare di internet e di almeno un componente che abbia competenze digitali elevate. La definizione di questi indicatori è basata sulla sintesi delle informazioni rilevate direttamente sulle persone che compongono la famiglia. Per quanto riguarda le competenze digitali è stato adottato il nuovo quadro concettuale definito dalla Commissione europea in collaborazione con gli istituti nazionali di statistica, che individua e descrive le competenze digitali in termini di conoscenze, abilità e atteggiamenti. Il modello declina la competenza digitale in 21 competenze specifiche organizzate in quattro domini: informazione, comunicazione, software skill e problem solving. (Vuorikari R, Punie Y, Carretero Gomez S and Van Den Brande G 2016)

Tavola 1 – *Il framework sulle competenze digitali*

Informazione	Comunicazione	Problem solving	Software skill
Copiare o muovere un file o una cartella	Spedire o ricevere email	Trasferire file tra computer e/o da altri dispositivi	Utilizzare software per elaborazione testi
Usare servizi di archiviazione condivisione	Partecipare a social network	Installare software o applicazioni	Usare fogli elettronici per il calcolo
Ottenere informazioni da siti web della P.A.	Telefonare su internet	Modificare le impostazioni di un software	Utilizzare software per modificare foto
Cercare informazioni su merci o servizi	Caricare contenuti di propria creazione	Comprare o ordinare merci e/o servizi	Creare presentazioni
Cercare informazioni sanitarie		Vendere merci o servizi	Scrivere in linguaggio di programmazione
		Svolgere attività di e-learning	Usare le funzioni avanzate dei fogli elettronici per il calcolo
		Usare servizi bancari via Internet	

Per definire il livello di competenza associato a ciascun dominio sono state selezionate una serie di attività svolte nei tre mesi precedenti l'intervista. A seconda del dominio di competenza il numero di attività varia da un minimo di quattro a un massimo di sette (Tavola 1).

Il livello di competenza relativo a ciascun dominio viene definito sulla base del numero di attività svolte ed assume tre modalità: nessuna competenza, competenza di base, competenze al di sopra di quelle base. Mediante i quattro indicatori specifici viene costruito un indicatore complessivo con quattro modalità: nessuna competenza, competenze digitali basse, competenze digitali di base, competenze digitali elevate.

Per analizzare le relazioni esistenti tra la probabilità che in famiglia sia presente almeno un componente con competenze digitali elevate e le caratteristiche socio economiche, demografiche e territoriali è stato applicato un modello logistico lineare. Il modello adottato esprime la trasformata logit della probabilità in funzione lineare degli effetti semplici di un set di variabili esplicative. Le variabili esplicative considerate sono:

- X_1=tipologia familiare, definita sulla base della classe di età del suo componente più giovane: 1=meno di 18 anni, 2=18-44, 3=45-64, 4=65 e più;
- X_2= titolo di studio più elevato in famiglia: 1= laurea, 2= diploma di scuola secondaria di secondo grado, 3= licenza di scuola secondaria di primo grado o inferiore;
- X_3= condizione professionale del capofamiglia: 1=Dirigente, imprenditore, libero professionista, 2=direttivo, quadro, impiegato, intermedio, 3=lavoratore in proprio, 4=operaio, 5=altra condizione;
- X_4 = ripartizione territoriale: 1=Italia settentrionale, 2=Italia centrale, 3=Italia meridionale e insulare;
- X_5 = tipologia del comune di residenza: 1= area metropolitana, 2=comuni con più di 50.000 ab., 3= 10.001-50.000 ab., fino a 10.000 ab.

Poiché le variabili esplicative sono di tipo categorico si è proceduto alla loro trasformazione in variabili dummy. A ciascuna modalità delle cinque variabili esplicative è stata associata una variabile dicotomica che assume il valore 1 se la modalità è presente e il valore 0 altrimenti.

Il modello logistico lineare è dato da:

$$logit(P) = ln\left(\frac{P}{1-P}\right) = \beta_0 + \sum_{r=1}^{5}\sum_{s=1}^{h_r}\beta_{rs}\cdot X_{rs} \qquad (1)$$

dove: P = probabilità che nella famiglia sia presente almeno un componente con competenze digitali elevate; X_{rs} = valore della variabile dummy relativa alla s.ma

modalità della r.ma variabile esplicativa; β_0 = intercetta; β_{rs} = coefficienti di regressione logistica.

La stima dei parametri è stata effettuata con il metodo della massima verosimiglianza, utilizzando la procedura proc logistic del Sas. La procedura fornisce le stime dei parametri e i corrispondenti errori standard, calcola i test relativi alla significatività complessiva del modello e delle singole variabili esplicative, consente di effettuare i confronti tra le diverse modalità di ciascuna variabile esplicativa. I confronti tra coppie di modalità sono basati sull'odds ratio, ossia sul rapporto tra gli odds delle due modalità. Indicando con X_i e X_j le due modalità a confronto e con P_i e P_j le corrispondenti probabilità, i due odds e l'odds ratio sono dati da:

$$odds(X_i) = \frac{P_i}{1-P_i}; \quad (2)$$

$$odds(X_j) = \frac{P_j}{1-P_j}; \quad (3)$$

$$\text{odds ratio} = \frac{odds(X_i)}{odds(X_j)} = \frac{\frac{P_i}{1-P_i}}{\frac{P_j}{1-P_j}} \quad (4)$$

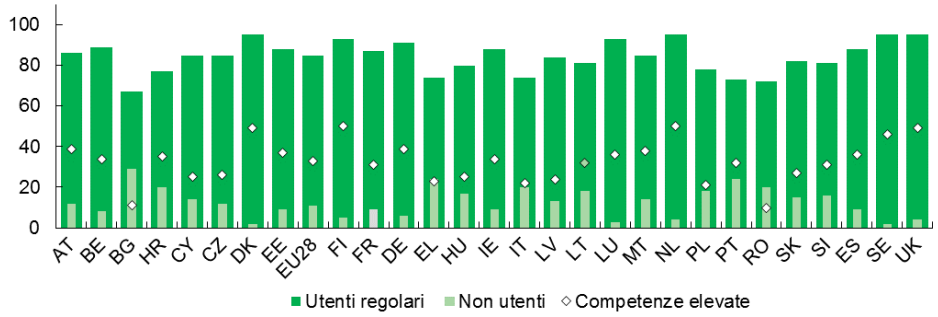
L'ipotesi nulla di uguaglianza dei due odds viene testata mediante il test di Wald basato sul rapporto tra le funzioni di verosimiglianza.

3. Principali risultati

3.1 L'Italia nel contesto europeo

La disponibilità di un accesso ad internet rappresenta la precondizione affinché le famiglie possano usufruire dei servizi online e sviluppare le competenze necessarie per poter utilizzare appieno le opportunità offerte dalle ICT (Di Maggio P.J., Hargattai, E. 2002) L'Italia, a riguardo, sconta un ritardo rispetto agli altri Paesi dell'Unione europea. Infatti in Italia nel 2019 internet è utilizzato regolarmente dal 74% degli individui tra i 16 e i 74 anni, con un aumento di cinque punti percentuali negli ultimi tre anni. Nell'insieme dei 28 paesi Ue, invece, tra il 2017 e il 2019 gli utenti sono saliti dall'81% all'85%. Un altro elemento che testimonia il nostro ritardo è la quota di non utenti pari al 20% contro l'11% della media europea. Tale divario si riscontra anche per le competenze digitali, il 22% degli utenti di 16-74 anni residenti in Italia ha competenze digitali elevate contro il 33% della media EU28 (Figura 1).

Figura 1 – *Famiglie con almeno un componente in età 16-74 anni che dispongono di un accesso a Internet da casa, 2019*



Fonte Eurostat Survey on ICT usage by household and individuals.2019

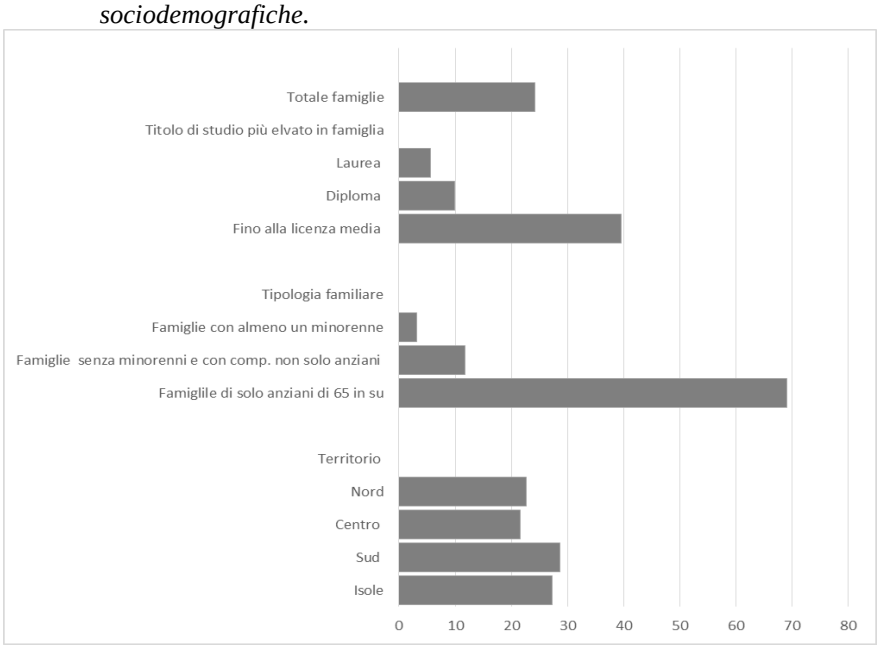
3.2 Il ruolo della famiglia per la diffusione delle ICT

Nel corso degli ultimi mesi, con la pandemia da Covid-19, l'adozione delle tecnologie digitali ha assunto una particolare rilevanza nei diversi ambiti della vita quotidiana, quindi la presenza di almeno un componente tecnologicamente attrezzato diventa una risorsa per l'intera nucleo familiare per poter meglio svolgere le attività della vita di tutti i giorni: dal fare la spesa al mantenere relazioni sociali amicali e lavorative, alla didattica, alla pratica di hobby. (Erker T., Hartam M., Punie Y., 2005)

Con riferimento al complesso delle famiglie, nel 2019 sono 6 milioni 175 mila (24,2%) le famiglie ancora escluse dalla Rete nelle quali nessun componente ha utilizzato internet negli ultimi 12 mesi. Le famiglie più svantaggiate sono quelle costituite da soli anziani e quelle con un basso titolo di studio (al massimo la licenza media). Sensibili anche le differenze legate al territorio, la percentuale di famiglie in cui nessun componente usa internet sfiora infatti il 30% al Sud (Figura 2).

In un momento in cui è sempre più richiesto l'utilizzo del digitale, le famiglie in cui non sono presenti internauti rischiano di essere completamente escluse non potendo usufruire al proprio interno di un "mediatore" che possa costituire un link diretto con la Rete.

Figura 2 – *Famiglie in cui nessun componente usa internet e caratteristiche sociodemografiche.*



Fonte Elaborazioni dell'Autrice su dati ISTAT Indagine Multiscopo "Aspetti della vita quotidiana". Anno 2019

Sono il 73,5% le famiglie in cui è presente almeno un componente che si è connesso negli ultimi 3 mesi, ma anche tra loro si evidenziano alcune criticità. Una situazione ancora meno favorevole si riscontra per le abilità d'uso, infatti sono meno del 40% le famiglie in cui è presente almeno un internauta con competenze digitali elevate, il 25% ha almeno un componente con competenze digitali di base, il 33% invece ha competenze basse. Vi è inoltre una nicchia di famiglie il 2,7%, pari a 511 mila 556, in cui nessun dei suoi componenti ha competenze digitali (Tavola 2).

Tra le famiglie si osserva un forte divario tecnologico da ricondurre a fattori sociali, generazionali e territoriali, che trova conferma nei risultati del modello di regressione logistica. Il test per la valutazione dell'adattamento del modello ai dati osservati mostra un'elevata significatività, così come altamente significativi risultano gli effetti di tutte le variabili esplicative ad esclusione della tipologia del comune di residenza (Tavola 3).

Tavola 2 – *Competenze digitali dei componenti la famiglia.*

Competenze digitali dei componenti la famiglia	N. famiglie	% su famiglie con almeno utente regolare
Almeno un componente con competenze elevate	7.384.674	39,4
Almeno un componente con competenze di base	4.701.874	25,1
Almeno un componente con competenze di basse	6.151.169	32,8
Nessun componente ha competenze digitale	511.556	2,7
Totale famiglie con almeno un utente regolare	18.749.273	100

Fonte: Elaborazioni dell'Autrice su dati Istat Indagine Multiscopo "Aspetti della vita quotidiana". Anno 2019

Tavola 3 – *Modello logistico: Test di adattamento del modello e analisi degli effetti delle variabili esplicative.*

Effetti	Gradi di libertà	Wald Chi-quadro	Pr > Chi-quadro
Modello	16	3627,09	<0,001
Età del componente più giovane	3	995,24	<0,001
Titolo di studio più elevato	2	1305,61	<0,001
Condizione professionale del capo famiglia	4	362,88	<0,001
Ripartizione territoriale	4	104,21	<0,001
Comune di residenza	3	5,76	0,1241

Fonte: Elaborazione dell'Autrice su dati Istat Indagine Multiscopo "Aspetti della vita quotidiana". Anno 2019

Gli effetti delle modalità delle 5 variabili esplicative sono stati stimati mediante l'odds ratio, calcolato come rapporto tra l'odd di ciascuna modalità e l'odd di una modalità scelta come riferimento. La significatività dell'odds ratio è stata valutata utilizzando il test chi-quadro di Wald (Tavola 4). L'odds ratio anche se esprime il rapporto tra due odds può essere utilizzato per interpretare il verso del rapporto tra le due probabilità, ma non l'intensità di tale rapporto. Così un valore dell'odds ratio uguale 2 significa che la probabilità a numeratore è maggiore di quella al denominatore, ma non è corretto dire che sia il doppio.

Tavola 4 – *Modello logistico: Odds ratio, errore standard e test di Wald*

Confronti	Odds ratio	Errore standard	Wald Chi-quadro	Pr > Chi-quadro
Età del componente più giovane (Rif: 65 anni e più				
Almeno un minore di 18 anni	4,08	0,448	163,66	<,0001
Senza minori e almeno un comp. 18-44	4,88	0,513	228,20	<,0001
Senza comp 0-44 e almeno un comp. 45-64	1,91	0,210	34,48	<,0001
Titolo di studio più elevato (Rif: Licenza media)				
Laurea	8,05	0,602	776,46	<,0001
Diploma	3,34	0,232	299,85	<,0001
Condizione professionale del capo famiglia (Rif: Operai)				
Dirigenti, Impr, Lib.prof,	2,02	0,152	87,38	<,0001
Diret,Quadri,Impieg,Inter	1,97	0,115	132,62	<,0001
Lavoratori in proprio, coad,	1,25	0,097	8,62	0,0033
Altra condizione	1,03	0,163	0,03	0,8583
Ripartizione territoriale (Rif: Isole)				
Nord-est	1,63	0,121	42,99	<,0001
Nord-ovest	1,54	0,120	31,03	<,0001
Centro	1,43	0,109	21,36	<,0001
Sud	1,14	0,088	2,73	0,0983
Comune di residenza (Rif: Fino a 10.000 ab.)				
Area metropolitana	1,18	0,038	2,82	0,0928
Più di 50.000 ab.	1,09	0,035	1,99	0,1583
10.001-50.000 ab.	1,02	0,034	1,76	0,1843

Fonte: Elaborazioni dell'Autrice su dati Istat Indagine Multiscopo "Aspetti della vita quotidiana". Anno 2019

La probabilità che in famiglia sia presente almeno un componente con competenze digitali elevate è significativamente minore nelle famiglie composte da soli anziani ed è più elevata nelle famiglie in cui è presente almeno un componente di 18-44. La probabilità risulta significativamente crescente con il titolo di studio, l'odd delle famiglie con laureati è otto volte maggiore di quello delle famiglie i cui i componenti hanno al massimo la licenza della scuola media inferiore. Si riscontra un odd doppio nelle famiglie in cui il capofamiglia è un dirigente imprenditore o libero professionista rispetto a quelle con capofamiglia operaio. In riferimento al

territorio differenze significative si evidenziano tra le regioni settentrionali e centrali rispetto a quelle insulari, mentre non emergono differenze significative tra il meridione e le isole. Come si è già detto il modello non mette in evidenza differenze significative tra le quattro tipologie del comune di residenza.

4. Alcune considerazioni finali

L'analisi svolta con questo studio ha messo in luce come per una parte delle famiglie italiane l'emergenza sanitaria potrebbe rappresentare l'occasione per un avanzamento nell'utilizzo delle tecnologie digitali nella vita quotidiana. Si tratta delle famiglie in cui è presente almeno un componente tecnologicamente attrezzato, con competenze digitali elevate, in grado di assumere il ruolo di mediatore per l'accesso alla comunicazione, ai servizi, alla socialità che la rete offre, potendo svolgere per l'intero nucleo familiare la funzione di traino, socializzando con gli altri membri le proprie competenze per meglio svolgere le attività della vita di tutti i giorni. In altri segmenti la mancanza di adeguate condizioni potrebbe rendere più difficoltoso l'accesso ai vantaggi derivati dall'uso della Rete. In questa area si trovano le famiglie che hanno competenze digitali basse, che non potendo contare sulla presenza di alcun membro già in possesso di adeguati skill nella fruizione delle tecnologie digitali, possono vivere con maggiore disagio la gestione del quotidiano. Le famiglie che non usano internet restano certamente quelle più esposte ad una condizione di grave criticità, impreparate e senza strumenti per accedere ai servizi online. Se i bisogni emergenziali connessi alla pandemia da Covid-19 spingono verso una maggiore diffusione degli strumenti digitali occorrerà operare in modo che ciò non acuisca i divari digitali già esistenti. Al contrario sarà importante cogliere e supportare eventuali segnali di crescita delle competenze tecnologiche e di mutamento negli usi delle ICT eventualmente sviluppati pur in un contesto di mezzi limitati.

Riferimenti bibliografici:

ERKER T., HARTMAN M., PUNIE Y. 2005. *Domestication of Media and* Technology. N. Berkshire (UK) Open University Press.

HARGATTAI E., HSIEH Y.P. 2013. Digital inequality. In W. H. DUTTON (Ed) *Oxford Handbook of Internet Studies* Oxford, UK. Oxford University Press, pp.129-150.

DI MAGGIO P.J., HARGATTAI E. 2002. From the 'digital divide' to 'digital inequality' studying internet use as penetration increases. *Working Papers 47,*

Princeton University, Woodrow Wilson School of Public and International Affairs, Center for Arts and Cultural Policy Studies.
HARGATTAI E. 2002. Second-Level Digital Divide: Differences in People's Online Skills, *First Monday*, Vol. 7, No. 4.
ISTAT Indagine Multiscopo "Aspetti della vita quotidiana". ANNO 2019.
VUORIKARI R., PUNIE Y., CARRETTERO GOMEZ S., VAN DEN BRANDE G. 2016. DigComp 2.0: *The Digital Competence Framework for Citizens. Update Phase 1: the Conceptual Reference Model* Luxembourg (Luxembourg): Publications Office of the European Union.

SUMMARY

The digital divide before the Covid-19 pandemic: households between possibility and digital exclusion

During this past year, with the Covid-19 pandemic, the adoption of digital technologies has taken on particular relevance in the different areas of daily life, opening up the need for a greater diffusion of the use of ICT (Information Communication Technology) for Italian households. This paper draws on micro-data from community survey on ICT usage by households and individuals conducted for Italy by Istat to develop an analysis on digital divide. The unit of analysis was household. If there is a technologically equipped member of the household, he or she could act as a " mediator " for the other members in the usage of ICT. The digital divide has been analysed through two dimensions, one related to access and the other to digital skills.

Laura Zannella, Istat, laura.zannella@istat.it

SYNCHRONIZATION AMONG REAL BUSINESS CYCLES OF U.S. STATES

Margherita Gerolimetto, Stefano Magrini

1. Introduction

Understanding the level of synchronization among real business cycles is relevant from various perspectives. For example, when areas are in a monetary union, the issue of whether their business cycles are coordinated is of crucial importance. If the business cycles are not synchronized, i.e. are in different phases (expansion or recession), then different policies would be required to bring major indicators in these economies on the optimal path. Thus, while an economy entering a recession would require an expansionary policy with a decrease in the interest rate, a booming economy requires a tighter monetary stance. In such a case, a common monetary policy would be suboptimal and lead to higher costs.

In this paper we present a synchronization analysis of the real business cycles of the U.S. states between 2005 and 2019. We employ data on real GDP covering the period stretching from 2005:Q1 to 2020:Q2. These data are somewhat innovative with respect to previous analyses (among others, Beckwork, 2010; Partridge and Rickman, 2005), because they are elaborated making use of Regional price parities (RPPs) by state and offers a more in-depth analysis with respect to studies where prices are assumed to be homogeneous across spatial units, as it happens by adopting a national deflator.

From the methodological point of view, to isolate the business cycles we adopt a recently developed MatLab function for signal extraction. This function, called CiSSA (Circulant Singular Spectrum Analysis), proposed by Bógalo *et al*. (2021a) extracts the underlying signals in a time series identifying their frequency of oscillation in an automated way. Bógalo *et al*. (2021b) adopted CiSSA in case of macroeconomic time series, in particular to obtain revisions at the end of the sample; they show via simulations that the algorithm performs well. CiSSA represents an alternative to the traditional Hodrik-Prescott filter (Hodrik and Prescott, 1997) and has the advantage of being very versatile and working well both in case of stationary and non-stationary time series.

Once the business cycles of the U.S. states are extracted using CiSSA, we study their level of synchronization with the national one using various indices of

synchronization. In particular, we will focus on the concordance index (Harding and Pagan, 2006) and the cross-correlations.

Moreover, we also investigate the possibility that immigrant flows might affect the level of synchronization of the state cycles with respect to the national one. This is based on the idea that immigrants tend to exhibit a higher propensity to migrate in reaction to labour market opportunities (Molloy *et al.*, 2011). Consequently, it can be of interest to investigate whether their movements can help synchronizing the business cycles. To do this, we study the correlation between on the variation (in absolute value) between the average share of non-US natives over 2005-2009 with respect to the same share over 2000-2005 and the level of synchronization of the regional cycles with respect to the national one.

The structure of the paper is as follows. We will present the algorithm of signal extraction called CiSSA in the second section and recall two measures of synchronization in the third section. The empirical analysis will be presented in the fourth section while the fifth section concludes.

2. Signal extraction with CiSSA

Circulant Singular Spectrum Analysis, CiSSA, is an algorithm for signal extraction (Bógalo *et al.*, 2021). More in general, Singular Spectrum Analysis (SSA) is a nonparametric procedure to reconstruct the original time series as the sum of orthogonal components of known frequencies. It works in two distinct phases. In the first phase, called decomposition, the original time series is transformed into a related trajectory matrix made of pieces of length L and then elementary matrices are obtained via the single value decomposition; according to which method is adopted to perform the single value decomposition, there are different versions of the SSA. In the second phase, called reconstruction, the elementary matrices are grouped and the time series is recovered.

More in details, given a time series $X_1,\dots X_T$, finite realization of a zero-mean stochastic process $\{X_t\}$ and let L be a possible integer, called the window width (such that $1< L < T/2$), the decomposition and reconstruction stages of SSA are carried out in 4 steps.

1. *Embedding*: a trajectory matrix is built by putting together lagged pieces of size L of the original series
2. *Decomposition*: the trajectory matrix is decomposed in elementary matrices of rank 1, associated to various frequencies
3. *Basic frequency grouping*: elementary matrices are grouped by frequency
4. *Reconstruction*: the matrices obtained in step 3 are transformed into M signals of the same length as the original series for frequencies w_k, where $k = 1,..., L$.

The decomposition phase requires the computation of the eigenvectors by diagonalizing the matrix of second moments with the trajectory matrix. Traditionally, the identification of the frequencies associated with the trajectory matrices is made after the time series have been reconstructed. CiSSA replaces the variance covariance matrix of the Toeplitz version with a related circulant matrix. The advantage of adopting circulant matrices is that they allow for exact identification between the eigenvectors and eigenvalues obtained in terms of the frequency of the components they represent. As a result, in the reconstruction phase it is possible to group the matrices by frequency and identify the signal associated to each frequency.

In this study we will adopt a MatLab function for CiSSA algorithm, whose outcome is a set of *M* time series, one for each frequency w_k. The MatLab function also allows to implementing different grouping strategy that will lead to obtain the default signals typically required in an economic time series: trend, business cycle, seasonality.

3. Business cycle synchronization

Roughly speaking, business cycles synchronization means that their turning points occur at approximately the same points in time. More specifically turning points are defined as local maxima and minima in the sample path of the time series. Synchronization can be measured and here we will focus in particular on two complementary approaches: concordance index and cross-correlation.

The concordance index introduced by Harding and Pagan (2006) measures the amount of time that in which two economies are in the same business cycle phase. For a generic economy *X* suppose that S_{Xt} is a binary variable taking value 1 when the economy *X* at time *t* is in recession and value 0 when it is in expansion. Similarly it is defined S_{Yt} for another economy *Y*. So, over a period of *T* instants, the concordance index C_{XY} that measures the degree of synchronization of the two economies *X* and *Y* is:

$$C_{XY} = \frac{1}{T}\sum_{t=1}^{T}[S_{Xt}S_{Yt} + (1 - S_{Xt})(1 - S_{Yt})] \quad (1)$$

When C_{XY} = 1 there is perfect concordance, i.e. perfect synchronization. Conversely, when C_{XY} = 0 there is perfect discordance, i.e. cycles are always is opposite phases. Values of the index ranging between 0.5 and 1 indicate weak to perfect synchronization (procyclical) and values of the index ranging between 0 and 0.5 indicate perfect to weak discordance (countercyclical).

Following Harding and Pagan (2006), it is possible to re-write C_{XY} as a monotonic transformation of ρ_{xy} the correlation between the time series Y_t and X_t. Specifically, Harding and Pagan (2006) reparametrize C_{XY} as follows:

$$C_{XY} = 1 + 2\hat{\rho}_S \left(\hat{\mu}_{S_X}(1 - \hat{\mu}_{S_X})\right)^{\frac{1}{2}} \left(\hat{\mu}_{S_Y}(1 - \hat{\mu}_{S_Y})\right)^{\frac{1}{2}} + 2\hat{\mu}_{S_X}\hat{\mu}_{S_Y} - \hat{\mu}_{S_X} - \hat{\mu}_{S_y} \quad (2)$$

where $\hat{\rho}_S$ is the estimated correlation coefficient between S_{Xt} and S_{Yt}, $\hat{\mu}_{S_X}$ and $\hat{\mu}_{S_Y}$ are, respectively, the estimated E(S_{Xt}) and E(S_{Yt}). As pointed out by Harding and Pagan (2006), due to the binary nature of S_{Xt} and S_{Yt}, the estimated standard deviations are $\hat{\sigma}_{S_X} = \sqrt{(\hat{\mu}_{S_X} - \hat{\mu}_{S_X}{}^2)}$ and $\hat{\sigma}_{S_Y} = \sqrt{(\hat{\mu}_{S_Y} - \hat{\mu}_{S_Y}{}^2)}$. Harding and Pagan (2006) further clarify that the concordance index has a maximum (value 1) when S_{Xt} = S_{Yt} and a minimum (value 0) when S_{Xt} = (1-S_{Yt}) and when either of these holds $\hat{\sigma}_{S_X}\hat{\sigma}_{S_Y} = \hat{\sigma}^2_{S_X}$. Therefore C_{XY}=1 corresponds to $\hat{\rho}_S = 1$ while $C_{XY} = 0$ corresponds to $\hat{\rho}_S = -1$. Moreover, with similar arguments Harding and Pagan (2006) show that it is possible to test synchronization between two cycles with the following regression

$$\frac{S_{Yt}}{\hat{\sigma}_{S_Y}\hat{\sigma}_{S_X}} = \nu + \rho_S \left(\frac{S_{Xt}}{\hat{\sigma}_{S_Y}\hat{\sigma}_{S_X}}\right) + \epsilon_t \quad (3)$$

If ρ_S is not significantly different from zero, there is no synchronization (which is equivalent to C_{XY}= 0.5). Note that because of the possible problems with the error model, the Newey West heteroskedasticity and autocorrelation consistent (HACC) adjusted standard error must be used when testing the significance of ρ_S.

In a complementary perspective to the concordance index, cross-correlations are also capable of providing valuable information about the level of synchronization. Cross-correlations measure linear dependence between two time series at different time lags. In general, two types of co-movements can be analysed with the cross-correlation coefficients. First, contemporaneous co-movements, which could be pro-cyclical (zero lag correlation is positive), countercyclical (zero lag correlation is negative), a-cyclical (zero lag correlation is nonsignificant). Second, non contemporaneous co-movements, or phase shifts, which suggest that a series is leading if the largest absolute value of cross-correlation is on negative lag; a series is coincidental if the largest absolute value of cross-correlation is on zero lag; a series is lagging if the largest absolute value of cross-correlation is on positive lag.

So, supposing X_t. is the US GDP and Y_t is a state GDP, the cross-correlation at lag zero indicates if one state movement is pro-cyclical, counter-cyclical or a-cyclical with respect to the national cycle. On the other hand, the maximum of the

absolute value of the cross-correlation indicates if one state cycle is leading, coincident or lagging the national one.

4. Empirical analysis

We now present our empirical analysis. We employ data (in logarithm) on real GDP covering the period stretching from 2005:Q1 to 2020:Q2 from BEA (Bureau of Economic Analysis) for the U.S. states and the overall U.S. These data are somewhat innovative with respect to previous analyses because they are elaborated making use of Regional price parities (RPPs) by state. RPPs measure the differences in price levels across states for a given year and cover all consumption goods and services, including housing rents. Thus, this indeed allow us to offer a more accurate analysis with respect to studies where prices are assumed to be homogeneous across spatial units, as it happens by adopting a national deflator.

Moreover, we employ annual data from ACS (American Community Survey), for the share of non-natives covering the period 2000-2010 for the information about the flows of immigrant, for the second part of the analysis[1].

4.1. Synchronization analysis

Our synchronization analysis starts from the extraction of the business cycles using CiSSA algorithm, presented in the second section. This is done for each of the U.S. states and for the overall national economy. The second step consists in dating all the business cycles, i.e. in identifying the turning points for all the cycles. For this step we implement a simple algorithm, based on the following set-up:

- a minimum distance between consecutive peaks (troughs): 7 years
- a minimum prominence of peaks (troughs): 0.2%
- alternation of peaks and troughs

These rules have been thought to represent a behavior that, firstly, mirrors a cycle with alternation of peaks and troughs, and, secondly, reflects the behavior of business cycle of the U.S. economy analyzed by the National Bureau of Economic Research (NBER). Indeed, in the period between 2001 and 2020 that includes the time span under our examination, the average duration of the peak-to-peak period and trough-to-trough is 10 years, so we choose as minimum distance between consecutive peaks (troughs) to give some flexibility to the behavior of the states. Also for the prominence, we followed the same logic.

[1] Data have been accessed from IPUMS, Integrated Public Use Microdata Series (Ruggles et al. 2020).

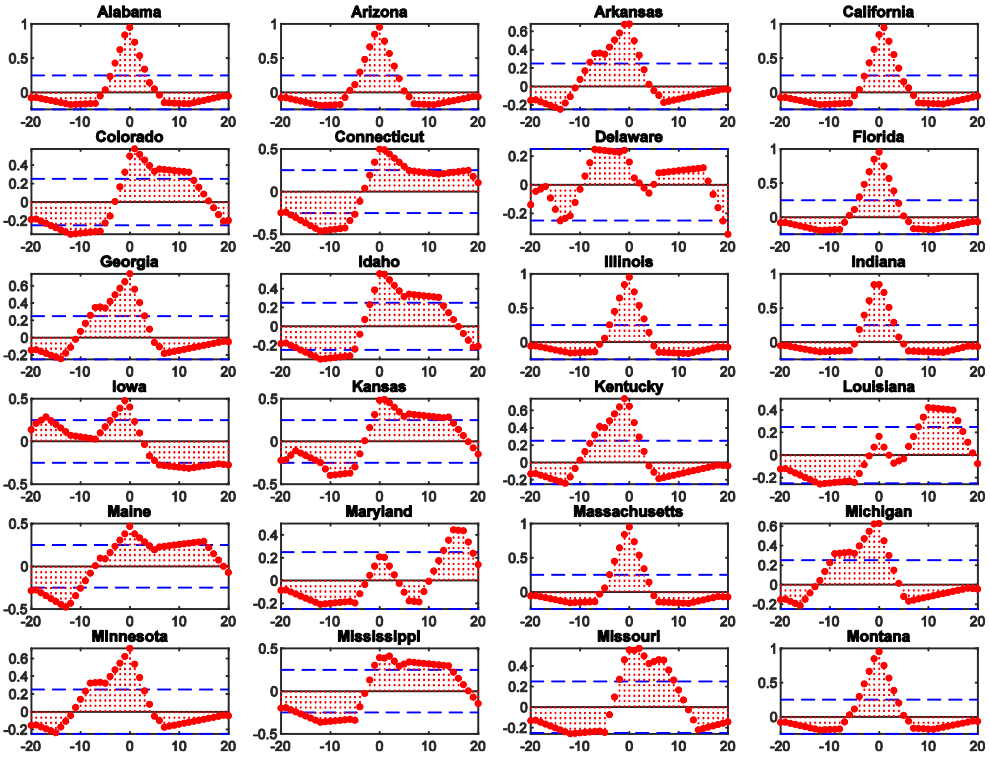
We present in Table 1 the results on the concordance index C_{xy}, coupled with the ρ_S coefficient obtained estimating regression (2) for each pairing of states economies with the U.S. economy. The double stars in the Table indicate strong significance (p-value smaller than 0.001).

Table 1 – *Concordance Index*

State	C_{xy}	ρS	State	C_{xy}	ρS
Alabama	0.984	0.917**	Nebraska	0.766	0.519**
Arizona	0.984	0.981**	Nevada	0.906	0.885**
Arkansas	0.875	0.782**	New Hampshire	0.906	0.756**
California	0.953	0.814**	New Jersey	0.875	0.782**
Colorado	0.781	0.603**	New Mexico	0.656	0.513**
Connecticut	0.703	0.635**	New York	0.766	0.519**
Delaware	0.375	0.167	North Carolina	0.734	0.545**
Florida	0.984	0.981**	North Dakota	0.453	0.006
Georgia	0.891	0.865**	Ohio	0.875	0.846**
Idaho	0.797	0.686**	Oklahoma	0.609	0.519**
Illinois	0.984	0.917**	Oregon	0.766	0.647**
Indiana	0.953	0.75**	Pennsylvania	0.969	0.833**
Iowa	0.703	0.506**	Rhode Island	0.891	0.737**
Kansas	0.734	0.609**	South Carolina	1	1**
Kentucky	0.875	0.718**	South Dakota	0.281	0.051
Louisiana	0.703	0.186	Tennessee	0.906	0.885**
Maine	0.672	0.596**	Texas	0.953	0.942**
Maryland	0.734	0.224	Utah	0.922	0.904**
Massachusetts	0.984	0.917**	Vermont	0.797	0.686**
Michigan	0.844	0.744**	Virginia	0.578	0.481**
Minnesota	0.875	0.846**	Washington	0.891	0.865**
Mississippi	0.734	0.481**	West Virginia	0.906	0.564**
Missouri	0.844	0.615**	Wisconsin	0.906	0.821**
Montana	0.984	0.981**	Wyoming	0.703	0.378*

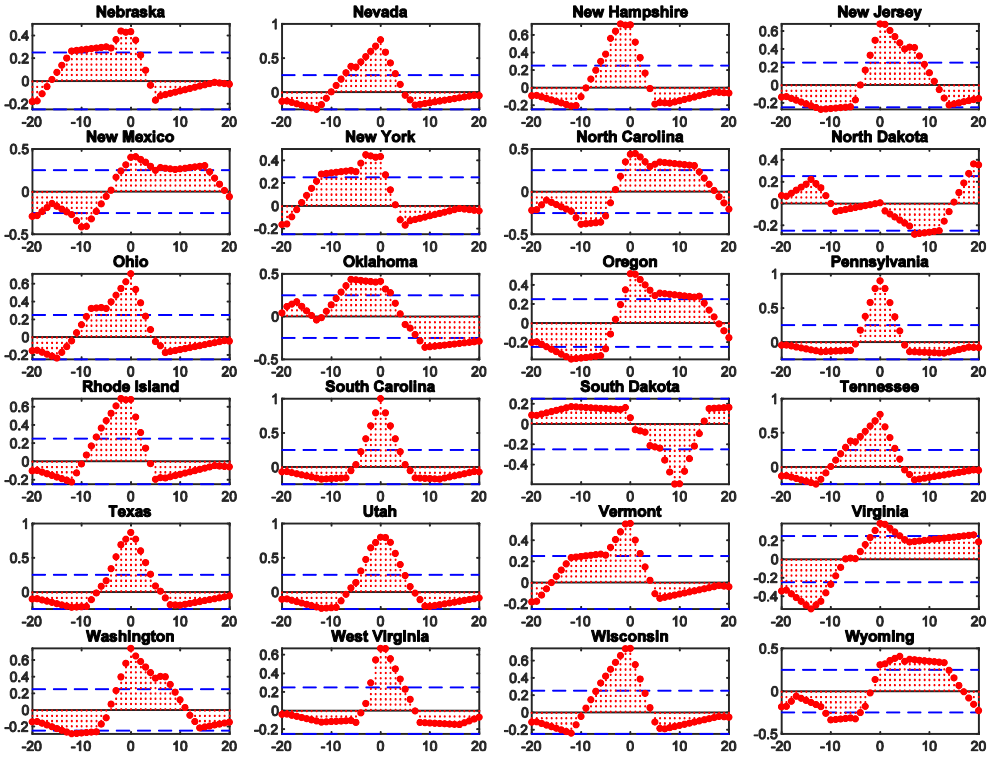
Then, we proceed by analysing the degree of synchronization between the cycle of each state and the cycle of the national economy. This step consists of the calculation of the concordance index between each state and the national economy, and then in the determination of the cross-correlations between state and the national economy.

Figure 1a – *Cross-correlations of each state and the national economy*.



As expected, states that are well synchronized with the national economy, such as Alabama, exhibit a cross-correlation with respect to overall U.S. characterized by a very visible positive peak on the zero lag, and this pattern repeats regularly for all states whose concordance index is high. Instead, for those states that are a-synchronized with the national economy, such as Luisiana, the cross-correlation is very irregular and no clear pattern is visible.

Figure 1b – *Cross-correlations of each state and the national economy*



4.2. The role of internal migrations

Immigrants tend not to distribute homogeneously across states. For instance, based on American Community Survey data, in 2010 about two-thirds (65%) of the total foreign-born population lived in just six states (California, New York, Texas, Florida, New Jersey and Illinois) and over one-fourth (25.4%) lived in California. In addition, the skill distribution for immigrants is characterized by a strong polarization as most of them either acquired a low level of schooling or hold a graduate degree. For instance, in 2010 about 32% of immigrants had not completed the equivalent of high-school education, compared with only 11% of natives. At the same time, immigrants are as likely as natives to be highly educated, with 27% of immigrants and 28% of natives having completed a bachelor's degree. In contrast, are underrepresented in the middle of the skill distribution, among workers with high-school or some college education (41% for immigrants, 61% for natives). Also

due to these characteristics, immigrants tend to be more reactive to labour market opportunities (Peri, 2013) and so their movements might help smoothing differences between business cycles, thus improving synchronization.

To study the role of immigrants on the degree of synchronization, we focus on the variation (in absolute value) between the average share of non-US natives over 2005-2009 with respect to the same share over 2000-2005. In particular, we firstly calculate the correlation between the degree of synchronization and the level of mobility; then, we run a simple regression between the same variables.

Table 2 – *Correlation and regression analysis results.*

Correlation	Estimated regression coeff	Standard error	p-value
0.349	10.395	3.931	0.011

The results are presented in Table 2; they clearly show that an increase in the share of non-natives within the population is associated with an increase in the degree of synchronization with the national cycle, thus confirming the intuition that, given their larger level of mobility in response to labour market opportunities, they can effectively improve business cycles synchronization.

5. Concluding remarks

In this paper, we present a synchronization analysis of the real business cycles of U.S. states. We employ data on real GDP covering the period stretching from 2005:Q1 to 2020:Q2. Once the business cycles of the U.S. states are extracted using CiSSA, we study their level of synchronization with the national one using two measures of synchronization.

Moreover, based on the literature that documents how workers with different levels of education and human capital exhibit different level of propensity to migrate, we also investigate the possibility that immigrant flows might affect the level of synchronization of the state cycles with respect to the national one.

Results show that there is a remarkable variability in the degree of synchronization across states. More in details, Delaware, Luisiana, Maryland, North and South Dakota are a-synchronized with the national economy. Instead, Alabama, Arizona, Florida, Illinois, Massachusetts, Montana, South Carolina are strongly synchronized. In addition, an increase of the share of non-natives in the population is associated with an increase in the degree of synchronization with the national cycle.

Acknowledgements

The authors wish to thank warmly an anonymous referee for the valuable comments to an earlier version of this paper. The authors are also grateful to the Editor for the suggestions. The usual disclaimer applies and the views are the sole responsibility of the authors.

References

BECKWORTH D. 2010. One nation under the Fed? The asymmetric effect of US Monetary Policy and its implications for the United States as an Optimal Currency Area, *Journal of Macroeconomics*, Vol. 32: pp. 732-746.

BOGALO J., PONCELA P. and SENRA E. 2021a. Circulant spectrum analysis: a new automated procedure for signal extraction, *Signal Processing*, Vol. 179, 2021. doi:10.1016/j.sigpro.2020.107824

BOGALO J., PONCELA P. and SENRA E. 2021b. Circulant Singular Spectrum Analysis to Monitor the State of the Economy in Real Time, *Mathematics*, Vol. 9, 1169. https://doi.org/10.3390/math9111169

HARDING D, PAGAN A. 2006. Synchronization of cycles, *Journal of Econometrics*, Vol. 132, pp. 59-79.

HODRIK R., PRESCOTT E. C. 1997. Postwar US business cycles: an empirical investigation. *Journal of Money Credit and Banking*, Vol. 29, pp. 1-16.

MOLLOY R., SMITH C.L. and WOZNIAK A. 2011. Internal migration in the United States, *Journal of Economic Perspectives*, Vol. 25, pp. 173-196.

PERI G. 2013. The economic benefits of immigration. *Berkeley Review of Latin American Studies, Fall 2013*, pp. 14-19.

PARTRIDGE M., RICKMAN D.S. 2005. Changes in asymmetric shocks in an optimal currency area: an analysis using US state data, *Oxford Economic Papers*, Vol. 57, No. 3, pp. 373-397.

RUGGLES S., FLOOD S., GOEKEN R., GROVER J., MEYER E., PACAS J. and SOBEK M. 2020. *IPUMS USA: Version 10.0 [dataset]*. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V10.0

SUMMARY

Synchronization among real business cycles of U.S. states

In this paper we present a synchronization analysis of the real business cycles of U.S. states. We employ data on real GDP covering the period stretching from 2005:Q1 to 2020:Q2. From the methodological point of view, to isolate the business cycles we adopt a recently developed MatLab function for signal extraction. Once the business cycles of the U.S. states are extracted using CiSSA, we study their level of synchronization with the national one using various indices of synchronization. Moreover, based on the literature that documents how workers with different levels of education and human capital exhibit different level of propensity to migrate, we study the correlation between immigrant flows and the level of synchronization of the regional cycles with respect to the national one.

Margherita GEROLIMETTO, Università Ca' Foscari Venezia, Dipartimento di Economia. E-mail: margherita.gerolimetto@unive.it

Stefano MAGRINI, Università Ca' Foscari Venezia, Dipartimento di Economia. E-mail: stefano.magrini@unive.it

SOCIETÀ E RIVISTA ADERENTI AL SISTEMA ISDS
ISSN ASSEGNATO: 0035-6832

Direttore Responsabile: Prof.ssa CHIARA GIGLIARANO

Iscrizione della Rivista al Tribunale di Roma del 5 dicembre 1950 N. 1864

Associazione all'Unione Stampa Periodica Italiana

TRIMESTRALE

La copertina è stata ideata e realizzata da Pardini, Apostoli, Maggi p.a.m.@tin.it – *Roma*

Stampato da CLEUP sc
"Coop. Libraria Editrice Università di Padova"
Via G. Belzoni, 118/3 – Padova (Tel. 049/650261)
www.cleup.it

ATTIVITÀ DELLA SOCIETÀ

RIUNIONI SCIENTIFICHE

XXXVII	La mobilità dei fattori produttivi nell'area del Mediterraneo (Palermo, 15-17 giugno 2000).
XXXVIII	Qualità dell'informazione statistica e strategie di programmazione a livello locale (Arcavacata di Rende, 10-12 maggio 2001).
XXXIX	L'Europa in trasformazione (Siena, 20-22 maggio 2002).
XL	Implicazioni demografiche, economiche e sociali dello sviluppo sostenibile (Bari, 15-17 maggio 2003).
XLI	Sviluppo economico e sociale e ulteriori ampliamenti dell'Unione Europea (Torino, 20-22 maggio 2004).
XLII	Sistemi urbani e riorganizzazione del territorio (Lucca, 19-21 maggio 2005).
XLIII	Mobilità delle risorse nel bacino del Mediterraneo e globalizzazione (Palermo, 25-27 maggio 2006).
XLIV	Impresa, lavoro e territorio nel quadro dei processi di localizzazione e trasformazione economica (Teramo 24-26 maggio 2007).
XLV	Geopolitica del Mediterraneo (Bari, 29-31 maggio 2008).
XLVI	Povertà ed esclusione sociale (Firenze 28-30 maggio 2009).
XLVII	Un mondo in movimento: approccio multidisciplinare ai fenomeni migratori (Milano 27-29 maggio 2010).
XLVIII	150 anni di Statistica per lo sviluppo del territorio: 1861-2011. (Roma 26-28 maggio 2011).
XLIX	Mobilità e sviluppo: il ruolo del turismo. (San Benedetto del Tronto, 24-26 maggio 2012).
L	Trasformazioni economiche e sociali agli inizi del terzo millennio: analisi e prospettive (Università Europea di Roma, 29-31 maggio 2013).
LI	Popolazione, sviluppo e ambiente: il caso del Mediterraneo (Università Federico II di Napoli, 29-31 maggio 2014).
LII	Le dinamiche economiche e sociali in tempo di crisi (Università Politecnica delle Marche, 28-30 maggio 2015).
LIII	Mutamento economico e tendenze socio-demografiche tra sfide e opportunità (Università degli Studi Internazionali di Roma, 26-28 maggio 2016).
LIV	Mobilità territoriale, sociale ed economica: modelli e metodi di analisi (Università degli Studi Internazionali di Catania, 25-26 maggio 2017).
LV	Coesione sociale, welfare e sviluppo equo e sostenibile (Università degli Studi dell'Insubria, Varese 24-25 maggio 2018).
LVI	Benessere e Territorio: Metodi e Strategie (Università Politecnica delle Marche, Ascoli Piceno 23-24 maggio 2019).
LVII	Metodi e misure politiche per una ripartenza sostenibile (Evento on-line, 27 maggio 2021).