

VOLUME LXXI – N. 4

OTTOBRE-DICEMBRE 2017

# RIVISTA ITALIANA DI ECONOMIA DEMOGRAFIA E STATISTICA



**DIRETTORE**

CHIARA GIGLIARANO

**COMITATO SCIENTIFICO**

GIORGIO ALLEVA, GIAN CARLO BLANGIARDO, LUIGI DI COMITE, MAURO GALLEGATI, GIOVANNI MARIA GIORGI,  
ALBERTO QUADRO CURZIO, CLAUDIO QUINTANO, SILVANA SCHIFINI D'ANDREA

**COMITATO DI DIREZIONE**

CHIARA GIGLIARANO, CLAUDIO CECCARELLI, PIERPAOLO D'URSO, SALVATORE STROZZA, ROBERTO ZELLI

**REDAZIONE**

LIVIA CELARDO, MARIATERESA CIOMMI, ANDREA CUTILLO, GIUSEPPE GABRIELLI, ALESSIO GUANDALINI,  
SIMONA PACE, GIUSEPPE RICCIARDO LAMONICA, ANDREA SPIZZICHINO

*Sede Legale:* C/O Studio Associato Cadoni, Via Ravenna n.34 – 00161 ROMA.  
sieds.new@gmail.com, rivista.sieds@gmail.com

**SIEDS**  
**SOCIETÀ ITALIANA**  
**DI ECONOMIA DEMOGRAFIA E STATISTICA**

**CONSIGLIO DIRETTIVO**

*Presidenti Onorari:* LUIGI DI COMITE, GIOVANNI MARIA GIORGI

*Presidente:* FRANCESCO MARIA CHELLI

*Vice Presidenti:* CLAUDIO CECCARELLI, PIERPAOLO D'URSO,  
ROBERTO ZELLI

*Segretario Generale:* MATTEO MAZZIOTTA

*Consiglieri:* EMMA GALLI, CHIARA GIGLIARANO, STEFANIA GIRONE, LUCIANO NIEDDU,  
STEFANIA RIMOLDI, SILVANA MARIA ROBONE, SALVATORE STROZZA, CECILIA VITIELLO

*Segretario Amministrativo:* ALESSIO GUANDALINI

*Revisori dei conti:* FABIO FIORINI, SIMONE POLI, DOMENICO SUMMO

*Revisori dei conti supplenti:* MARGHERITA GEROLIMETTO, GIUSEPPE NOTARSTEFANO

**SEDE LEGALE:**

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

sieds.new@gmail.com

rivista.sieds@gmail.com

## INDICE

Andrea Fasulo, Alessio Guandalini, Marco D. Terribili <i>Google trends for nowcasting quarterly household consumption expenditure</i> .....	5
Claudio Ceccarelli, Diego Chianella, Filomena De Filippo, Cinzia Graziani, Alessio Guandalini, Marco Lattanzio, Loriga, Alessandro Martini, M. Dionisio Terribili <i>Quality improvements in variance estimation for the labour force survey</i> .....	15
Silvia Loriga, Andrea Spizzichino, Oriol Gisbert Martì, Ilaria Franco <i>Popolazione e titolo di studio: nuovi dati per analisi di lungo periodo</i> .....	27
Margherita Gerolimetto, Stefano Magrini <i>A finite sample appraisal of the distribution dynamics approach</i> .....	39
Simona Mastroluca, Simona Pace <i>L'armonizzazione delle indagini sociali nella nuova strategia del Censimento permanente</i> .....	49
Ciro Baldi, Sara Gigante, Silvia Pacini <i>Territorial wage inequality: the effect of employer, employee and job characteristics</i> .....	61
Federico Brogi, Valeria Quondamstefano, Fabrizio Solari <i>Il Censimento permanente: prime analisi sugli obiettivi di diffusione a livello comunale</i> .....	73
Maria Carmela Miccoli, Antonella Biscione <i>Famiglie, povertà e ricchezza: Italia e Spagna a confronto</i> .....	83
Andrea Ciccarelli, Adriana Monte, Gabriella Schoier <i>La produzione in proprio di beni alimentari. aspetti economici e sociali</i> .....	95
Andrea Cutillo, Eleonora Meli <i>Le scelte di consumo delle famiglie italiane</i> .....	107

Domenica Fioredistella Iezzi, Francesco Zarelli <i>Environmental education: a web mining analysis</i> .....	119
Mariateresa Ciommi, Chiara Gigliarano, Stefania Taralli, Francesco M. Chelli <i>The equitable and sustainable well-being at local level: a first attempt of time series aggregation</i> .....	131
Francesco Bozzo, Stefania Girone, Vincenzo Fucilli, Alessandro Petrontino, Ruggiero Sardaro <i>The consumption of fish in old age: a statistical analysis on the dataset “multiscopo”</i> .....	143

## **GOOGLE TRENDS FOR NOWCASTING QUARTERLY HOUSEHOLD CONSUMPTION EXPENDITURE**

Andrea Fasulo, Alessio Guandalini, Marco D. Terribili

### **Introduction**

Italian National Institute of Statistics (Istat) carries out an important sample survey about Italian households expenditures on goods and services; the observed expenditure is obviously considered in the computation of the aggregates on which national accounts are based, first and foremost the Gross Domestic Product (GDP). For this reason, in the calculation of expenditure estimates, the quickness represents an important issue: estimating these indicators in a swifter way would involve a remarkable improvement of the process which leads to the final statistical indicator. Exploiting innovative data sources potentiality can allow to introduce forecasting and nowcasting in the estimation process.

Big data exploitation is actually a reality in the official statistics field: web data is under study for the production of official statistics indicators, such as unemployment rate, forecasted using data from the web (D'Alo *et al.*, 2015), and consumer price index, calculated using prices collected using webscraping techniques (Polidoro *et al.*, 2015).

The "3Vs" (volume, velocity, variety) definition model for describing big data can lead to better estimates which exploit their informative potentiality, also in complex fields as small area estimation (Falorsi *et al.*, 2017) or forecasting.

One of the most innovative data source used for this purpose is Google Trends, a web facility provided by the web search, which allows every user to evaluate how the research flows of a set of keywords change during an observation period. This longitudinal data can increase the predictive power of a time series baseline model. Google Trend data has many important assets: they are weekly update, they can be referred to one or several keywords and they can be circumscribed in the space (referring to a country or to a region) and in the time (referring to a period).

The paper is organized as follow: in Section 1 the household budget survey carried out by the Italian National Institute of Statistics is described. In Section 2 the Google trend tool is presented and the keywords chosen for catching the trends of expenditure are listed. In section 3 the methodology used for the estimation of

model for nowcasting and the results are discussed. Finally, in Section 4 conclusions and further research.

## 1. The household budget survey

The household budget survey (HBS)<sup>1</sup> supplies information on the evolution of level and composition of household consumption expenditure, according to their main socioeconomic and territorial characteristics. It provides the official estimates of relative and absolute poverty in Italy and the measure of inflation by household expenditure classes.

The HBS is a continuous sample survey carried out by ISTAT in each week of the whole year. The sample is a two-stage sample of municipalities and household. The municipalities are stratified by typology<sup>2</sup> and population size within regions. The size of the yearly sample consists in around 500 municipalities and 20 thousand households.

The survey collects data on all expenditures incurred by resident household to purchase goods and services exclusively devoted to household consumption. The data are collected through three questionnaires:

- *internal interview*, in which the main socio-demographic characteristics of the household, each member and the house are collected; furthermore, the possession of durable goods, expenditure for transport and communication are reported;
- *daily diary*, in which all the members indicate the expenses incurred during a period of 14 days, the amount of consumed or donated self-produced goods and the place of purchase of goods and services;
- *final interview*, in which information on other expenditure are gathered.

The survey provides data on the household consumption expenditure for different consumption purposes (ECOICOP)<sup>3</sup>, geographical areas and household typologies. The yearly estimates are released in July of each year and are related at the consumption expenditure of the previous year ( $t - 1$ ). Furthermore, quarterly estimates are provided to Nation Account for compiling the national accounts

---

<sup>1</sup> See <http://siqua.istat.it/SIQual/visualizza.do?id=0021002>.

<sup>2</sup> The municipalities are classified in Metropolitan area, neighborhood of metropolitan area and municipalities with more than 50 thousand inhabitants, municipalities with less than 50 thousand of inhabitants not included in the neighborhood of metropolitan area.

<sup>3</sup> Classification of individual consumption by purpose. For further information see <https://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=5>.

statistics. This data are available at 45/50 days after the end of the quarter, then with a quarter of delay.

## 2. Google Trend

The time series query share extracted from Google Trends is used as an auxiliary variable to improve the model-based nowcast.

GT is a web tool that provides sample time series computed as ratio between the number of queries for a keywords at time  $t$  weighted with the total volume of searches at time  $t$  and the maximum number of queries for the keywords throughout the period considered, weighted with the volume of total searches.

Actually, GT supplies a query index data for a group of ten keywords at most. Therefore, a group of keywords has been selected according to the topics in the HBS interviews. Then, among these, ten keywords with the most powerful trends have been included in the query index data. The list of the keywords used is as follow:

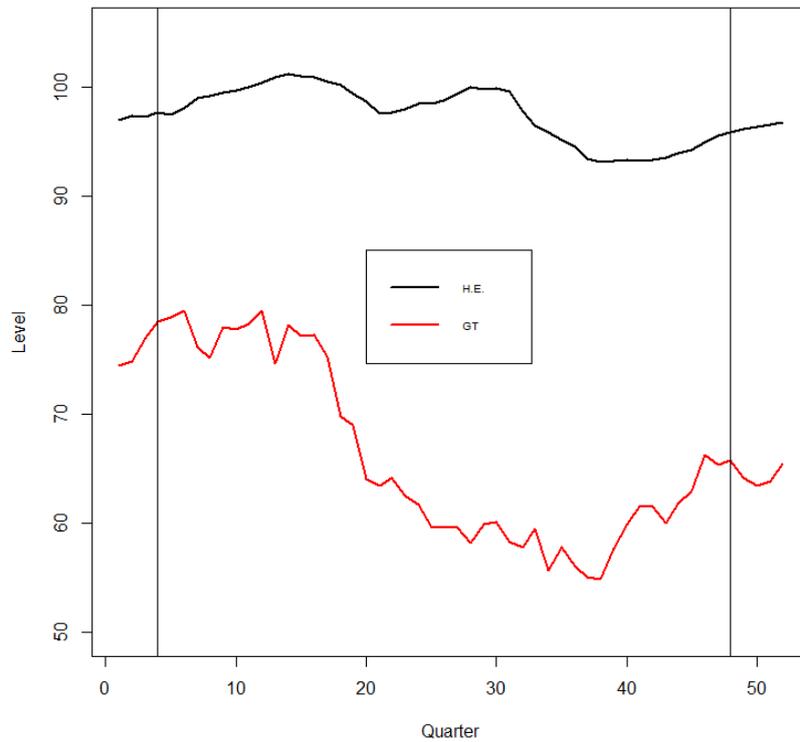
- *abbigliamento* (clothing);
- *bollette* (bills);
- *finanziamento* (financing);
- hotel;
- IKEA;
- *orologi* (watches);
- *prestito* (loan);
- *preventivo* (estimate);
- *ristrutturazione* (refurbishment);
- *voli* (flights).

## 3. Methodology

In this section, we will describe the relevant methodological statistical approach used for the study. Our statistical model is implemented in R software. The paper focused on the study of nowcasting quarterly households consumption expenditures, using the seasonally adjusted data sample series from first quarter of 2004 to fourth quarter of 2016. Furthermore, GT data for the keywords described in the section 2 is used for the query index data. The GT series has been seasonally adjusted using the X-13ARIMA-SEATS method. Figure 1 shows the structural

trends for the two series, the household consumption expenditure (black line) and the GT series (red line). The time series are index number and the same base is used (3<sup>rd</sup> quarter of 2006). A very simple baseline forecasting model has been studied, based on the series from first quarter of 2005 to fourth quarter of 2015, modelling the disturbance terms as an Auto Regressive Integrated Moving Average (ARIMA) model with order (1,1,0).

**Figure 1** – Level of household consumption expenditure (H.E.) vs. Google Trend series level (GT). 1<sup>st</sup> quarter of 2005 – 4<sup>th</sup> quarter of 2015, Italy.



Note: time series level are index number and the same base is used (3<sup>rd</sup> quarter of 2006).

We next add to the baseline model the information coming from GT. The GT predictive power has been validated by comparing the results obtained with the baseline model and other models that take into account important leading indicators of the household consumption series (Vosen & Schmidt, 2011a, 2011b).

In particular has been added information about GDP, interest rate of the Italian Ordinary Treasury Bills (OTB) and Consumer Confidence index (CC). All the auxiliary variables have been tested with different lag structures considering information at the time  $t$ ,  $t - 1$ ,  $t - 2$ ,  $t - 3$  and  $t - 4$ .

The goodness of fit of the models were carried out studying both the residual correlation and distribution by means of the Ljung-Box (8 lags) and the Jarque-Bera test. To evaluate the explanatory power of the different external variables considered in the application and so to compare the quality of each model has been considered also the AIC, BIC and log-likelihood values.

The 1-step ahead nowcast of the four quarters of the 2016 has been carried out using a rolling regression procedure for all the models and the results has been compared by means of two classic indicators of forecast accuracy, the Relative Root Mean Squared Error (RRMSE) and the Mean Percentage Error (MPE). The indicators are formulated as follows:

$$RRMSE = 100 \frac{\sqrt{\frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{T}}}{\bar{y}}$$

$$MPE = \frac{100}{T} \sum_{t=1}^T \frac{(y_t - \hat{y}_t)}{y_t}$$

where  $T$  are the quarters available,  $y_t$  and  $\hat{y}_t$  are respectively the actual estimates for the survey and the forecast value at the time  $t$ . All the models that will be compared in the next section are summarized in the list below:

- Model 1: the ARIMA benchmark model;
- Model 2: the augmented benchmark model including GT;
- Model 3: the augmented benchmark model including GDP;
- Model 4: the augmented benchmark model including OTB;
- Model 5: the augmented benchmark model including CC.

A significance comparison will be carried out through the Harvey-Leybourne-Newbold (1997) modification of the Diebold-Mariano (1995) test statistic for equal forecast accuracy. This aim of the test is to understand if the differences between the MSEs are statistically significant.

#### 4. Results

The case-study follow the steps identified by the Box-Jenkins approach (Box & Jenkins, 1970). The figure 1 shows an absence of stationarity for the household expenditure consumption series. The presence of a unit root is confirmed by the augmented Dickey-Fuller test (p-value=0.9) and the KPSS stationarity test (p-value<0.01), so a first difference has been made on the data. The model that best fits the correlation and auto-correlation plot is an ARIMA (1,0,0).

The goodness-of-fit of the models is shown in Table 2, in which Model 1 refers to the benchmark model, Model 2 utilizes the lagged GT series, while Model 3-5 include the lagged series respectively of GDP, OTB, and CC.

**Table 1** – Goodness-of-fit results.

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
AR (1)	0.67***	0.65***	0.68***	0.66***	0.66***
GT (t – 1)		0.05 **			
GDP (t – 3)			-0.05		
OTB(t – 3)				0.002	
CC(t – 3)					0.0007
AIC	54,50	52,68	52,59	52,08	57,36
BIC	58,22	51,16	57,94	57,43	62,91
Log-Lik	-25,28	-23,34	-23,35	-23,04	-25,68
Residual Diagnostics					
Mean	-0,0030	-0,0004	-0,0010	-0,0080	-0,0040
Skewness	-0,96	-0,87	-0,92	-0,9	-0,96
Kurtosis	1,87	1,47	1,8	2,35	2,04
Jb normal test	0.0003	0,001	0,000009	0,0000004	0,000001
LB test (lag 8)	>0.05	>0.05	>0.05	>0.05	>0.05

Significance codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

In all the models the AR component is significant and with positive values, while just the GT auxiliary variables shows a significant coefficient (p=<0.05). Model 2 shows very good results in terms of AIC, BIC and log-likelihood. Very good performance are provided by Model 4 too, but the OTB coefficient estimate is not significant (p>0.1). Not good results are showed from all the models for the residuals distribution. In fact all the models refuse the Jarque-Bera null hypothesis

of normality ( $p\text{-value} < 0.1$ ). Looking the results for the Ljung-Box test is possible to accept for all the models the null hypothesis that the residual are independently distributed, so the correlations in the data, up to 8 lags, are supposed to be 0.

The four quarters of the 2016 have been nowcasted with the five models. Table 3 shows the comparison of the forecast made by means of the RRMSE and MPE indicators. Very interesting results are obtained by Model 2 and Model 5 both in terms of RRMSE and MPE. A direct comparison between the baseline model and the GT model shows that using GT information the Mean Squared Error reduce by 14%. It's important to underline that just the models 1-2 present statistically significant coefficients.

**Table 2** – RRMSE and MPE for 1-step ahead prediction.

	Model 1	Model 2	Model 3	Model 4	Model 5
RRMSE (%)	0,068	0,063	0,080	0,079	0,060
MPE (%)	0.058	0.054	0.069	0.056	0.049

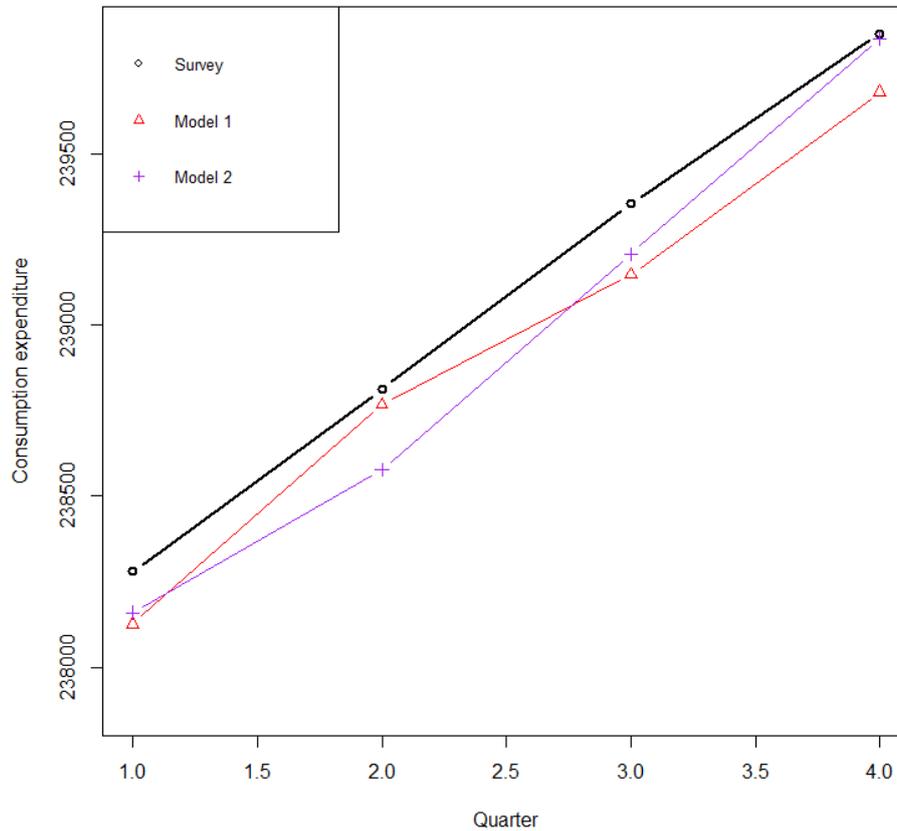
Significance level is determined in Table 3 through the Harvey-Leybourne-Newbold modification of the Diebold-Mariano test. The table shows that although the Google model shows almost always better results of all the other indicators, the Diebold-Mariano statistics are not significant.

**Table 3** – Diebold-Mariano test.

	Model 2/Model 3	Model 2/Model 4	Model 2/Model 5
DM test	1.19	-0.14	0.35

Significance codes: \*\*\*\* 0.01 \*\*\* 0.05 \*\* 0.1 \* 1

The household expenditure consumption series with the nowcast estimates obtained using Model 1 and Model 2 for 1-step ahead prediction are plotted in Figure 2. Both the models underestimate the true value in all the four quarters studied and the model including GT information outperforms the benchmark model over three quarters of four.

**Figure 2** – 1-step ahead nowcast estimates. Four quarters of the 2016, Italy.

## 5. Conclusions and further research

Google trends data exploitation which has been shown in this paper represents a promising approach for nowcasting quarterly households expenditures estimation. The proposed models give good results both in terms of predictive capability and in terms of goodness of fit.

Nowcasting would guarantee to fill the informative gap between the issue of estimations and the reference period, providing reliable and useful estimates to the users in a least time.

Further developments to this paper could be a deeper study of the analyzed keywords of which google trends data have been downloaded and exploited, in order to further improve the predictive models.

Moreover applying this method to other statistical field (mainly Gross Domestic Product and Consumer Confidence Indicator) can be interesting and it may lead to results as good as those presented in the present paper.

## References

- ASKITAS N., ZIMMERMANN K.F. 2009. Google Econometrics and Unemployment Forecasting. *Applied Economics Quarterly*, Vol. 55, No. 2, pp. 107-120.
- BORTOLI C., COMBES S. 2015. Contribution from Google Trends for forecasting the short-term economic outlook in France: limited avenues. *Insee – Conjoncture in France*, March 2015, pp. 43-55.
- BOX G.E.P., JENKINS G.M. 1970, Time Series Analysis, *Forecasting and Control*. Holden Day, San Francisco
- D'ALÒ M., FALORSI S., FASULO A. 2015, Monthly unemployment rate prediction with Google Trends data: does Google search data improve the nowcast of Italian labour market?. *Big Data and the complexity of Labour Market Policies: New Approaches in Regional and Local Labour Market Monitoring for Reducing Skills Mismatches*, Larsen C., Rand S., Schmid A., Mezzanica M., Dusi S. (Eds.), Reiner Hamp Verlag. 99-114.
- FALORSI S., FASULO A., NACCARATO A., PRATESI M. 2017, 61<sup>st</sup> World Statistics Congress ISI 2017, *Small Area model for Italian regional monthly estimates of young unemployed using Google Trends Data*  
[https://www.researchgate.net/publication/320554956\\_Small\\_Area\\_model\\_for\\_Italian\\_regional\\_monthly\\_estimates\\_of\\_young\\_unemployed\\_using\\_Google\\_Trends\\_Data](https://www.researchgate.net/publication/320554956_Small_Area_model_for_Italian_regional_monthly_estimates_of_young_unemployed_using_Google_Trends_Data)
- POLIDORO F., GIANNINI R., LO CONTE R., MOSCA S., ROSSETTI F. 2015. Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation. *Statistical Journal of the IAOS*, Vol. 31, no. 2, pp. 165-176, 2015
- VOSEN S., SCHMIDT T. 2011a. Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, Vol. 30, No. 6, pp. 565-578.
- VOSEN S., SCHMIDT T. 2011b. A monthly consumption indicator for Germany based on internet search query data. *Ruhr Economic Papers*, No. 208. ISSN: 1864-4872 (online).

## SUMMARY

### **Google Trends for Nowcasting Quarterly Household Expenditure**

During last years, several studies focused on the predictive capability of web data to forecast statistical indicators. Google Trends is a free web tool that quantify search-term volume on the search engine. The aim of this work is to forecast the household expenditures for consumption in Italy, using Google Trends related to particular expenditure keywords. Several prediction models have been tested, also including relevant leading indicators correlated to the household expenditures behavior, on a time series survey data from 2004 to 2016. The ARIMA model has been chosen and models with different lag structures have been tested. The model comparison shows that the results including Google data outperform those of both benchmark and augmented models.

---

Andrea FASULO, Italian National Institute of Statistics, fasulo@istat.it

Alessio GUANDALINI, Italian National Institute of Statistics, alessio.guandalini@istat.it

Marco D. TERRIBILI, Italian National Institute of Statistics, terribili@istat.it

## **QUALITY IMPROVEMENTS IN VARIANCE ESTIMATION FOR THE LABOUR FORCE SURVEY**

Ceccarelli C., Chianella D., De Filippo F., Graziani C., Guandalini A.,  
Lattanzio M., Loriga S., Martini A., Terribili M.D.

### **1. Introduction**

The analysis of the results of a sample survey should always include an assessment of the accuracy of the estimates, a measure of their dispersion around the true, unobserved value of the target variable in the population. Furthermore it is a good practice for national statistical institutes to estimate sampling errors and then to disseminate them in a transparent and clear way for the users.

Dissemination strategy for Italian Labour Force Survey (LFS) include a quite large set of accuracy indicators for the main results produced. Monthly, quarterly and yearly press releases include evaluation of precision for the main indicators on labour market. These precision indicators are currently transmitted to Eurostat for both the quarterly and the yearly Quality Accuracy Reports. Istat received recently solicitations coming from Eurostat and the European Commission's Directorate-General for Employment, Social Affairs and Equal Opportunities to broaden the accuracy evaluations provided, considering also variance estimations for LFS-based indicators' annual net changes. The purpose of this document is to present and explain the methodology developed for the calculation of the annual net changes variance's estimate for Italian LFS and the first results obtained.

### **2. The Italian LFS sampling design**

The Italian LFS (Istat, 2006) sampling design is a two stage sampling design with stratification of the primary units (PSUs). The PSUs are the municipalities, while the second stage units (SSUs) are the households. In each NUTS3 domain, large municipalities with population over a certain threshold, are always included in the sample (self-representative strata: SR-strata). Smaller municipalities (not self-representative: NSR-strata) are grouped into strata, composed by municipalities having similar population size; then one municipality is selected in

each stratum with probability proportional to its population. Households are randomly selected (SRS) from the population frame, in all the municipalities selected at the first stage<sup>1</sup>.

The quarterly sample is uniformly spread over the 13 weeks of the quarter. The representativeness of the sample, that is the presence of all the strata, is guaranteed for each month (group of 4 or 5 weeks).

**Table 1 - Main features of the Italian LFS sampling design.**

*Q1:2015 and Q1:2016, mean values.*

	2015	2016
Overall number of municipalities	8.057	8.047
Number of municipalities in the sample	1.260	1.264
Theoretical sampling size (households)	71.536	71.536
Effective sampling size (households)	65.151	65.236
Population size (households)	25.792.738	25.766.327

*Source: Italian LFS.*

About 10% of the smallest municipalities is rotated every year: a new municipality is randomly selected from each of these strata (with probability proportional to its population). The households are rotated according to a 2-(2)-2 rotation scheme: households are interviewed during two consecutive quarters. Then, after a two-quarters break, they are again interviewed twice in the corresponding two quarters of the following year. As a result, each household is included in four waves of the survey along a 15-months period.

### 3. The estimation procedure

The calibration estimator (Deville & Särndal, 1992 and Särndal, 2007) is used to obtain Italian LFS estimates. The final weights are computed as follows:

1. First, design weights  $\delta_j$  are obtained as the inverse of the inclusion probabilities of any household in the sample.
2. Then, correction factors  $k_j$  are worked out as the reciprocal of the response ratios (computed for several kind of households and territorial domains), in order to take into account households non-response. Intermediate weights,

<sup>1</sup> For methodological details see Cochran (1977) and Conti & Marella (2011).

corrected for non-response, are then computed multiplying design weights by these correction factors ( $d_j = \delta_j k_j$ ).

3. Then, starting from intermediate weights  $d_j$ , final weights  $w_j$  are obtained solving a minimization problem under constraints. The function to be minimized is a distance between final and intermediate weights; the constraints regard the estimates of some auxiliary variables that have to be equal to the totals in the reference population derived by external sources. The main constraints are: population by gender and 14 5-years age groups (at NUTS2 level), population by gender and 5 age groups (at NUTS3 level and for the 13 biggest municipalities), monthly population by gender (at NUTS2 level), number of households by wave (at NUTS2 level), foreigner population (Male, Female, EU and Not EU at NUTS2 level). Final weights ensure that all members of a given household have the same weight. Through the calibration estimator, applying final weights, the sample reproduces the same distribution of the population according to the chosen auxiliary variables.

The estimate of the total of the variable  $Y$  is then computed as follows:

$$\hat{Y} = \sum_{j \in s} y_j w_j = \sum_{j \in s} y_j d_j \gamma_j = \sum_{j \in s} y_j d_j k_j \gamma_j.$$

Weights are computed on quarterly basis, whereas annual estimates are computed as averages of quarterly estimates.

#### 4. The variance estimation procedure

With every sampling figure, an estimate of its accuracy must be provided, so a measure of the variability of the estimates around the unknown “true” value of the parameter in the reference population is calculated, depending on both the partial observation of the phenomenon in the sample instead of the whole population and the adopted estimation procedure.

For each quarter, an evaluation of the accuracy of the survey figures is given by the variance of the estimator. Considering that the Italian LFS sampling design deals with SR and NSR strata and that the different strata are independent, the variance estimation is given by:

$$\hat{V}ar(\hat{Y}) \cong \hat{V}ar(\hat{Z}) = {}^{SR}\hat{V}ar(\hat{Z}) + {}^{NSR}\hat{V}ar(\hat{Z})$$

where  $Z$  is the linearization of  $Y$ ,  $h$  is the stratum index,  $H_{SR}$  is the set of  $SR$  strata,  $H_{NSR}$  for  $NSR$  strata taking into account that the calibration estimator is not linear.

In  $SR$  strata the estimation of the variance is computed among the observation of the variable  $Y$  in the households, according to this formula:

$${}^{SR}\hat{V}ar(\hat{Z}) = \sum_{h \in H_{SR}} M_h \frac{(M_h - m_h)}{m_h(m_h - 1)} \sum_{j=1}^{m_h} (z_{hj} - \bar{z}_h)^2$$

where  $\bar{z}_h = \frac{1}{m_h} \sum_{j=1}^{m_h} z_{hj}$ . In  $NSR$  strata, the corresponding formula for the variance estimation is given by:

$${}^{NSR}\hat{V}ar(\hat{Z}) = \sum_{h \in H_{NSR}} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left( \hat{z}_{hi} - \frac{\hat{z}_h}{n_h} \right)^2$$

where  $\hat{z}_{hi} = \sum_{j=1}^{m_{hi}} z_{hij} w_{hij}$  and  $\hat{z}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} z_{hij} w_{hij}$ . that is in  $SR$  strata the variance is computed among the households (SSUs), while in  $NSR$  strata it is computed among the municipalities (PSUs) that is considering only the leading contribution.

Considering that the LFS sample is selected on quarterly basis, the variance estimation is conducted for each quarterly estimate.

Istat along the years has developed specific IT tools in order to provide accuracy estimation for sampling surveys estimates, that implement the methodology above mentioned. Genesees (Generalized Software for Sampling Estimates and Errors in Surveys) was a SAS-based application to calculate estimates and sampling errors for complex designs, recently ReGenesees (R Evolved Generalized Software for Sampling Estimates and Errors in Surveys<sup>2</sup>) is a full-fledged R software for design-based and model-assisted analysis of complex sample surveys.

In this work we used ReGenesees, as we currently do in the context of the LFS.

## 5. The net changes variance estimation

The net change  $\Delta_Y$  is the difference between yearly estimates. The plug-in estimator for net changes can be written as:

$$\hat{\Delta}_Y = {}^t\hat{Y} - {}^{t-1}\hat{Y} \quad (1)$$

---

<sup>2</sup> See Zardetto (2015).

where  ${}^t\hat{Y}$  is the yearly total estimator of  $Y$  for the year  $t$  and  ${}^{t-1}\hat{Y}$  is the same estimate but related to the previous year. In the Italian LFS, the yearly estimates are obtained as average of four quarterly estimates of totals:

$${}^t\hat{Y} = \frac{1}{4} \sum_{q=1}^4 {}^t_q\hat{Y} \quad (2)$$

where  $t$  is the year and  $q$  ( $= 1, 2, 3, 4$ ) are the quarters. The Italian LFS, as described in paragraph 2, adopts a 2-(2)-2 rotation scheme. It means that a rotation group of households remains in the sample for two consecutive quarters, then it leaves the sample for two quarters, and then it is considered again in the sample for other two consecutive quarters. Then it is dropped out definitely. The theoretical overlapping is equal to 50% for two consecutive quarterly samples and for same quarters in two consecutive years, it's equal to 25% with a lag of three or five quarters, otherwise there is no overlapping.

**Figure 1** – Rotating design of LFS and overlapping of a quarterly theoretical sample in two years. Overlapping,  $o_{q'(\blacksquare)}^{q(\blacksquare)}$ , has been expressed considering the quarter 1 of year 2 as base.

Year	Quarter	Rotation groups																$o_{q'(\blacksquare)}^{q(\blacksquare)}$
		D	E	F	G	H	I	L	M	N	O	P	Q	R	S	T		
1	4	D4	E3			H2	I1											50%
2	1		E4	F3			I2	L1										---
2	2			F4	G3			L2	M1									50%
2	3				G4	H3			M2	N1								0
2	4					H4	I3			N2	O1							25%
3	1							I4	L3		O2	P1						50%
3	2								L4	M3		P2	Q1					25%
3	3									M4	N3		Q2	R1				0
3	4										N4	O3		R2	S1			0
4	1											O4	P3		S2	T1		0

Source: Italian LFS.

The variance estimator of the yearly estimate is given by:

$$\hat{V}ar(\ ^t\hat{Y} ) = \hat{V}ar\left(\frac{1}{4}\sum_{q=1}^4\ ^t\hat{Y}_q\right) = \frac{1}{16}\left(\sum_{q=1}^4\hat{V}ar(\ ^t\hat{Y}_q) + 2\sum_{q\neq q'}\hat{C}ov(\ ^t\hat{Y}_q, \ ^t\hat{Y}_{q'})\right). \quad (3)$$

It depends on the  $\hat{V}ar(\ ^t\hat{Y}_q)$ , the variance estimator of quarter  $q$ , and  $\hat{C}ov(\ ^t\hat{Y}_q, \ ^t\hat{Y}_{q'})$ , that is the covariance estimator between two quarterly estimates. The covariance estimator can be written as the product of the auto-correlation between estimates of two quarters and the square root of the product of the quarterly variance estimates:

$$\hat{C}ov(\ ^t\hat{Y}_q, \ ^t\hat{Y}_{q'}) = \hat{\delta}_{q'(t)}^{q(t)} \hat{\rho}_{q'(t)}^{q(t)} \sqrt{\hat{V}ar(\ ^t\hat{Y}_q)\hat{V}ar(\ ^t\hat{Y}_{q'})} \quad q = 1, 2, 3, 4$$

where  $\hat{\delta}_{q'(t)}^{q(t)}$  is the partial overlapping of quarterly samples of year  $t$  due to the rotation scheme used in LFS and  $\hat{\rho}_{q'(t)}^{q(t)}$  is the auto-correlation between estimates of two quarters. For a more accurate estimate of  $\hat{C}ov(\ ^t\hat{Y}_q, \ ^t\hat{Y}_{q'})$  we compute the effective overlapping between quarters, that is lower than the theoretical one due to the attrition. The auto-correlations,  $\hat{\rho}_{q'(t)}^{q(t)}$ , are computed without sampling weights. In the calculation are taken into account three auto-correlations at one quarter and one auto-correlation at three quarters.

Then, the estimator of variance of net change in (1) is:

$$\hat{V}ar(\hat{\Delta}_Y) = \hat{V}ar(\ ^t\hat{Y}) + \hat{V}ar(\ ^{t-1}\hat{Y}) - 2\hat{C}ov(\ ^t\hat{Y}, \ ^{t-1}\hat{Y}) \quad (4)$$

where  $\hat{V}ar(\ ^t\hat{Y})$  and  $\hat{V}ar(\ ^{t-1}\hat{Y})$  are the variance estimators in (3), while  $\hat{C}ov(\ ^t\hat{Y}, \ ^{t-1}\hat{Y})$  is the estimator of the covariance between yearly estimates. In this case, in the calculation of the covariance term are taken into account seven auto-correlations at one quarter, five auto-correlations at three quarters, four auto-correlations at four quarters and three auto-correlations at five quarters.

All the expression formulas above hold also for the linearized version of  $\hat{Y}$ .

**Table 2 – 2015-2016 Net changes accuracy evaluation for the list of 23 indicators requested (part 1).**

INDICATOR DESCRIPTION	V_2015	V_2016	SE	VAR	CHANGE			SIGNIFICANT
					POINT ESTIMATION	C.I.		
						LB	UB	
Newly employed, age group 20-64, in %	10,64	11,01	0,1233	0,0152	0,4	0,1	0,6	YES
Employment rate of older workers, age group 55-64, in %	48,24	50,33	0,1553	0,0241	2,1	1,8	2,4	YES
Temporary employees as percentage of employees, age group 20-64, in %	13,81	13,81	0,1128	0,0127	0,0	-0,2	0,2	NO
Youth unemployment rate, age group 15-24, in %	40,32	37,77	0,6019	0,3623	-2,6	-3,7	-1,4	YES
Unemployment rate, age group 15-74, in %	11,93	11,72	0,0963	0,0093	-0,2	-0,4	0,0	YES
NEET rate, age group 18-24, in %	27,91	26,04	0,3113	0,0969	-1,9	-2,5	-1,3	YES
Employment rate 20-64, total, in %	60,53	61,56	0,0663	0,0044	1,0	0,9	1,2	YES
Employment rate 20-64, men, in %	70,60	71,68	0,0977	0,0095	1,1	0,9	1,3	YES
Employment rate 20-64, women, in %	50,62	51,59	0,0915	0,0084	1,0	0,8	1,2	YES
Early leavers from education and training, age group 18-24, in %	14,70	13,78	0,1952	0,0381	-0,9	-1,3	-0,5	YES
Tertiary educational attainment, age group 30-34, in %	25,35	26,17	0,2626	0,0690	0,8	0,3	1,3	YES

V\_2015 = estimated value in 2015; V\_2016 = estimated value in 2016; SE = relative error; VAR = Sampling Variance; CHANGE (Point Estimation) = V\_2016 - V\_2015; CHANGE (C.I.) = 95% Confidence Interval of Change, lower bound (LB) and upper bound (UB); SIGNIFICANT = YES if the change is significantly different from 0, NO otherwise.

Source: Italian LFS

**Table 2 – 2015-2016 Net changes accuracy evaluation for the list of 23 indicators requested (part 2).**

INDICATOR DESCRIPTION	V_2015	V_2016	SE	VAR	CHANGE			SIGNIFICANT
					POINT ESTIMATION	C.I. LB	C.I. UB	
Youth employment rate, age group 20-29, in %	40,29	42,06	0,2453	0,0602	1,8	1,3	2,3	YES
Employment rate of low skilled persons, age group 20-64, in %	49,39	50,32	0,1178	0,0139	0,9	0,7	1,2	YES
Employment rate of non-EU nationals, age group 20-64, in %	60,72	61,52	0,2846	0,0810	0,8	0,2	1,4	YES
Part-time employment rate, age group 20-64, in %	18,24	18,49	0,0832	0,0069	0,3	0,1	0,4	YES
NEET rate, age group 15-24, in %	21,39	19,87	0,2261	0,0511	-1,5	-2,0	-1,1	YES
Share of adult population with upper secondary or tertiary education, age group 25-64, in %	59,90	60,14	0,0514	0,0026	0,2	0,1	0,3	YES
Lifelong learning - Percentage of adult population aged 25-64 participating in education and training	7,33	8,26	0,0792	0,0063	0,9	1,1	0,8	YES
Youth unemployment rate, age group 15-24, in %	10,56	10,05	0,1943	0,0377	-0,5	-0,9	-0,1	YES
Long term unemployment rate, age group 15-74, in %	6,93	6,72	0,0782	0,0061	-0,2	-0,4	-0,1	YES
Involuntary temporary employment, age group 20-64, in %	10,20	10,13	0,1780	0,0317	-0,1	-0,4	0,3	NO
Involuntary part-time employment as percentage of the total part-time employment, age group 20-64, in %	65,45	64,24	0,3400	0,1156	-1,2	-1,9	-0,5	YES

V\_2015 = estimated value in 2015; V\_2016 = estimated value in 2016; SE = relative error; VAR = Sampling Variance; CHANGE (Point Estimation) = V\_2016 - V\_2015; CHANGE (C.I.) = 95% Confidence Interval of Change, lower bound (LB) and upper bound (UB); SIGNIFICANT = YES if the change is significantly different from 0, NO otherwise.

Source: Italian LFS

## **6. First results**

As presented at the December 2014 meeting of the Eurostat Labour Market Statistics Working Group (LAMAS), DG EMPL of the European Commission asked Eurostat to provide from 2017 onwards estimates of the variance of annual net changes for a number of LFS based indicators. The detailed list of indicators for which variance estimates of annual net changes have been requested was defined in the annex 1 of LAMAS document Eurostat/F3/LAMAS/29/15.

According to these definitions and the above described methodology the level estimates for 2015 and 2016 as well as the annual net change 2015-2016 have been calculated.

In particular the same SAS macro statement for indicators calculation and linearization of the target variables have been used.

Quarterly standard errors have been calculated applying the usual tool (ReGenesees) and methods in order to guarantee coherence with the accuracy evaluation currently defined and disseminated.

## **7. Conclusions**

With every sampling figure, an estimate of its accuracy must be provided, so a measure of the variability of the estimates around the unknown “true” value of the parameter in the reference population is calculated, depending on both the partial observation of the phenomenon in the sample instead of the whole population and the adopted estimation procedure.

The aim of this work is developing the estimation of the sampling variance of net changes between yearly estimates of the Italian LFS. This information is fundamental in order to analyze the significance of the difference year-to-year of a certain variable, properly taking into account of the sampling error and it has been required by Eurostat and the European Commission for the main policy indicators computed over LFS data.

According to specific National needs, the goal of the work has been furtherly enlarged in order to develop the estimation of the sampling variance of net changes between a quarterly estimate and the estimate in the corresponding quarter of the previous year, in order to allow the evaluation of the significance of annual changes, each quarter. Moreover the territorial domain of the estimates for which these variances have been computed is NUTS2.

## Acknowledgements

The present work has been realized within the grant for the project “Quality improvements for the Labour Force survey” supported by Eurostat and the European Commission's Directorate-General for Employment, Social Affairs and Equal Opportunities.

## References

- COCHRAN W. 1977. *Sampling Techniques*. Wiley & Sons Ltd., New York, 3rd edition.
- CONTI P.; MARELLA D. 2011. *Campionamento da popolazioni finite: teoria e tecnica*. Springer, Milano.
- DEVILLE J.C., SÄRNDAL C.E. 1992. Calibration Estimators in Survey Sampling. *Journal of the American Statistical Association*, Vol. 87, 376–382.
- ISTAT. 2006. La Rilevazione sulle Forze di Lavoro: Contenuti, Metodologie, Organizzazione. Collana Metodi e norme, No. 32.
- LE COROLLER G. 2017. *Development of methodology for the estimate of variance of annual net changes for LFS-based indicators - Deliverable 2 (Annex) – Start guide (Final)*. Available on-line: [https://circabc.europa.eu/webdav/CircaBC/ESTAT/labmarstatwg/Library/statistics\\_working/2015/2.%207-8%20December%20\(Section%20A\\_%20LFS\)/2.%20Documents/Doc%2029%20Item%204.1%20Annex%20II%20SAS%20programs%20explanation.pdf](https://circabc.europa.eu/webdav/CircaBC/ESTAT/labmarstatwg/Library/statistics_working/2015/2.%207-8%20December%20(Section%20A_%20LFS)/2.%20Documents/Doc%2029%20Item%204.1%20Annex%20II%20SAS%20programs%20explanation.pdf)
- OSIER G., RAYMOND V. 2017. *Development of methodology for the estimate of variance of annual net changes for LFS-based indicators - Deliverable 1 - Short document with derivation of the methodology (FINAL)*. Available on-line: [https://circabc.europa.eu/webdav/CircaBC/ESTAT/labmarstatwg/Library/statistics\\_working/2015/2.%207-8%20December%20\(Section%20A\\_%20LFS\)/2.%20Documents/Doc%2029%20Item%204.1%20Annex%20II.pdf](https://circabc.europa.eu/webdav/CircaBC/ESTAT/labmarstatwg/Library/statistics_working/2015/2.%207-8%20December%20(Section%20A_%20LFS)/2.%20Documents/Doc%2029%20Item%204.1%20Annex%20II.pdf)
- SÄRNDAL C.E. 2007. The calibration approach in survey theory and practice. *Survey Methodology*, Vol. 33, No. 2, pp. 99–119.
- ZARDETTO D. 2015. ReGenesees: an Advanced R System for Calibration, Estimation and Sampling Error Assessment in Complex Sample Surveys. *Journal of Official Statistics*, Vol. 31, No. 2, pp. 177–203.

## SUMMARY

The Italian National Institute of Statistics, along with other European countries, is involved in the improvement efforts for the Labour Force Survey (LFS) in the context of the new IESS (Integrated European Social Statistics) framework regulation.

The IESS framework regulation asks for several actions that have to be undertaken by National Institutes of Statistics in order to achieve improvements in LFS. Among these actions, additional indicators for the annual point and annual net changes are request. Furthermore, the related estimates of variance of the estimates must be produced.

The requested indicators are mainly simple ratios (annual point) or difference of ratios (annual net changes), therefore nonlinear parameter must be linearized to derive the estimates of variance. Moreover, when computing the sampling variance the overlapping among the samples, due to the rotation scheme adopted in the Italian LFS, has to be taken into account. In this work the methodology used to produce annual point, annual net changes and the related variance estimations is illustrated.

---

Claudio CECCARELLI, Istituto nazionale di statistica, cceccar@istat.it

Diego CHIANELLA, Istituto nazionale di statistica, chianella@istat.it

Filomena DE FILIPPO, Istituto nazionale di statistica, defilipp@istat.it

Cinzia GRAZIANI, Istituto nazionale di statistica, cingraziani@istat.it

Alessio GUANDALINI, Istituto nazionale di statistica, alessio.guandalini@istat.it

Marco LATTANZIO, Istituto nazionale di statistica, lattanzio@istat.it

Alessandro MARTINI, Istituto nazionale di statistica, alemartini@istat.it

Silvia LORIGA, ISTAT, Istituto nazionale di statistica, siloriga@istat.it

Marco D. TERRIBILI, ISTAT, Istituto nazionale di statistica, terribili@istat.it



## **POPOLAZIONE E TITOLO DI STUDIO: NUOVI DATI PER ANALISI DI LUNGO PERIODO**

Silvia Loriga, Andrea Spizzichino, Oriol Gisbert Martí, Ilaria Franco

### **1. Introduzione**

L'interesse per serie storiche sempre più lunghe, riferite a diversi temi, che consentano di studiare le trasformazioni avvenute nella nostra società è sempre maggiore. Per la popolazione e nello specifico per il mercato del lavoro riveste grande importanza l'analisi dei principali aggregati per titolo di studio.

Come noto i cambiamenti avvenuti nel tempo nella rilevazione sulle forze di lavoro (RFL) hanno reso non confrontabili dati riferiti a periodi in cui le tecniche o le definizioni erano differenti. Per consentire analisi di medio e lungo periodo l'Istat si è impegnato nella ricostruzione di serie storiche al fine di eliminare i break causati da cambiamenti nella rilevazione.

Il lungo processo di disaggregazione delle serie ricostruite ha riguardato sia variabili socio demografiche come il genere, l'età e il territorio, sia variabili strettamente legate al mercato del lavoro tra cui la posizione nella professione, il settore d'attività economica, la tipologia di orario lavorativo etc.

In questo lavoro viene presentato lo step effettuato recentemente nel processo di ricostruzione di serie storiche relative al mercato del lavoro. Sono state disaggregate tutte le serie già prodotte negli ultimi anni rispetto alla variabile titolo di studio, al fine di conoscere le caratteristiche dei principali aggregati del mercato del lavoro anche con riferimento al massimo livello di istruzione raggiunto.

Questo lavoro illustra la metodologia adottata per la ricostruzione ponendo l'attenzione in particolare sulle scelte fatte per garantire un'informazione che sia in linea con le caratteristiche contemporanee dei livelli di istruzione. Vengono inoltre analizzati i principali risultati mettendo in evidenza l'evoluzione della distribuzione per titolo di studio della popolazione e la relazione di questa con la partecipazione al mercato del lavoro.

## 2. Quadro d'insieme

Tra i cambiamenti che nel corso degli anni sono intervenuti nella RFL, il più importante è sicuramente l'ultimo avvenuto nel 2004. Il cambio di indagine è stato caratterizzato da una profonda riorganizzazione del processo di produzione dei dati, resasi necessaria in parte per l'adeguamento alle nuove direttive europee, in parte per tenere conto dei cambiamenti intervenuti sul mercato del lavoro e, infine, per aggiornare le tecniche e le metodologie statistiche d'indagine.

L'effetto congiunto di queste innovazioni ha reso non confrontabili i dati prodotti dalla nuova RFL con i dati rilevati dalla RFL stessa prima del 2004, generando un break nelle serie storiche di dati e creando gravi problemi agli utenti per la continuità dell'analisi sia di lungo periodo sia di breve periodo. Per colmare tale *gap* informativo l'Istituto Nazionale di Statistica ha ritenuto opportuno effettuare una ricostruzione/raccordo delle serie storiche dei principali indicatori del mercato del lavoro relative al periodo precedente il IV trimestre 2003.

Per il raccordo delle serie si è adottato un approccio a livello macro, *model based* e per componenti; livello macro è stato scelto perché non è possibile effettuare una ricostruzione dei dati individuali della RFL a causa dell'assenza in questi delle informazioni necessarie alle nuove definizioni degli aggregati; si è inoltre scelto un approccio *model based* basato sull'utilizzo di tecniche econometriche e statistiche di analisi dei dati temporali.

L'approccio seguito, da ultimo, utilizza la struttura per componenti delle serie in quanto ricostruisce separatamente la componente di lungo periodo, quella di periodo annuale e la componente di breve periodo che tradizionalmente prendono i nomi di ciclo-trend, stagionalità e componente erratica.

Risultati della ricostruzione, pubblicati in modo sempre più dettagliato a partire dall'uscita del primo dato in nuova indagine (I trimestre 2004), sono le serie trimestrali dei principali aggregati del mercato del lavoro distinti per sesso, regione e classi quinquennali d'età, con dettagli molto approfonditi sulle variabili strettamente legate al lavoro: settore d'attività (a 11 modalità), posizione nella professione, carattere dell'occupazione e tipologia di orario per gli occupati, e durata della disoccupazione per le persone in cerca di lavoro.

In questo quadro si inserisce il presente lavoro: il dettaglio informativo già raggiunto viene ulteriormente approfondito per il titolo di studio. La metodologia utilizzata per la disaggregazione è però diversa da quella originale: l'approccio scelto per questo lavoro può essere definito per strutture in quanto sfrutta le distribuzioni demo-socio-economiche estratte dalla RFL; la scelta di cambiare tecnica è dovuta all'alto numero di serie da disaggregare.

L'esercizio sembra estremamente attuale visto il crescente interesse per dati sul mercato del lavoro con alto livello di disaggregazione, necessari per la definizione

di indicatori sempre più dettagliati; in particolare l'analisi per titolo di studio delle serie di occupati, disoccupati e non forze lavoro consente di interpretare l'evoluzione della partecipazione al mercato del lavoro in relazione alla crescente propensione alla formazione scolastica riscontrata negli ultimi anni.

Tale esercizio va inoltre a colmare un vuoto informativo che nasce in parte dall'impossibilità di confrontare dati storici di fonte forze lavoro rilevati in modo diverso nel tempo e in parte dalla cadenza decennale dei dati censuari che necessariamente non consentono di cogliere con continuità i mutamenti nel livello d'istruzione della popolazione.

La rilevanza di un'analisi di lungo periodo della struttura per titolo di studio della popolazione emerge dai dati riportati nella tabella 1: si osserva come in trenta anni il livello d'istruzione si sia modificato profondamente. La tabella mostra affinità e differenze tra dati rilevati dalla RFL e il censimento della popolazione riferiti agli stessi periodi<sup>1</sup>

**Tabella 1** – *Popolazione di 25-64 anni per titolo di studio (frequenze relative).*

RFL	1981	1991	2001	2011
Elementari	59,2	37,9	21,7	10,5
Licenza media	23,1	33,2	35,2	33,4
Senza diploma	<b>82,3</b>	<b>71,1</b>	<b>56,9</b>	<b>43,9</b>
Scuola secondaria	13,4	22,7	33,4	40,8
Laurea o più	4,2	6,2	9,8	15,3
Con diploma	<b>17,7</b>	<b>28,9</b>	<b>43,1</b>	<b>56,1</b>
<b>Totale</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>
<b>CENSIMENTO</b>	<b>1981</b>	<b>1991</b>	<b>2001</b>	<b>2011</b>
Elementari	61,5	41,1	22,7	11,3
Licenza media	20,8	30,6	34,3	33,5
Senza diploma	<b>82,4</b>	<b>71,6</b>	<b>57,0</b>	<b>44,9</b>
Scuola secondaria	13,0	21,8	32,1	38,7
Laurea o più	4,6	6,6	11,0	16,5
Con diploma	<b>17,6</b>	<b>28,4</b>	<b>43,0</b>	<b>55,1</b>
<b>Totale</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>

*Fonte Rilevazione sulle forze lavoro e censimento della popolazione.*

Si osserva che distinguendo la popolazione tra individui con e senza diploma i risultati delle due fonti sono piuttosto simili mentre, rispetto alla disaggregazione a quattro modalità, i titoli centrali sono più frequenti per forze lavoro mentre quelli

<sup>1</sup> Il censimento fa sempre riferimento a un preciso giorno della prima metà di ottobre; per la rilevazione sulle forze lavoro si è considerato il dato riferito al IV trimestre che nel 1981, 1991 e 2001 considerava come settimana di riferimento la prima di ottobre e nel 2011 tutte le settimane del trimestre che comunque non è caratterizzato da cambi di titolo di studio se non per i livelli più alti.

estremi per il censimento. Tutto ciò è dovuto probabilmente alla tendenza delle indagini campionarie a rilevare determinate tipologie di individui a causa dell'autoselezione dei rispondenti. Si può ipotizzare per una rilevazione campionaria la diffidenza a partecipare da parte di persone con livello d'istruzione particolarmente basso e la difficoltà nel raggiungere coloro che hanno titoli più alti, che tendenzialmente risultano più mobili sul territorio.

In questo quadro la classificazione del titolo di studio ha richiesto un'attenta analisi della qualità dei microdati a disposizione, dei cambiamenti avvenuti nella rilevazione e nel sistema scolastico.

I cambiamenti nella rilevazione sono iniziati con il passaggio, nel I trimestre 2003, dalla vecchia RFL, in cui la variabile titolo di studio assumeva 8 modalità, alla nuova indagine con 10 modalità oltre alla possibilità di non risposta, fino ad arrivare dopo tre trimestri alla distinzione in 12 titoli di studio e l'eliminazione della possibilità di non rispondere e alle 15 modalità nella rilevazione corrente. Ai cambiamenti nella rilevazione si è aggiunta la riforma universitaria (DM 509/99), operativa dall'A.A. 2001/02, con il passaggio dalle lauree di 4 o 5 anni (vecchio ordinamento) a quelle di primo livello (3 anni) e la possibilità di conseguire successivamente una laurea specialistica (2 anni).

Considerando questi fattori si è scelto di disaggregare tutte le serie già ricostruite rispetto alla variabile titolo di studio utilizzando due classificazioni diverse, ma confrontabili, a seconda del periodo.

In particolare tra il I trimestre 1977 e il III 1992 abbiamo considerato la disaggregazione a quattro modalità illustrata nella tabella 1, tra il IV trimestre 1992 e il IV 2003 si considerano sei modalità con la distinzione di coloro che hanno conseguito un titolo di scuola secondaria di 2-3 anni e 4-5 anni e dei laureati tra diploma di laurea o laurea triennale e laurea specialistica o di vecchio ordinamento.

### **3. Metodologia**

La metodologia utilizzata per questa ricostruzione differisce profondamente da quella adottata nella definizione delle altre variabili ricostruite; ciò è dovuto a delle modifiche intervenute sulla variabile titolo di studio nel periodo di passaggio dalla vecchia alla nuova rilevazione (2003) che non consentono l'utilizzo delle tecniche per componenti.

Nelle fasi di progettazione di questa ricostruzione si è deciso di far ricorso a tecniche macro in quanto:

1. Le informazioni a disposizione consentono l'utilizzo di ricostruzioni tramite strutture senza fare ricorso a imputazioni sui singoli record.

2. Per coerenza con le tecniche di ricostruzione su dati di fonte forze di lavoro adottate in passato.

La tecnica utilizzata è una generalizzazione dalla metodologia statistica per la stima indiretta di piccole aree mediante l'analisi di zone più vaste che le contengono, e più esattamente dal metodo SPREE<sup>2</sup> (J.N.K. Rao 2000). Questo approccio utilizza informazioni che sono in relazione con la variabile da stimare e che dividono la popolazione nelle cosiddette *crossclasses*. La scelta di tali variabili gioca naturalmente un ruolo fondamentale nella procedura di stima. Il punto focale della procedura è la determinazione di due relazioni riferite ai dati a disposizione:

1 l'*association structure*, vale a dire la relazione tra la variabile da stimare e quella (o quelle) a essa associata, con riferimento alla singola piccola area;

2 l'*allocation structure*, vale a dire la relazione tra la variabile da stimare e quella (o quelle) a essa associata con riferimento alla grande area, composta dall'unione delle piccole (Purcell 1980)

Nel nostro caso si tratta di una generalizzazione in quanto invece di lavorare su aree geografiche più piccole alle quali vengono associate le strutture di aree geografiche più grandi che le contengono, consideriamo serie disaggregate rispetto a più variabili (n) alle quali vengono associate strutture di serie disaggregate per meno variabili (n-k) che risultano particolarmente esplicative rispetto al complesso delle variabili.

Nella pratica le serie ricostruite sono prese come vincolo per ottenere un risultato perfettamente coerente con i dati diffusi dall'Istat. Sulle serie vincolo vengono proiettate le strutture demo-socio-economiche derivanti dai dati di indagine. Il metodo è applicato in modo differente a seconda dei periodi: per i trimestri dal IV 1992 al IV 2003 vengono estratte le strutture per 6 titoli di studio dalla vecchia RFL e applicate agli aggregati ricostruiti. Le strutture per titolo di studio vengono costruite come segue<sup>3</sup>:

$$P_{\text{anno,trim,cond,reg,sex,eta,set,pos,dur,tit}} = \frac{Pop_{\text{anno,trim,cond,reg,sex,eta,set,pos,dur,tit}}}{Pop_{\text{anno,trim,cond,reg,sex,eta,set,pos,dur}}} \quad (1)$$

Dove pop indica la popolazione per ogni anno, trim=trimestre, cond=condizione occupazionale, reg=regione, sex=sesso, set=settores d'attività economica, pos=posizione professionale, eta=classe d'età quinquennale, dur=durata della disoccupazione e tit=titolo di studio.

<sup>2</sup> Structure preserving estimation

<sup>3</sup> La formula 1 è riferita alla disaggregazione di variabili del periodo IV trimestre 1992-IV trimestre 2003; per il periodo I trimestre 1977-III trimestre 1992 le variabili che contribuiscono al calcolo delle strutture sono meno e con modalità diverse poiché è minore il numero di serie ricostruite.

Per costruzione risulta:

$$\sum_{tit=1}^W P_{anno,trim,cond,reg,sex,eta,set,pos,dur,tit} = 1 \quad (2)$$

Con  $W$  pari al numero di modalità della variabile titolo di studio che è pari a 4 per il periodo I trimestre 1977-III trimestre 1992 e a 6 per il periodo successivo.

Tali strutture, come già accennato, non vengono calcolate su tutte le  $n$  variabili ricostruite ma solo sulle  $n-k$  che risultano più correlate con il titolo di studio per evitare distorsione. La scelta delle variabili da includere, e il numero di modalità in cui suddividere ogni variabile, è un'operazione piuttosto delicata poiché aggiungendo variabili non sempre aumenta il livello di qualità delle strutture che si ottengono.

Per esempio per il periodo IV trimestre 1992-IV trimestre 2003 è risultato più performante non includere le variabili carattere dell'occupazione e tipologia di orario di lavoro e ridurre le modalità del settore d'attività economica da 11 a 4.

Per ottenere le stime ricostruite, le strutture vengono moltiplicate per tutte le serie caratterizzate dalle specifiche modalità delle  $n-k$  variabili considerate per la loro definizione<sup>4</sup>:

$$Pop_{anno,\dots,dur,car,piepar,tit}^* = Pop_{anno,\dots,dur,car,piepar} * P_{anno,\dots,dur,tit} \quad (3)$$

Il risultato ( $Pop^*$ ) è l'abbinamento e disaggregazione di ogni serie ricostruita, rispetto a tutte le  $k$  variabili, in 4/6 sotto-aggregati per livello di istruzione massimo raggiunto. Per un totale di: 962 serie tra il I trimestre 1977 e il III trimestre 1992 e 67550 serie dal IV 1992 al IV 2003

#### 4. Principali risultati

La base dati prodotta consente di fare analisi di lungo periodo in quanto abbiamo a disposizione serie ricostruite con diversi livelli di disaggregazione a partire dal 1977 o dal 1992. Per il periodo 1977-1992, come già detto, abbiamo tre classi d'età e quattro livelli d'istruzione, dal 1992 al 2003 classi di età quinquennali e titolo di studio a sei modalità.

L'analisi prevede uno studio della distribuzione della popolazione e dell'andamento dei tassi di attività, occupazione e disoccupazione dei dati dal 1977 e solo per alcune sottopopolazione dal 1992.

<sup>4</sup> La popolazione ( $Pop$ ) ha in più, rispetto al peso ( $P$ ), la disaggregazione per le variabili  $car$  (carattere dell'occupazione) e  $piepar$  (tipologia di orario di lavoro) oltre a 11 modalità della variabile settore d'attività economica.

Vista la lunghezza delle serie ci siamo concentrati principalmente sugli anni 1977, 1987, 1997, 2007 e 2016 e su individui appartenenti alla classe d'età 25-64, in quanto hanno avuto la possibilità di concludere la loro carriera di studi.

Sfruttando il maggiore dettaglio disponibile dal 1992, ci siamo concentrati sull'analisi più dettagliata per età di coloro che hanno raggiunto un titolo di studio superiore al diploma. Anche in questo caso pur avendo a disposizione le medie dei dati dal 1993 al 2016, l'analisi è stata riferita solo ad alcuni anni (1993, 2000, 2008, 2016).

Come noto negli ultimi anni si è registrato un progressivo innalzamento dei livelli di istruzione della popolazione italiana (OECD 2016).

Questo fenomeno è ben evidente dai dati in tabella 2; per la popolazione compresa tra i 25 e 64 anni, si è registrato un dimezzamento della quota di coloro che hanno raggiunto al massimo la licenza media passando dall'86,4% del 1977 al 40,3% del 2016. Al contrario, si evidenzia un aumento di coloro che detengono un titolo di studio superiore al diploma di oltre 46 punti percentuali, si passa infatti dal 13,6 % nel 1977 al quasi 60% nel 2016.

**Tabella 2** – *Popolazione residente in Italia di 25-64 anni per titolo di studio (valori percentuali). Anni 1977, 1987, 1997, 2007, 2016.*

Titolo di studio	1977	1987	1997	2007	2016
Elementari	68,4	48,1	27,2	14,8	6,7
Licenza media	18,1	28,3	35,3	33,5	33,5
Senza diploma	<b>86,4</b>	<b>76,4</b>	<b>62,5</b>	<b>48,3</b>	<b>40,3</b>
Scuola secondaria	10,2	18,4	29,2	38,2	42,0
Laurea o più	3,4	5,2	8,2	13,5	17,7
Con diploma	<b>13,6</b>	<b>23,6</b>	<b>37,5</b>	<b>51,7</b>	<b>59,7</b>
Totale	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b>

Fonte Rilevazione sulle forze lavoro.

Tutto ciò risulta ancora più evidente se ci concentriamo sulla classe d'età compresa tra i 25 e i 34 anni.

A seguito del processo di riforma dei sistemi di istruzione superiore dell'Unione Europea che ha avuto inizio nel 1997 e che si pone l'obiettivo di integrare i sistemi universitari dei paesi aderenti entro il 2010, è possibile notare un forte incremento della quota di popolazione che raggiunge un livello d'istruzione superiore al diploma<sup>5</sup>. In particolare tra il 1993 e il 2016 la percentuale di 25-34enni che hanno la laurea risulta quasi quintuplicata passando dal 5,4% al 24,8%.

<sup>5</sup> Per l'Italia tale riforma è sancita dal Processo di Bologna e dello Spazio Europeo dell'Istruzione Superiore del 1999, dal DM 509/99, ed è operativa dall'A.A. 2001/02

A questo costante aumento del livello di scolarizzazione della popolazione italiana è associato un mondo lavorativo sempre più competitivo e profondamente caratterizzato per genere.

Dalla tabella 3 emerge come la partecipazione maschile raggiunga tassi molto più alti rispetto a quella femminile ma anche quanto questa differenza tenda a diminuire sia nel tempo sia all'interno dello stesso anno al crescere del livello del titolo di studio.

Il primo fenomeno è legato dall'aumento della partecipazione al mondo del lavoro delle donne: guardando ai totali la differenza tra i due sessi passa da più di 51 punti percentuali nel 1977 a oltre 22 punti percentuali nel 2016.

Il secondo fenomeno è dovuto alla forte discrepanza tra i tassi di attività femminile per i diversi titoli di studio, il che mette in evidenza quanto il livello d'istruzione sia più discriminante tra le donne. Concentrando l'attenzione sui livelli estremi di istruzione, per entrambi i sessi, il tasso di attività di coloro che non raggiungono la licenza media diminuisce molto nel tempo, passando dall'84,1 % al 58,3 % per i maschi e dal 30,0% al 20,3% per le femmine.

**Tabella 3** – Tassi di attività dei 25-64enni per sesso, titolo di studio (valori percentuali). Anni 1977,1987,1997,2007,2016.

Titolo di studio	Maschi					Femmine				
	1977	1987	1997	2007	2016	1977	1987	1997	2007	2016
Elementari	84,1	74,9	59,7	55,1	58,3	30,0	28,7	22,9	18,6	20,3
Licenza media	93,3	91,9	85,7	82,1	79,1	39,9	48,4	47,7	47,7	47,6
Scuola secondaria	89,0	90,2	87,3	86,8	87,3	62,0	68,5	67,6	68,3	67,7
Laurea o più	93,9	94,0	91,5	89,2	90,3	85,9	87,0	82,6	79,1	81,8
<b>Totali</b>	<b>87,0</b>	<b>84,3</b>	<b>80,5</b>	<b>81,5</b>	<b>83,1</b>	<b>35,7</b>	<b>43,2</b>	<b>48,5</b>	<b>55,1</b>	<b>60,8</b>

Fonte Rilevazione sulle forze lavoro.

Al contrario, per titoli di studio superiori al diploma, nel 1977, i tassi di attività sono più elevati che nel 2016, ciò lascia supporre che nel 1977, l'aver conseguito un titolo di studio universitario rappresentasse di per se un elemento a favore di una maggiore probabilità di collocazione nel mercato del lavoro. Nel 2016, possedere un titolo di studio universitario si accompagna comunque a una maggiore partecipazione al mercato del lavoro, ma in termini comparativi il vantaggio si è ridotto. Passando all'analisi dell'occupazione, tabella 4, possiamo notare il progressivo aumento della differenza tra il tasso di occupazione di coloro che hanno un titolo di studio inferiore alla licenza media e quello dei laureati, nel 1977 tale differenza ammontava a 33 punti percentuali e nel 2016 supera i 50 punti percentuali. Risulta di particolare interesse l'aumento nel tempo del tasso di occupazione complessivo a fronte di una progressiva diminuzione dei tassi per tutti i titoli di studio. Questo fenomeno è attribuibile all'effetto composizione

dell'occupazione che come detto in precedenza vede una sempre più forte partecipazione femminile e di persone con titoli di studio sempre più elevati.

**Tabella 4** – Tassi d'occupazione 25-64enni titolo di studio (valori percentuali). Anni 1977,1987,1997,2007,2016.

Titolo di studio	1977	1987	1997	2007	2016
Elementari	53,1	46,6	34,0	31,2	29,2
Licenza media	67,0	67,6	61,7	62,3	55,7
Scuola secondaria	74,0	75,4	72,1	74,4	70,7
Laurea o più	86,1	87,4	81,6	80,2	79,8
<b>Totali</b>	<b>58,8</b>	<b>60,0</b>	<b>58,8</b>	<b>64,8</b>	<b>64,5</b>

Fonte Rilevazione sulle forze lavoro.

Approfondendo l'analisi rispetto al territorio, nella tabella 5, possiamo notare un forte divario tra i livelli del tasso di occupazione delle regioni del nord e quelle del sud e che quest'ultimo diventa sempre più evidente col passare del tempo. Nel 1977 infatti la differenza era solo di circa 4 punti percentuali nel 2016 arriva fino a 23 punti percentuali.

**Tabella 5** – Tassi d'occupazione 25-64enni titolo di studio e ripartizione geografica (valori percentuali). Anni 1977,1987,1997,2007,2016.

Titolo di studio	1977	1987	1997	2007	2016
		Nord			
Elementari	54,5	49,2	36,2	33,8	35,0
Licenza media	69,1	72,0	68,2	69,8	65,6
Scuola secondaria	75,8	79,2	78,1	80,8	78,2
Laurea o più	85,3	87,7	83,8	84,1	84,8
<b>Totali</b>	<b>60,5</b>	<b>63,2</b>	<b>64,0</b>	<b>71,8</b>	<b>73,4</b>
		Centro			
Elementari	52,7	48,0	36,7	34,6	35,2
Licenza media	66,5	67,5	63,2	66,3	60,9
Scuola secondaria	72,9	75,8	71,2	75,1	72,7
Laurea o più	85,3	87,1	80,6	80,7	81,8
<b>Totali</b>	<b>59,0</b>	<b>61,9</b>	<b>60,9</b>	<b>68,3</b>	<b>69,6</b>
		Mezzogiorno			
Elementari	51,3	42,7	30,1	27,7	24,5
Licenza media	63,2	60,7	52,2	51,1	42,2
Scuola secondaria	72,0	69,8	63,4	63,3	57,3
Laurea o più	87,8	87,2	79,3	73,8	69,8
<b>Totali</b>	<b>56,3</b>	<b>54,5</b>	<b>50,6</b>	<b>53,3</b>	<b>49,8</b>

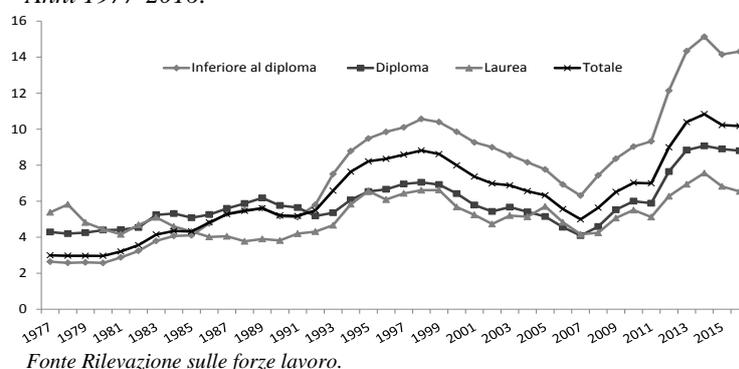
Fonte Rilevazione sulle forze lavoro.

Si osserva anche che i tassi di occupazione delle regioni settentrionali e centrali di coloro che hanno un titolo di studio superiore al diploma, rimangono piuttosto stabili nel tempo a differenza di quelli delle regioni meridionali che passano

dall'87,8 % nel 1977 al 69,8% nel 2016: il 2006 e il 2016, nel periodo di crisi economica, il titolo di studio ha rappresentato una garanzia per l'occupazione nel centro nord molto più che nel mezzogiorno.

Spostando l'attenzione sulla disoccupazione, nella figura 1 sono riportati i dati riguardanti le serie dei tassi di disoccupazione dal 1977 al 2016 per la classe di età 25-64 anni. Da questi dati risulta evidente come il tasso di disoccupazione per coloro che hanno un titolo di studio inferiore al diploma sia aumentato notevolmente passando dal 2,6% al 14,3%.

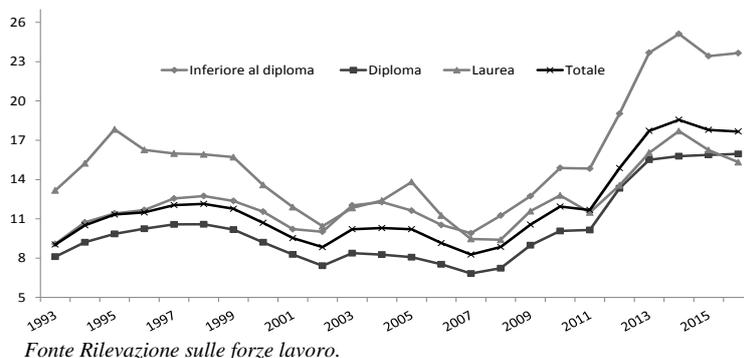
**Figura 1** – Tassi di disoccupazione dei 25-64enni per titolo di studio (valori percentuali). Anni 1977-2016.



Più articolato risulta invece l'andamento per coloro che hanno il diploma o una laurea. I tassi di disoccupazione di queste due categorie infatti rimangono sostanzialmente vicini per l'intero periodo discostandosi leggermente solo negli ultimi 5 anni. Si nota inoltre che fino al 1985 i diplomati e fino al 1991 i laureati avevano livelli di tasso di disoccupazione superiore al totale per poi diventare costantemente inferiori a tale livello; al contrario coloro che non hanno raggiunto il diploma fanno segnare dal 1991 tassi di disoccupazione che sono sempre più alti del totale e progressivamente più distanti da questo.

Restringendo il periodo di riferimento al 1992-2016, possiamo analizzare in modo più approfondito questo fenomeno facendo riferimento alla classe d'età 25-34 anni. Dalla figura 2 si nota che il livello dei tassi di disoccupazione per coloro che hanno un diploma è più basso di quello di chi ha almeno una laurea e che questa tendenza si inverte solo negli ultimi due anni. Questo fenomeno, più intenso fino al 2011, è apparentemente in contrasto con ciò che ci si aspetterebbe ed è dovuto principalmente al fatto che, tra i 25 e i 34 anni, coloro che hanno solo il diploma hanno avuto più tempo per cercare lavoro, a ciò si aggiunge la maggiore aspettativa in ambito lavorativo dei laureati, che li porta ad essere più selettivi rispetto alle offerte di lavoro e ad aumentare di conseguenza il tempo di ricerca.

**Figura 2** – Tassi di disoccupazione dei 25-34enni per titolo di studio (valori percentuali). Anni 1992-2016.



## 5. Conclusioni

Questo lavoro ha illustrato l'ultimo contributo fatto dall'Istat in termini di ricostruzioni di serie storiche di fonte RFL con l'aggiunta della variabile titolo di studio a tutte le serie prodotte in passato.

Rispetto ai precedenti lavori di ricostruzione di serie storiche è stata utilizzata una tecnica differente in quanto non considera l'impatto delle diverse componenti della serie ma le strutture di alcune variabili rispetto alla nuova variabile da ricostruire. La decisione di cambiare metodo per la ricostruzione era obbligata a causa di alcune modifiche della variabile titolo di studio intercorse negli anni. Seppur nuova per forze lavoro quella utilizzata è comunque una tecnica consolidata in Istat. Il risultato di tali elaborazioni sono serie di dati lunghe 40 anni interamente confrontabili. Le analisi delle serie prodotte mettono in evidenza quanto il livello di istruzione sia sempre più legato al lavoro; in particolare la partecipazione è sempre più riservata a coloro che hanno un livello di istruzione più elevato e questo è stato negli ultimi anni una garanzia rispetto alla crisi che ha colpito i mercati.

Questo lavoro si inserisce in un processo di aggiornamento e integrazione delle serie ricostruite e di dati storici sul mercato del lavoro diffusi dall'Istat. E' prevista infatti la pubblicazione delle principali serie ricostruite per titolo di studio dal 1977 all'interno del datawarehouse I.stat, e il rilascio di microdati storici dal 1977 contenenti i dati originariamente rilevati dall'indagine.

## Riferimenti bibliografici

ISTAT 2006. La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione. Istat, *Metodi e Norme*.

- ISTAT 2013. Occupati e disoccupati: dati ricostruiti dal 1977, *Statistiche report*
- OECD 2016. Education at a Glance 2016:OECD Indicators.
- PURCEL N.J., KISH L. 1980. Postcensal estimates for local Areas (or domains), *International statistical review*, 48, 3-18.
- RAO J.N.K. 2000. Statistical methodology for in direct estimations in small areas, 39, Eostat.

## SUMMARY

### Population and educational level: new data for long period analysis

The interest in longer time series, referring to different themes, is increasing because they allow to study the transformation of society. Over time the availability of the main aggregates by educational level is very important for the population and especially for the labour market analysis. The changes occurred over time in the Labour Force Survey have caused breaks in the time series, making incomparable data referring to periods in which techniques and definitions were different. To allow middle and long-term analysis, ISTAT has been working on the back-recalculation of time series always longer and more disaggregated to remove the breaks caused by changes in the Survey.

In this work the final step of the process of time series back-recalculation about labour market is shown. Exploiting the time series already back recalculated by main socio-demographic variables and thanks to the reconstruction techniques already developed, it was possible to reach a higher level of disaggregation of the series already built in the last years regarding the variable educational level. In this way we are allowed to know the characteristics of the main aggregates of labour market with regard to the higher educational level attained.

This work shows the methodology used to reconstruct the time series focused on the choices made to ensure an information consistent with the characteristics of the educational level. The main results have also been analyzed highlighting how much the distribution of the population by educational level is changed and how the level of education has become a discriminating factor related to the participation in the labour market.

---

Silvia LORIGA, ISTAT, siloriga@istat.it  
Andrea SPIZZICHINO, ISTAT, spizzich@istat.it  
Oriol GISBERT MARTI, Università di Pisa, oriol.gisbert@gmail.com  
Ilaria FRANCO, Università LA SAPIENZA, illyf93@gmail.com

## **A FINITE SAMPLE APPRAISAL OF THE DISTRIBUTION DYNAMICS APPROACH**

Margherita Gerolimetto, Stefano Magrini

### **1. Introduction**

In recent years the distribution dynamics (DD) approach to the analysis of convergence (Quah, 1996a, Quah, 1996b and Quah, 1997) has become more commonly adopted. Despite this more widespread use, to our knowledge no attempt has been made so far to evaluate the ability of the DD approach to capture, in finite samples, the features of the true dynamics underlying the evolution of per capita income disparities over time. To fill this gap, in this work we propose a Monte Carlo experiment whose aim is to show the finite sample performance of the DD approach as well as the actual improvement when a mean bias adjusted estimate of the conditional density (Hyndman et al., 1996) is employed (Fischer and Stumpner, 2008; Basile, 2010; Magrini et al., 2013; Gerolimetto and Magrini, 2014).

In contrast to the more traditional regression approach (Barro and Sala-i-Martin, 1995) that focuses on the transition dynamics of a representative economy towards its steady-state growth path, the distinctive feature of the DD approach is to examine directly the evolution of the entire cross-sectional distribution of per capita income (Durlauf and Quah, 1999; Magrini, 2004). In practice, this implies that convergence dynamics are analysed by visually inspecting the plot of the stochastic kernel estimate or comparing the features of the ergodic distribution to those of the distribution at the beginning of the considered period. In particular, when the ergodic distribution displays a tendency towards a point mass, this is interpreted as convergence towards equality while the formation of two (o more) separate modes signals a tendency toward polarization (stratification). Clearly, ascertaining the features of the convergence process in such a way is not feasible within a Monte Carlo experiment; consequently, we simulate alternative distribution dynamics that produce unambiguous features on the theoretical ergodic distribution and then test for the presence of these features in the estimated ergodic distributions.

The paper is structured as follows. In the second section, the DD approach is recalled. In the third section the Monte Carlo experiment is presented. The fourth section concludes..

## 2. Distribution Dynamics

Let the random variables  $X$  and  $Y$  represent per capita income (relative to group average) across a set of  $N$  economies at time  $t$  and  $t+s$  respectively and  $F(X)$  and  $F(Y)$  represent the corresponding distributions. Assuming that the distributions admit a density, denoted respectively with  $f(X)$  and  $f(Y)$ , and that their dynamics can be modelled as a first order process, the density at time  $t+s$  is:

$$f(Y) = \int_{-\infty}^{\infty} f(Y|X)f(X)dX \quad (1)$$

in which  $f(Y|X)$  is the stochastic kernel, a conditional density function that maps the density at time  $t$  into the density at time  $t+s$ . The stochastic kernel is the key element of the DD approach: convergence is analysed directly from the shape of a plot of the stochastic kernel estimate or, assuming that the process behind (1) follows a time homogenous markov process, by comparing the shape of the initial distribution to the ergodic distribution, i.e. the limit of  $f(Y)$  as  $s \rightarrow \infty$ .

A common way to obtain an estimate of the stochastic kernel in equation (1) is through the kernel density estimator. However, Hyndman et al. (1996) argue that this estimator might have poor bias properties and, among the possible solutions, suggest to adjust the estimate of the mean function (i.e., the mean of the conditional density function) implicit in the kernel density estimator with one obtained from a smoother with better bias properties. In the present analysis, therefore, we apply this mean-bias adjustment and, in particular, employ the local linear estimator (Loader, 1999) to obtain a superior estimate of the mean function.

## 3. Monte Carlo Experiment

The aim of this simulation experiment is to assess whether the distribution dynamics approach is capable of recognizing the long-run tendencies underneath the evolution of a cross-sectional distribution. To do this, we simulate pairs of vectors representing per capita income (relative to group average) of a group of  $N$  economies at time  $t$  and  $t+s$  as follows: the vector at time  $t$ ,  $X$ , is drawn from a Gaussian distribution,  $N(1,0.252)$ ; the vector at time  $t+s$ ,  $Y$ , is obtained according to

$$Y = M(X) + \varepsilon \quad (2)$$

where  $M(X)$  represents a specific mean function  $M(X)$  and  $\varepsilon$  is an error term. In particular, the error term is generated as a vector of normally distributed random variables,  $N \sim (1, \sigma^2)$ , where  $\sigma$  is set to obtain three alternative levels for the pseudo- $R^2$  (0.75, 0.8, 0.9); as for  $M(X)$ , we consider two alternative functions, represented in Fig. 1, which characterise the underlying process as one of (a) club-convergence coupled with a variance increase, or (b) absolute convergence. In the experiment, the sample size  $N$  takes on the values 50, 100 and 200 and the number of Monte Carlo replications is 1000. As for the bandwidths in the stochastic kernel estimate, we consider two opposite set-ups: in the first, we use a fixed bandwidth based on the cross-validation selection criterion both for the conditional density estimator and for the mean function estimator within the mean-bias adjustment procedure; in the second set-up, we use 70% nearest neighbours bandwidths for both estimators.

To ascertain whether the DD approach recognizes the correct mean function, and consequently reaches an accurate representation of the ergodic distribution thus leading to correct inference on the simulated features of the convergence process, we resort to a battery of tests. Specifically, we use Kolmogorov-Smirnov (KS), Epps-Singleton (ES; Epps-Singleton, 1986), and Azzalini-Bowman (AzB; Azzalini and Bowman, 1997) tests for the null hypothesis that the initial distribution is equal to the ergodic distribution. Moreover, to complement the output of the previous tests, we employ the Ansari-Bradley (AnB; Ansari and Bradley, 1960) two-sample (both two-tail and one-tail) test, for a difference in scale parameters. This, in particular, allows concluding whether the variance of the ergodic distribution is smaller (in case of convergence; AnB-C) or larger (in case of divergence; AnB-D) compared to that of the initial distribution. Finally, to check if the DD approach is able to recognize club-convergence dynamics, we use the unimodality test by Silverman (S; Silverman, 1986). To carry out all these tests, data from the ergodic distribution are needed. Clearly, in empirical applications these data are not available. A way out is to generate a sample from the estimated ergodic distribution as suggested in Silverman (1986) which, in turn, is based on the principles of the smoothed bootstrap proposed by Efron (1981).

Tables 1 and 2 report, for the two bandwidths set-ups, rejection percentages of the considered tests. In general, since we focus on (a) club-convergence coupled with a variance increase and (b) absolute convergence, we expect the DD approach to estimate an ergodic distribution that, in the majority of the Monte Carlo replications, is significantly different from the initial both globally (KS, ES and AzB tests) and in its second moment (AnB-C test for process  $a$ , convergence, and AnB-D test for process  $b$ , divergence or club-convergence). In addition, in relation

to process  $b$  only, we expect the estimated ergodic distribution to be bimodal (S test).

Rejection counts appear substantially high even for small  $N$  and, as expected, increase both in  $N$  and pseudo- $R^2$ . This means that the DD procedure is in the majority of the cases capable of recognizing the feature of the distribution that characterize a specific type of dynamic. In addition, we observe a performance improvement when a variable bandwidth is employed; this improvement is particularly evident for  $N=50$ , thus suggesting the employment of this set-up for very small sample sizes.

**Table 1 – Fixed bandwidth: percentages of rejections**

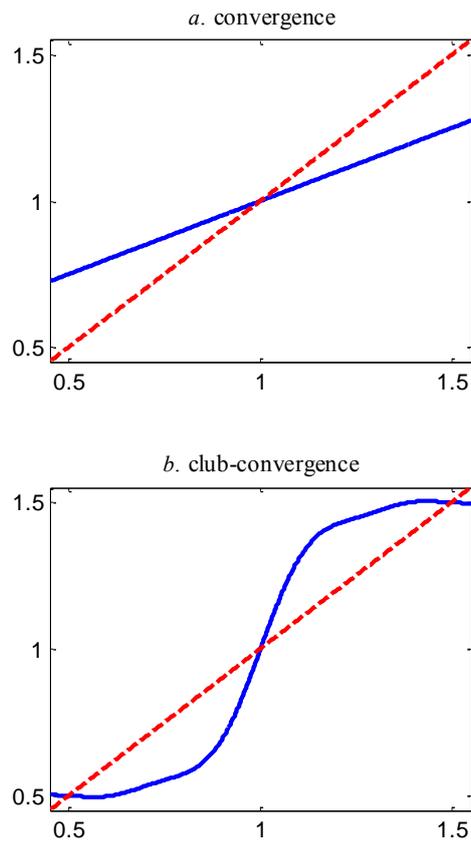
N	test	pseudo- $R^2 = 0.75$		pseudo- $R^2 = 0.8$		pseudo- $R^2 = 0.9$	
		$a$	$b$	$a$	$b$	$a$	$b$
50	KS	0.69	0.62	0.76	0.71	0.91	0.86
	ES	0.88	0.98	0.93	1	0.98	1
	AzB	0.85	0.95	0.92	0.98	0.98	1
	AnB	0.61	0.96	0.6	0.98	0.56	1
	AnB-C	0.98	-	1	-	1	-
	AnB-D	-	0.65	-	0.6	-	0.48
	S	-	0.39	-	0.57	-	0.81
100	KS	0.95	0.96	0.98	0.99	1	1
	ES	1	1	1	4	1	1
	AzB	1	1	1	1	1	1
	AnB	0.88	1	0.83	1	1	1
	AnB-C	1	-	1	-	0.01	-
	AnB-D	-	0.88	-	0.82	-	0.66
	S	-	0.56	-	0.74	-	0.91
200	KS	1	1	1	1	1	1
	ES	1	1	1	1	1	1
	AzB	1	1	1	1	1	1
	AnB	0.95	1	0.92	1	0.81	1
	AnB-C	1	-	1	-	1	-
	AnB-D	-	0.94	-	0.90	-	0.76
	S	-	0.64	-	0.81	-	0.96

Percentages of rejections of the null hypotheses of the various tests. The sign “-“ has been adopted when the test has not been implemented.

**Table 2** – Variable bandwidth: percentages of rejections

N	test	pseudo- $R^2 = 0.75$		pseudo- $R^2 = 0.8$		pseudo- $R^2 = 0.9$	
		<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>
50	KS	0.77	0.62	0.86	0.7	0.96	0.85
	ES	0.94	0.98	0.97	1	1	1
	AzB	0.93	0.95	0.96	0.98	0.99	1
	AnB	0.64	0.96	0.61	0.98	0.57	1
	AnB-C	0.97	-	0.99	-	1	-
	AnB-D	-	0.94	-	0.56	-	1
	S	-	0.55	-	0.69	-	0.85
100	KS	0.99	0.96	1	0.99	1	1
	ES	1	1	1	1	1	1
	AzB	1	1	1	1	1	1
	AnB	0.87	1	0.80	1	0.70	1
	AnB-C	1	-	1	-	1	-
	AnB-D	-	0.86	-	0.75	-	0.54
	S	-	0.74	-	0.88	-	0.92
200	KS	1	1	1	1	1	1
	ES	1	1	1	1	1	1
	AzB	1	1	1	1	1	1
	AnB	0.94	1	0.89	1	0.80	1
	AnB-C	1	-	1	-	1	-
	AnB-D	-	0.93	-	0.85	-	0.62
	S	-	0.89	-	0.96	-	0.96

Percentages of rejections of the null hypotheses of the various tests. The sign “-“ has been adopted when the test has not been implemented.

**Figure 1** – Simulated mean functions

Notes: Process (a) is:  
 $y = 0.5 + 0.5x$

Process (b) is:

$$y = 1 - 0.12 \cos(wx) - 0.59 \sin(wx) - 0.01 \cos(2wx) - 0.03 \sin(2wx) - 0.08 \cos(3wx) - 0.12 \sin(3wx) \\ - 0.02 \cos(4wx) - 0.02 \sin(4wx) - 0.04 \cos(5wx) - 0.03 \sin(5wx) - 0.01 \cos(6wx) - 0.003 \sin(6wx) \\ - 0.02 \cos(7wx) - 0.003 \sin(7wx) \\ \text{where } w=2.941$$

#### 4. Conclusions

In this paper we present a Monte Carlo experiment to show the capability of the DD approach to convergence analysis to capture the distinctive features of the dynamics underlying the evolution of per capita income disparities over time. The results of the Monte Carlo experiment disclose the ability of the DD approach to reveal, in finite samples, the main characteristics of such dynamics. This has been assessed by simulating *ad hoc* dynamics and providing, via several testing procedures, the number of times the specific feature is recognized. In particular, and even for very small sample sizes, the approach appears able to correctly identify the tendency of the underlying process both in global terms and with respect to the evolution of the second moment of the distribution; in addition, it also appears to be capable of recognizing the possible emergence of a second mode, that is a very interesting scenario, given its important economic meaning of club convergence.

#### References

- ANSARI A.R., BRADLEY R.A. 1960. Rank-sum tests for dispersions, *The Annals of Mathematical Statistics*, Vol. 31, pp. 1174–89.
- AZZALINI A., BOWMAN A.W. 1997. *Applied smoothing techniques for data analysis*, Oxford: Clarendon Press
- BARRO R.J., SALA-I-MARTIN X. 1995. *Economic growth*, New York: McGraw-Hill
- BASILE R. 2010. Intra-distribution dynamics of regional per-capita income in Europe: evidence from alternative conditional density estimators, *Statistica*, Vol. 70, pp 3–22.
- DURLAUF S.N., QUAH D. 1999. The new empirics of economic growth, in TAYLOR J.B. and WOODFORD M. (Eds), *Handbook of Macroeconomics*, Vol. 1A, Amsterdam: North Holland, pp. 231–304.
- EFRON B. 1981. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods, *Biometrika*, Vol. 68, pp. 589–99.
- EPPS T.W., SINGLETON K.J. 1986. An omnibus test for the two-sample problem using the empirical characteristic function, *Journal of Statistical Computation and Simulation*, Vol. 26, pp. 177–203.
- FISCHER M.M., STUMPNER P. 2008. Income distribution dynamics and cross-region convergence in Europe. Spatial filtering and novel stochastic kernel representations, *Journal of Geographical Systems*, Vol. 10, pp. 109–40.

- GEROLIMETTO M., MAGRINI, S. 2014. A novel look at long-run convergence dynamics in the United States, *International Regional Science Review*, Vol. 40, pp. 241-269.
- HYNDMAN R.J. BASHTANNYK D.M., GRUNWALD G.K., 1996. Estimating and visualizing conditional densities. *Journal of Computational and Graphical Statistics*, Vol. 5, pp. 315–36.
- LOADER C. 1999. Local regression and likelihood. Berlin: Springer
- MAGRINI S. 2004. Regional (di)convergence, in HENDERSON J.V. and THISSE J.-F. (Eds), *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: North Holland, pp. 2741–96.
- MAGRINI S., GEROLIMETTO M., DURAN H.E. 2013. Regional convergence and aggregate business cycle in the United States, *Regional Studies*, Vol. 49, pp. 251-272
- QUAH D.T. 1996a. Empirics for economic growth and convergence. *European Economic Review*, Vol. 40, pp. 1353–75.
- QUAH D.T., 1996b. Convergence empirics across economies with (some) capital mobility, *Journal of Economic Growth*, Vol. 1, pp. 95–124.
- QUAH D.T. 1997. Empirics for growth and distribution: stratification, polarization, and convergence clubs, *Journal of Economic Growth*, Vol. 2, pp. 27–59.
- SILVERMAN B.W. 1986. Density estimation for statistics and data analysis. London: Chapman and Hall.

## SUMMARY

### **A finite sample appraisal of the distribution dynamics approach**

The distribution dynamics approach has recently become of rather widespread use when convergence is analysed. Via Monte Carlo simulations, in this work we examine the ability of this approach to reveal, in finite samples, the main features of the dynamic process underlying the evolution of a distribution. The results of the experiment disclose the ability of the DD approach to recognize, in finite samples, the main features of these dynamics with respect to the evolution of per capita income disparities. In particular, and even for very small sample sizes, the approach appears able to correctly identify the tendency of the underlying process both in global terms and with respect to the evolution of the second moment of the distribution; in addition, it also appears capable of recognizing the possible emergence of a second mode.

---

MARGHERITA GEROLIMETTO, Università Ca' Foscari Venezia, Dipartimento di Economia. E-mail: [margherita.gerolimetto@unive.it](mailto:margherita.gerolimetto@unive.it)

STEFANO MAGRINI, Università Ca' Foscari Venezia, Dipartimento di Economia. E-mail: [stefano.magrini@unive.it](mailto:stefano.magrini@unive.it)



## **L'ARMONIZZAZIONE DELLE INDAGINI SOCIALI NELLA NUOVA STRATEGIA DEL CENSIMENTO PERMANENTE<sup>1</sup>**

Simona Mastroluca, Simona Pace

### **1. Introduzione**

Il processo di modernizzazione delle indagini sociali avviato dall'Istituto Nazionale di Statistica investe molti aspetti tra cui l'integrazione di dati provenienti da fonti amministrative e dati di indagine. Anche a livello internazionale la proposta di costruzione di un quadro di riferimento comune per le indagini sociali ha come obiettivo l'incremento della comparabilità e della coerenza delle statistiche sociali europee e l'armonizzazione delle variabili comuni.

L'integrazione tematica delle indagini sociali costituisce uno degli elementi fondanti del Sistema Integrato Censimento e Indagini Sociali (SICIS) in corso di realizzazione, che ha lo scopo di coprire il gap tra il sistema dei registri e gli obiettivi di diffusione del censimento permanente.

A tal fine occorre ridisegnare congiuntamente il censimento permanente e le attuali rilevazioni in campo sociale. Tale strategia consentirà di migliorare il livello di qualità complessivo delle informazioni acquisite a fronte di una riduzione dei costi e di ottenere una maggiore sinergia dei processi, innalzando anche il grado di coerenza delle statistiche prodotte in ambito sociale e censuario.

### **2. Il nuovo censimento permanente: caratteristiche e potenzialità**

A sei anni di distanza dall'ultimo Censimento della popolazione e delle abitazioni è in fase di programmazione la prima indagine a regime del nuovo progetto censuario che non sarà più a cadenza decennale ma permanente, con potenzialità di diffusione annuale dei dati.

Le profonde trasformazioni della società, la ormai copiosa disponibilità di dati di fonte amministrativa, gli elevati costi non più sostenibili, la necessità di ridurre il carico statistico sui rispondenti e, contemporaneamente, la volontà di garantire una

---

<sup>1</sup> I paragrafi 1, 2 e 4 devono attribuirsi a S. Mastroluca, i paragrafi 3, 5 e 6 a S. Pace.

fornitura tempestiva e continua dei dati hanno imposto una profonda rivisitazione della rilevazione. A livello internazionale sono numerosi i Paesi che effettuano un censimento totalmente *register based*, altri sono in fase di transizione e quindi utilizzano indagini campionarie a supporto dei registri, altri ancora per il 2021 hanno in programma un censimento di tipo tradizionale. In Italia già nel 2011 è stato realizzato un *register supported census* e alcune informazioni sono state rilevate solo su un campione di famiglie.

Il ridisegno dei censimenti in Europa - e nel mondo - avviene parallelamente al processo di modernizzazione delle indagini sociali e nella stessa direzione si sta muovendo anche l'Istat. Infatti, in Italia i principali aggregati censuari saranno prodotti a partire dal Registro base degli individui, dal Registro del Lavoro e dal Registro dei Luoghi. I dati raccolti sul campo non rappresentano più l'unica fonte di dati ma costituiscono una fonte sussidiaria per assicurare il completamento informativo e migliorare la qualità dei registri.

Il Sistema Integrato del Censimento permanente della popolazione e delle Indagini Sociali (SICIS) è una infrastruttura statistica complessa che sfrutta in modo integrato l'informazione di fonte amministrativa e quella proveniente dalle rilevazioni ed ha come obiettivo quello di colmare il gap informativo tra le variabili che a regime saranno disponibili nel Sistema Integrato dei Registri (SIR) e quelle necessarie per la diffusione dei dati del censimento permanente e delle principali indagini sociali dell'Istituto<sup>2</sup>.

Il SICIS si basa su un campione, il Master Sample (MS), progettato per supportare il SIR garantendo, da un lato, la produzione di dati statistici supplementari e migliorando, dall'altro, il livello di copertura e qualità dei Registri.

Si tratta di un disegno campionario congiunto per il censimento e le indagini sociali articolato in due fasi, la prima propedeutica agli obiettivi del Censimento permanente, la seconda a supporto delle altre rilevazioni.

La prima fase di indagine prevista nel periodo che va dal 15 settembre al 15 dicembre di ogni anno (dal 2018) si compone di due diversi schemi di rilevazione, uno Areale (400 mila famiglie<sup>3</sup>, 3000 comuni fino al 2021) e uno da Lista (1 milione di famiglie<sup>4</sup>, 3000 comuni fino al 2021), ed è finalizzata alla stima dei tassi di sovra e sotto copertura del Registro di base della popolazione, al completamento informativo in relazione agli obiettivi conoscitivi del Censimento permanente e a reperire notizie sulle famiglie campionate (es. telefono, mail) e facilitare così il contatto nella seconda occasione di indagine.

---

<sup>2</sup> Falorsi S. et al "Il Sistema integrato del Censimento permanente della popolazione e delle indagini sociali (SICIS)"- Documento interno.

<sup>3</sup> A partire dal 2022 le quantità verranno dimezzate.

<sup>4</sup> A partire dal 2022 le quantità verranno dimezzate.

La seconda fase di indagine viene effettuata l'anno successivo (e quindi dal 2019) ed è volta a rilevare variabili specifiche armonizzate su differenti sotto-campioni di famiglie (ovvero su quelle selezionate per le indagini: Multiscopo delle famiglie: aspetti della vita quotidiana; Rilevazione sulle Forze di lavoro; Indagine sulle spese delle famiglie; Indagine sulle condizioni di vita) ed a confermare le variabili strutturali in comune già acquisite nell'intervista di prima fase. Questo schema garantisce un quadro di coerenza tra le statistiche censuarie annuali e quelle, dello stesso tipo, prodotte con le indagini sociali, sfruttando il legame sulle stesse unità tra le variabili osservate con la prima fase e le stesse variabili osservate nella seconda fase.

Non solo. Il Sistema Integrato del Censimento permanente della popolazione e delle Indagini Sociali consentirebbe, a regime, di effettuare il *pooling* delle informazioni raccolte nelle due fasi su unità di rilevazione diverse allo scopo di rendere più nitida la fotografia dei fenomeni rilevati in maniera armonizzata nella prima e nella seconda occasione di indagine.

### **3. La modernizzazione delle indagini sociali nel contesto europeo**

Il sistema statistico europeo deve far fronte in misura sempre maggiore alla domanda di informazioni statistiche da parte di utilizzatori pubblici e privati non solo a fini di analisi e ricerca ma anche per la definizione delle politiche in campo economico e sociale. Tali dati devono rispettare gli elevati standard di qualità delle statistiche ufficiali, in particolare in termini di solidità, tempestività, pertinenza, capacità di adattamento alle nuove esigenze degli utenti, nonché di comparabilità ed efficienza. Le statistiche sociali utilizzate nell'UE provengono da una molteplicità di fonti: censimenti della popolazione, dati amministrativi, dati provenienti dalle imprese e dati su individui e famiglie ricavati da indagini campionarie.

L'attuale sistema di produzione di statistiche europee su individui e famiglie basate su indagini campionarie si fonda su una serie di regolamenti settoriali distinti che specificano con esattezza le tematiche da trattare e i requisiti tecnici per la rilevazione dei dati (ad esempio le dimensioni del campione, i criteri di qualità e i termini di trasmissione). Per lo svolgimento delle indagini sociali europee esistono attualmente cinque diverse basi giuridiche che disciplinano, rispettivamente, l'indagine sulle forze di lavoro (IFL), le statistiche europee sul reddito e sulle condizioni di vita (EU-SILC), l'indagine sull'istruzione degli adulti (AES), l'indagine europea sulla salute (EHIS) e l'indagine sull'uso delle tecnologie dell'informazione e della comunicazione (TIC) nelle famiglie (ICT-HH). Due indagini europee sono invece condotte in base a un accordo informale: l'indagine

sui bilanci delle famiglie (HBS) e l'indagine europea armonizzata sull'uso del tempo (HETUS).

La comunicazione della Commissione al Parlamento europeo e al Consiglio sul metodo di produzione delle statistiche UE "A vision for the next decade" ha messo in evidenza il crescente uso di dati provenienti da più fonti e di metodi innovativi di raccolta dei dati, l'importanza sempre maggiore dell'armonizzazione dei concetti e dei metodi statistici in tutti i settori, rilevando inoltre la necessità di nuove normative statistiche estese ad ambiti più ampi. Il *Framework Regulation Integrated European Social Statistics* (IESS) attualmente in discussione si muove all'interno di questo contesto ponendosi l'obiettivo di ridisegnare gli scenari per la produzione dei dati delle indagini sociali in un'ottica di integrazione. Con l'IESS ci sarà una sola base giuridica per le indagini sociali europee che garantirà la comparabilità, l'adattabilità e la coerenza dei dati sul lungo periodo tenuto conto anche della necessità di ridurre il carico statistico sui rispondenti nonché i costi complessivi delle indagini.

La proposta, così come è stata concepita, promuove l'uso di nuove forme di rilevazione dei dati e di fonti di dati alternative, compresi i dati amministrativi e le stime ottenute con la modellizzazione e organizza le statistiche oggetto del regolamento in settori e tematiche. Con il presente regolamento si definisce un quadro di riferimento comune per le statistiche europee su individui e famiglie provenienti da indagini campionarie e si specificano i dati e le informazioni che devono essere rilevati e presentati dagli Stati membri nonché i requisiti fondamentali di qualità che i dati devono soddisfare. Si stabilisce inoltre che specifiche tecniche più particolareggiate siano definite in atti delegati e misure di esecuzione. Si prevede l'integrazione delle varie rilevazioni di dati, tra loro e con l'uso dei dati amministrativi, consolidando e semplificando nel contempo la legislazione vigente.

Con questo nuovo *framework* di riferimento si perverrà anche a una migliore definizione dei ruoli dei diversi attori: i *policy makers* (PE e Consiglio) definiranno quali sono i dati da produrre (e quando dovrebbero essere prodotti); ai tecnici competerà il "come" devono essere prodotti i dati.

Il processo di modernizzazione delle indagini sociali investe dunque una pluralità di aspetti tra cui la standardizzazione delle variabili che rappresenta un requisito di fondamentale importanza che va ben oltre la fase di analisi. L'utilizzo di variabili standardizzate porta, infatti, dei vantaggi in tutte le fasi del processo di produzione del dato statistico: dal disegno dell'indagine, alla raccolta dei dati, dall'elaborazione alla fase di stima e di diffusione dei risultati.

Il progetto avviato da Eurostat su questa tematica ha come obiettivo quello di facilitare l'integrazione e la modernizzazione delle statistiche sociali mediante la standardizzazione di alcune variabili. Se in alcuni casi, infatti, potrebbe essere

necessario misurare concetti simili ma non identici in diverse indagini, in altri le differenze esistenti sono ingiustificate e non fanno che ostacolare l'integrazione tra le indagini a livello nazionale e europeo.

Già in passato Eurostat aveva intrapreso iniziative in questa stessa direzione: la *task force* sulle *Core Social Variables* aveva l'obiettivo di incrementare le possibilità di analisi trasversali e di promuovere l'utilizzo di alcune variabili sociali *core* in tutte le indagini sociali. Nel 2007, la *task force* ha prodotto una relazione finale in cui si raccomandava l'inclusione e l'armonizzazione di 16 variabili in tutte le indagini sociali. Nel 2011 è stato poi approvato dai Direttori delle Statistiche Sociali (DSS) un aggiornamento di tali variabili. Tuttavia, il lavoro condotto ha dimostrato che l'approccio alla standardizzazione doveva essere rafforzato per portare a risultati tangibili e che si dovevano fornire specifiche di dettaglio per garantire un livello di comparabilità tecnicamente sufficiente. Se in passato, l'approccio era quello di limitare la standardizzazione alle variabili ritenute *core* per tutte le indagini, si è convenuto che, ai fini della modernizzazione, fosse necessario ampliare la standardizzazione a un insieme di variabili più ampio. Nel novembre del 2013 il DSS ha avviato una serie di attività finalizzate a "rivedere le definizioni, le classificazioni e le linee guida per l'implementazione di tutte le variabili utilizzate in più indagini al fine di proporre, dove opportuno, una metodologia unica"<sup>5</sup>.

Il progetto di standardizzazione delle variabili sociali ha come obiettivo quello di sviluppare descrizioni standard per le variabili presenti in almeno due indagini sociali europee di microdati riguardanti famiglie e individui. L'elenco delle variabili standard include 28 variabili, di cui 18 sono variabili *core* (cioè da implementate in tutte le indagini sociali) mentre le altre 10 sono variabili incluse in alcune, ma non tutte, indagini. I lavori sulla standardizzazione sono stati organizzati in due diverse fasi corrispondenti rispettivamente alle variabili con priorità 1 e priorità 2 (P1 e P2). Nell'ottobre 2016 il DSS ha approvato le descrizioni standard proposte delle variabili P1, attualmente si sta lavorando sulle variabili P2.

Una volta approvate dal DSS, le versioni finali delle variabili standardizzate verranno raccolte in un unico manuale metodologico. Inoltre, alcuni elementi delle descrizioni delle variabili standardizzate - in primo luogo, il concetto e le categorie della variabile - saranno incluse nel regolamento di attuazione del regolamento quadro IESS. Il campo di applicazione del progetto sulle variabili standardizzate è relativo alle indagini sociali attualmente sotto regolamento o un accordo informale<sup>6</sup>. Per ciascuna di queste variabili è stata messa a punto una descrizione

<sup>5</sup> EUROSTAT. 2013. Doc. Eurostat/F/13/DSS/02/ 3.2/E Issues paper for DSS Board and DSS discussion.

<sup>6</sup> Cfr paragrafo 3.1.

standard che include la definizione, le caratteristiche delle popolazioni statistiche e delle unità osservate, le categorie standard e le linee guida per l'implementazione di ciascuna variabile. Tali descrizioni sono riferite alle variabili trasmesse ad Eurostat da parte degli Stati membri come file di microdati o variabili ricostruite direttamente da Eurostat o derivate<sup>7</sup>.

Eurostat vorrebbe successivamente estendere il lavoro sulla standardizzazione a variabili comuni a più di una raccolta di micro-dati sociali, ma non incluse attualmente nell'elenco delle P1 e P2.

#### 4. La diffusione censuaria tra regolamenti europei e tradizione italiana

Il Censimento della popolazione e delle abitazioni del 2011 è stato il primo normato da un Regolamento Quadro dell'Unione Europea (il 763 del 2008) e da tre Regolamenti di attuazione, uno sulle classificazioni e le specifiche tecniche delle variabili *core* (ovvero obbligatorie per tutti i Paesi membri), uno sul programma di diffusione dei dati e sui metadati e uno sugli indicatori di qualità volti a garantire in primo luogo contenuti informativi armonizzati e tempestività nella diffusione dei dati definitivi, obiettivi falliti con la formula del *Gentlemen's Agreement* adottata per il *round* del 2001.

Per il Censimento del 2021 il Regolamento Quadro di riferimento (che in allegato riporta l'elenco dei *core topics*, - tabella 1) è lo stesso del 2011, mentre i tre Regolamenti di attuazione della Commissione sono stati aggiornati in nome della semplificazione, grazie ai lavori svolti dalla *Task Force Eurostat on the future Eu censuses of population and housing* a cui ha partecipato attivamente anche l'Italia.

Rispetto alla precedente versione, sono state razionalizzate alcune delle classificazioni contenute nell'*Implementing Regulation on population and housing censuses as regards the technical specifications of the topics and of their breakdowns*<sup>8</sup> alla luce dell'esperienza maturata in occasione dell'ultimo censimento svolto. *L'Implementing Regulation establishing the reference year and the programme of the statistical data and metadata for population and housing censuses*<sup>9</sup> è stato profondamente rivisitato.

Il numero di ipercubi da rendere disponibili entro 27 mesi dalla data di riferimento del Censimento, articolati su diversi livelli di dettaglio tematico e territoriale, è stato sensibilmente ridotto (da 175 a 119 – tabella 2) e, laddove

---

<sup>7</sup> Le categorie standard non riguardano quindi la fase di raccolta dati o i dati diffusi sul sito Eurostat.

<sup>8</sup> Regolamento (CE) n. 2017/543 della Commissione del 22 Marzo 2017.

<sup>9</sup> Regolamento (CE) n. 2017/712 della Commissione del 20 Aprile 2017.

possibile, per alcuni di essi è diminuito il grado di complessità in termini di numero di variabili e/o di classificazioni di riferimento.

**Tabella 1 - Censimento della popolazione e delle abitazioni. Core Topics UE. Regolamento Quadro 763/2008.**

Etichetta	Variabile	Etichetta	Variabile	Etichetta	Variabile
AGE	Età	COC	Cittadinanza	TOH	Tipo riscaldamento
SEX	Sesso	POB	Luogo di nascita	TOB	Abitazioni per tipo di edificio
LMS	Stato civile	TSH	Titolo di godimento dell'alloggio	POC	Abitazioni per epoca di costruzione dell'edificio
HST	Posizione in famiglia	TLQ	Tipo di alloggio	TSH	Titolo di godimento dell'alloggio
FST	Posizione nel nucleo	OCS	Stato di occupazione	YAE	Anno di arrivo in Italia dal 1980
HAR	Numero residenti per tipo di alloggio	OWS	Alloggi di proprietà, in affitto, altro	YAT	Anno di arrivo in Italia dal 2010
LOC	Località	NOC	Numero di occupanti	ROY	Dimora abituale un anno prima del CP
CAS	Condizione professionale	UFS	Superficie	TFN	Tipo di nucleo
OCC	Professione	NOR	Numero di stanze		
EDU	Grado di istruzione	DFS/DRM	Metri quadri per occupante		
IND	Settore di attività economica	WSS	Disponibilità acqua		
SIE	Posizione nella professione	TOI	Disponibilità gabinetti		
LPW	Luogo di lavoro	BAT	Disponibilità vasca/doccia		

**Tabella 2 - Ipercubi di diffusione previsti dai Regolamenti di attuazione della Commissione sul programma di diffusione dei dati e dei metadati per il Censimento del 2011 e per quello del 2021.**

Territorio	2011				2021			
	Popolazione	Fam. e Nuclei	Alloggi	Totale	Popolazione	Fam. e Nuclei	Alloggi	Totale
NUTS1 (ripartizione)	9			9	11			11
NUTS2 (regione)	129	1	6	136	67		5	72
NUTS3 (provincia)	20	2	3	25	20	2	3	25
LAU2 (comune)	1	2	2	5	5	4	2	11
Totale	159	5	11	175	103	6	10	119

Ai 119 incroci obbligatori si aggiungono 13 nuovi indicatori (non previsti nel 2011) che dovranno essere forniti per griglia di territorio con celle da 1 chilometro quadrato. Si tratta di: Popolazione totale, Maschi, Femmine, Minori di 15 anni, 15-

64 anni, con più di 65 anni, occupati, Luogo di nascita nel Paese, Luogo di nascita in un altro Paese UE, Luogo di nascita fuori dall'UE, Dimora abituale un anno prima invariata, Dimora abituale un anno prima nel Paese, Dimora abituale un anno prima fuori dal Paese.

Nel corso del 2016 è stata fatta una lettura critica del piano di diffusione nazionale dello scorso Censimento propedeutica alla progettazione del piano per il 2021. In particolare, sono state realizzate alcune analisi volte a individuare i dati del Censimento del 2011 che sono stati maggiormente utilizzati o consultando direttamente le banche dati messe a disposizione sul sito dell'Istituto oppure avanzando richieste di elaborazioni personalizzate. E' stata perfezionata, inoltre, una vasta ricognizione della diffusione censuaria di alcuni paesi esteri, sia dal punto di vista dei contenuti che degli strumenti, utile per valutare l'introduzione di nuove proposte informative o di nuovi sistemi di diffusione che consentano una sempre maggiore fruibilità dell'output censuario da parte degli *stakeholders* italiani e stranieri.

Grazie alla copiosa disponibilità di informazione di fonte amministrativa e alle rilevazioni MS previste a partire dal 2018, alle tradizionali diffusioni decennali (europea e nazionale), verranno affiancate occasioni di diffusione annuale per rispondere ai continui fabbisogni informativi dei comuni e in generale degli utilizzatori del dato censuario per finalità di programmazione o ricerca. Non si tratta di un'offerta fissa nel tempo. La flessibilità della nuova strategia e l'arricchimento pressoché continuo dei Registri statistici consente di modulare anno per anno i contenuti del censimento e quindi gli incroci di dati consultabili.

## **5. Integrazione tematica delle indagini sociali: le prime analisi**

L'integrazione tematica delle indagini sociali costituisce uno degli elementi fondamentali della strategia che è alla base del Sistema integrato Censimento e indagini sociali – SICIS.

Per poter avviare le attività di integrazione è stata condotta un'analisi preliminare delle differenze attualmente esistenti in termini di definizioni, quesiti e popolazioni di interesse tra le variabili strutturali richieste per la produzione censuaria e presenti nelle indagini sociali. Sono state messe a confronto le variabili socio-demografiche del Censimento (Indagine Sperimentale D+ 2017) con le corrispondenti variabili osservate nelle principali indagini sociali (IS), ovvero quelle della seconda occasione del MS<sup>10</sup>).

---

<sup>10</sup> Cfr paragrafo 2.

**Tabella 3 -** *Quesiti dell'Indagine indagine sperimentale del Censimento D+, presenza nelle principali Indagini Sociali e corrispondenza con le variabili standardizzate proposte da Eurostat*

QUESITI D+ (*)	PRESENZA NELLE IS				Priorità Eurostat
	AVQ	RFL	SPESE	EU-SILC	
1.1 Relazione di parentela	x	x	x	x	
1.2 Sesso	x	x	x	x	P1_C
1.3 Data di nascita	x	x	x	x	P1_C
1.4 Luogo di nascita	x	x	x	x	P2_C
1.5 Stato civile	x	x	x	x	
1.6 Anno del matrimonio/unione civile	x	x	x	x	
1.7 Stato civile prec. ultimo matr./unione	x	x	x	x	
2.1 Cittadinanza	x	x	x	x	P2_C
2.2 Cittadinanza italiana dalla nascita	x	x	x	x	
2.3 Ottenimento cittadinanza italiana	x				
2.4 Luogo di nascita madre		x			P2_C
2.5 Luogo di nascita padre		x			P2_C
4.1 Il bambino frequenta	x			x	
4.2 Titolo di studio più elevato	x	x	x	x	P2_C
4.3 Titolo di studio conseguito all'estero					
4.4 Iscrizione attuale a un corso regolare di studi	x	x	x	x	P1
4.5 Tipo di corso a cui si è iscritti	x	x	x	x	P1
4.6 Frequenza corso di formazione professionale nella sett. di rif.		x		x	
5.1 Svolgimento di un'ora di lavoro nella sett. di rif.	x	x	x	x	
5.2 Eventuale assenza nella sett. di rif.	x	x	x	x	
5.3 Ricerca attiva nelle 4 sett. di rif.	x	x	x	x	
5.4 Disponibilità ad iniziare un lavoro entro due sett.	x	x	x	x	
5.5 Svolgimento nel passato di un'attività lavorativa	x	x	x	x	
5.6 Tipo di lavoro svolto	x	x	x	x	P2_C
5.7 Dipendenti retribuiti	x	x		x	P2_C
5.8 Professione svolta	x	x	x	x	P2_C
5.9 Settore di attività economica	x	x	x	x	P2_C
5.10 Cond. autopercepita sett. di rif.	x	x	x	x	P1_C
6.2 Luogo di studio/lavoro	x	x			
6.6 Orario di uscita da casa per recarsi al luogo di studio/lavoro	x				
6.7 Tempo impiegato per recarsi al luogo di studio/lavoro	x	x			
6.8 Mezzo di trasporto utilizzato per recarsi al luogo di studio/lavoro	x				
1.4 Numero dei componenti famiglia	x	x	x	x	P1_C
1.5 Titolo con cui la famiglia occupa l'alloggio	x		x	x	P1

Il confronto tra Censimento e indagini sociali è stato condotto rispetto a una serie di criteri, e cioè:

- **popolazione**, intesa come il target di riferimento della variabile, cioè l'insieme delle unità di rilevazione presso cui viene raccolta una determinata informazione;
- **concetto**, vale a dire il fenomeno rilevato dalla variabile;
- **riferimento temporale**, cioè il tempo cui si riferisce il concetto indagato;
- **quesito**, cioè la formulazione della domanda (*wording*) utilizzata per raccogliere una determinata informazione;
- **modalità**, ovvero tutti i valori assunti dalla variabile;
- **ordine della modalità**, cioè l'esatta collocazione delle modalità nel quesito.

Per ciascun criterio sono poi state definite delle modalità di classificazione (uguale, diversa armonizzabile, diversa non armonizzabile) e individuate delle regole univoche per l'attribuzione di un giudizio complessivo sull'armonizzabilità (armonizzabile, armonizzabile con accorgimenti, non armonizzabile).

L'analisi svolta sulle variabili ha evidenziato come attualmente esistano parecchie differenze tra le IS e la parte censuaria, meno tra le sole IS. In molti casi la non armonizzabilità di una variabile deriva dal fatto che vengono osservate popolazioni diverse; in altri casi i concetti rilevati sono "simili" ma non esattamente gli stessi; in altri casi, invece, la non armonizzabilità è determinata da modalità di risposta non confrontabili; infine in alcune situazioni si è rilevata la presenza di più cause che determinano la non "armonizzabilità". Le analisi fin qui condotte hanno costituito una prima riflessione sul tema dell'integrazione tematica delle IS e hanno offerto alcune indicazioni per definire una proposta di questionario armonizzato che tenga in considerazione le esigenze informative del Censimento e delle altre Indagini Sociali che entrano nella seconda occasione di indagine del MS.

Le variabili oggetto della proposta sono le variabili socio demografiche presenti nel questionario del censimento permanente e in almeno una delle quattro IS considerate (che corrispondono a 34 quesiti, di cui 16 fanno riferimento a variabili standardizzate di Eurostat - tab.3). Tale proposta dovrà però tenere conto anche delle definizioni e delle istruzioni contenute nei materiali a supporto delle indagini, nei quali si rilevano spesso differenze che non emergono dall'analisi delle classificazioni e delle modalità utilizzate. Successivamente la riflessione sull'armonizzazione prenderà in considerazione le altre 12 variabili standardizzate Eurostat, attualmente non presenti nel questionario del censimento permanente; alcune di queste potrebbero in futuro entrare, stabilmente o a rotazione, nella rilevazione.

## 6. Conclusioni

Le considerazioni avviate ad oggi suggeriscono che un'analisi completa dell'armonizzabilità non può prescindere da una serie di valutazioni dei fattori "esogeni" che potrebbero influire in modo determinante su eventuali cambiamenti nei quesiti. Si tratta di tutta una serie di vincoli ai quali ogni specifica indagine è soggetta: i regolamenti internazionali e nazionali ai quali la singola IS deve rispondere, le definizioni adottate, la struttura e i flussi dei questionari utilizzati, la diffusione programmata, l'organizzazione della rilevazione e i contratti in essere con le società esterne e così via. Alla luce di tutte queste considerazioni, l'armonizzazione potrebbe dunque rivelarsi più fattibile di quanto ipotizzato o, al contrario, più complessa se non addirittura non praticabile.

L'armonizzazione è dunque un obiettivo da costruire nel tempo e in modo integrato, rafforzando sin da ora i legami tra i diversi soggetti che partecipano alla costruzione del sistema delle statistiche sociali, i settori che producono le statistiche sociali, ma anche le strutture trasversali che contribuiscono in diversa misura a raccogliere e diffondere il patrimonio informativo in ambito sociale, sia in termini di progettazione e raccolta dati, sia in termini di definizione dei sistemi dei metadati di istituto, sia anche in termini di gestione e diffusione delle informazioni.

Per questo motivo sono stati avviati diversi tavoli di collaborazione all'interno dell'Istituto per fornire una riflessione congiunta su queste tematiche. L'attività di integrazione tematica così configurata sarà in grado di garantire il ridisegno complessivo delle indagini sociali in modo coerente a quanto proposto in sede europea, favorire l'ampliamento dal punto di vista tematico delle IS e armonizzare i processi statistici di produzione dei dati.

### Riferimenti bibliografici

EUROSTAT 2008. Regolamento (EC) N.763/2008 del Parlamento Europeo e del Consiglio sui censimenti della popolazione e delle abitazioni.

EUROSTAT 2013. Doc. Eurostat/F/13/DSS/02/ 3.2/E Issues paper for DSS Board and DSS discussion.

EUROSTAT 2016. DSS-2016-Oct-4.2 Standardisation of social variables - progress report.

EUROSTAT 2016. Proposta di regolamento del Parlamento europeo e del Consiglio che istituisce un quadro comune per le statistiche europee sulle persone e sulle famiglie, basate su dati a livello individuale ricavati da campioni, <https://ec.europa.eu/transparency/regdoc/rep/1/2016/IT/1-2016-551-IT-F1-1.PDF>

- EUROSTAT 2017. DSS-2017-Mar-4.2 Standardisation of social variables progress report v.2.
- EUROSTAT 2017. Regolamento di attuazione della Commissione (EU) N.2017/543 recante norme per l'applicazione del Regolamento (EC) N.763/2008 sulle specifiche tecniche e le classificazioni.
- EUROSTAT 2017. Regolamento di attuazione della Commissione (EU) N.2017/712 sull'anno di riferimento e il programma statistico dei dati e dei metadati di cui al Regolamento (EC) N.763/2008.
- EUROSTAT 2017. Regolamento di attuazione della Commissione (EU) N.2017/881 recante norme per l'applicazione del Regolamento (EC) N.763/2008 sui report di qualità e il formato tecnico di trasmissione dei dati.
- FALORSI S., et al. 2017. *Il Sistema integrato del Censimento permanente della popolazione e delle indagini sociali (SICIS)*, Roma, Documento interno.
- PACE S., et al. 2017. *L'integrazione delle indagini sociali e il censimento permanente: analogie e differenze nelle variabili strutturali*, Roma, Documento interno.

## SUMMARY

### **The harmonization of Social Surveys in the Permanent Census strategy**

The modernization of social surveys carried out by the Italian National Institute of Statistics involves many aspects among which the integration of data from administrative sources and survey data. Also at international level, the proposal to build a framework regulation for an integrated system of social surveys (European Parliament and Council Regulation) aims to increase the comparability and consistency of European social statistics and the harmonization of common variables. The thematic integration of social surveys is one of the founding elements of the Integrated Census and Social Surveys System (SICIS) which is underway. It is designed to cover the gap between the registry system and the dissemination of the permanent census. For this purpose it is necessary to redraw the current social surveys together with the permanent census. This strategy will allow to improve the overall quality of the information obtained while reducing costs and achieving greater synergy of processes, also raising the degree of consistency of social and census statistics.

---

Simona MASTROLUCA, Istituto nazionale di statistica, mastrolu@istat.it  
Simona PACE, Istat, Istituto nazionale di statistica, sipace@istat.it

## **TERRITORIAL WAGE INEQUALITY: THE EFFECT OF EMPLOYER, EMPLOYEE AND JOB CHARACTERISTICS<sup>1</sup>**

Ciro Baldi, Sara Gigante, Silvia Pacini

### **1. Introduction**

In Italy the wage inequality has a marked territorial characterization, with the regions of the South and Islands that record lower values with respect to the North and Center. The purpose of this paper is to contribute to this topic after a period of relative lack of interest of the economic literature with the notable exception of two recent contributions (Boeri, Ichino and Moretti, 2016; Pagani, 2017). It focuses on measuring the nominal wage inequality between and within areas on the basis of a new database available at Istat, the Register on Employees and Earnings (RACLI) and to assess the effect of observable variables.

### **2. Data**

This analyses is based on the new Register on Employees and Earnings (RACLI) produced by the Italian NSI with information at job level for the year 2014. It covers all enterprises with at least one employee and the relative work relationships in the private non agricultural sector (respectively about 1.5 and 15.8 millions). In the register, mainly based on social security data, there are detailed information on the relationships between the employee and the employer. Wages paid by enterprises and an estimation of paid hours are the main variables available together with classification variables related to the type of relationship such as the working time (full time/part time), type of the contract (temporary/permanent) and the contractual occupation<sup>2</sup>. Moreover, thanks to the link to the population and business registers and other administrative data available, other information are

---

<sup>1</sup> Istat bears no responsibility for the analysis and the conclusions which are solely those of the authors.

<sup>2</sup> More information on the RACLI register are available in a Statistical Report of the NSI (Istat, 2016), where can also be found a methodological note on the characteristics of the main variables. As the work on this database is still in progress, especially due to the building up of the new system of registers in Istat, the results should be considered provisional.

matched such as sex, age and level of education concerning the individual; nace activity code, geographical localization of local units and size class, concerning the enterprise where the job is active; the tenure (or seniority) regarding the job itself.

The analysis on the RACLI register may be based on different statistical units; in this work we choose to focus on is the work relationship between an employer and an employee. This relationship is characterized by an activation date and, if it is the case by a cessation date. This means that for the same employer and employee there can be more than one work relationships, if for instance, a temporary work contract is expired in the year and a new one is started in the same year<sup>3</sup>.

While most of the variables are derived from the social security data (Inps) that, in various forms, have been probably the main data source used in the literature to examine the wage inequality in Italy, this dataset has some interesting additional features like, the education level and the length of the work relationship, that are core variable among the wage determinants.

In this paper we work on a sample of about 1.3 million work relationships that is about the 8.2% of the total. The sample is drawn with a two stage methodology, first sampling the enterprises and secondly sampling the work relationships within the enterprises. The final sampling weights, are the product of those in the first and second stage.

The measure used through the paper is the hourly wage<sup>4</sup> (in logarithm), that is the wage rate of each hour paid in the work relationship. As we focus on the wage rate of each hour (Brandolini, Rosolia, 2016), the micro level wage rates are weighted by the number of hours paid in the work relationship. These employment weights thus are the product of the hours paid and the sampling weights.

### *2.1. Descriptive results*

Table 1 display some statistical measures to synthesize the distribution of log wage per hour in the main five NUTS1 Italian geographical areas: North-West (NW), North-Est (NE), Centre (CE), South (SO) and Islands (IS).

Looking at the mean, it is evident that it decreases from North-West to South and Islands with the Centre closer to the North than to the South. The difference

---

<sup>3</sup> The transformations of the contract (for instance from part-time to full-time or from temporary to permanent) without any cessation imply the continuation of the work relationships.

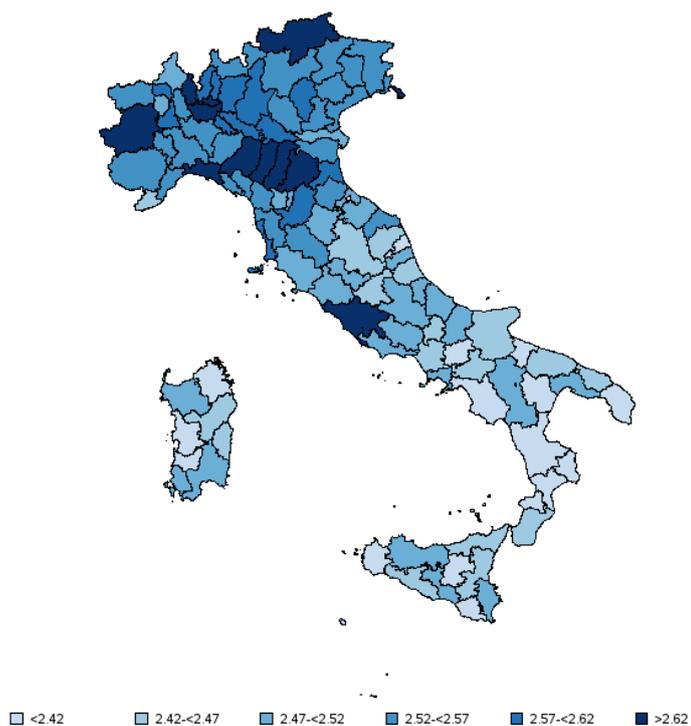
<sup>4</sup> The focus on hourly wage, among other advantages, allows to include seamlessly the part-time workers in the analysis. The number of hours paid is an estimate derived from a procedure that takes into account a number of variables of administrative and survey data. The use of different variables allows a correction, at least partial, of the possible errors of the administrative data.

between the North-West and the South is 19 log points. The pattern of the medians confirm what has been found with the means.

**Table 1** – Measures of location and variability for the areas of the country.

	Geographical areas					
	NW	NE	CE	SO	IS	IT
Mean	2,64	2,58	2,56	2,45	2,46	2,57
Median	2,56	2,51	2,46	2,39	2,39	2,49
Variance	0,19	0,14	0,18	0,12	0,12	0,17
p9010	1,01	0,87	0,98	0,79	0,79	0,94
p9050	0,65	0,56	0,66	0,5	0,52	0,61
p5010	0,36	0,31	0,32	0,29	0,27	0,33

**Figure 1** – Log hourly wage by provinces. Year 2014.



The polarization between areas of the country is even clearer when looking at a map by provinces (Figure 1) where the provinces of the North have darker colours,

denoting higher hourly wage rate, than South and Islands with the Centre somewhat in between the two.

**Table 2** – *Composition structure of the workforce, weighted by hours paid, in enterprises with at least one employee. Year 2014. Percentage values.*

	Geographical areas					
	NW	NE	CE	SO	IS	IT
	Sex					
Female	37.2	37.2	38.1	29.7	32.0	36.0
Male	62.8	62.8	61.9	70.3	68.0	64.0
	Age (years)					
15-29	14.9	16.1	14.3	15.8	15.8	15.3
30-49	60.5	59.8	60.1	59.4	59.3	60.0
50-high	24.6	24.1	25.6	24.8	24.9	24.7
	Educational level					
Primary	34.9	34.3	31.5	39.5	41.3	35.0
Secondary	46.1	49.3	47.8	45.7	45.0	47.1
Tertiary	15.5	12.9	15.8	11.8	10.8	14.1
NA*	3.6	3.5	5.0	3.0	3.0	3.7
	Tenure (years)					
0-2	27.4	28.3	31.4	40.2	39.1	30.9
3-8	30.4	30.3	30.6	28.2	30.1	30.1
9-high	42.2	41.4	38.0	31.6	30.8	39.0
	Occupation					
Apprentices	3.4	4.2	4.6	3.0	3.1	3.7
Managers	1.7	0.8	1.1	0.3	0.3	1.1
White collars	47.0	40.9	43.3	34.1	38.6	42.4
Blue collars	47.9	54.1	51.0	62.6	58.0	52.7
	Working Time					
Full-time	85.4	84.7	80.5	78.6	75.0	82.7
Part-time	14.6	15.3	19.5	21.4	25.0	17.3
	Type of Contract					
Temporary	9.2	11.7	10.5	12.6	12.7	10.8
Permanent	90.8	88.3	89.5	87.4	87.3	89.2
	Nace sector					
Industry	35.4	40.2	28.1	28.2	19.4	33.2
Construction	6.3	6.5	7.1	10.1	10.4	7.3
Services	58.3	53.3	64.8	61.7	70.2	59.5
	Size class of enterprise (n. of employees)					
0-4	12.7	13.3	16.4	21.9	24.7	15.6
5-14	16.4	17.6	18.4	19.5	20.5	17.8
15-49	16.1	18.3	15.6	16.3	14.3	16.5
50-249	20.6	20.8	16.4	14.7	14.1	18.6
250+	34.2	30.0	33.3	27.6	26.5	31.6

\*NA: Information not available in the register.

Turning the attention to the measures of dispersion, first of all Table 1 shows the variance of log wage per hour in the five NUTS1 areas. The dispersion of the

measure is maximum in North-West (0,19) and minimum in the South and Islands (0,12). Taken Italy as a whole as the reference the North-West reports a variance that is, in the scale used, almost 12% higher and the difference between North-West and Islands is of 58%.

Measuring the dispersion in terms of the difference between the 90 and 10 percentiles does not change the picture (Table 1). The North-West reports a 90-10 differential that is almost 30% higher than the South and Islands. It is worthwhile to note that the difference between regions is concentrated in the top half of the distribution. In fact while the difference between the 50 and 10 percentile is quite homogenous, hovering around 30 log points, the 90-50 differential is quite different with the one in North-West being about 1.3 times higher than the one in South and Islands.

The aggregate measures of wages (in terms both of location and dispersion) are likely to be affected by the composition of the employment in the different areas.

Table 2 shows the percentage composition, in terms of hours paid, of different characteristics of the workers, the jobs and the firms.

On each area, we can observe many characteristics at the employee's level, such as sex and level of education, related to the work relationship, such as type of the contract, working time, occupation and tenure, while at the firm's level we know NACE and size.

The share of female employment is 29.7% and 32% respectively in the South and Islands, that is at least 4 percentage points under the Italian share, with a distance between the South and the northern regions of 7-8 percentage points.

South and Islands are characterized by a higher incidence of employees with a primary education and a lower incidence of tertiary education and the tenure tends to decrease from North to South. In particular in the South and Islands the employees with a tenure of maximum two years are about 10 percentage point over the Italian average that is 30.9%.

Structural differences between the Northern and Southern areas are also evident on the work relationship characterizations: permanent and full-time contracts appear to be less frequent in South areas. The Northern areas present also the largest number of white collars.

Moreover the smaller enterprises, with a number of employees under fifteen, are more concentrated in the South and Islands: the share of work in enterprises with less than 15 employees is 50% more than respect to Northern regions. In the South and Islands the much higher percentage of employment (70% in the Islands versus the 59.5% of the Italian average) is occupied in the Services sectors.

### 3. Wage differentials between areas

Since this paragraph we make a step forward in the analysis trying to explain the differences in the wage inequality across areas. In order to do so we simplify the setting and divide the country in two macro areas: North-Centre (NC) and South-Island (SI). The aggregation of the Centre with the North is justified mainly by the fact that, as seen from Table 1, the Center is closer to the Northern areas both in terms of location and dispersion of hourly wage. We start analyzing what lies behind the average difference between areas, in terms of how much of the observed differences can be ascribed to the different composition of the employment. The tool used for this purpose is the classical decomposition due to Oaxaca (1973) and Blinder (1973) (hereafter OB decomposition)<sup>5</sup>.

The wage differential is thus decomposed in two components: a part due to differences in the characteristics of the areas, the explained component or composition effect, and a part due to the differences in the wage rates paid for each characteristic, the unexplained component or wage structure effect. The model takes into account demographic variables (gender, education in 4 classes and age in quadratic form), job related variables (tenure in quadratic form, occupations in 4 classes, type of working time, type of contract), enterprises variables (36 nace sectors and 5 size classes).

The total gap of the average log wage per hour in the North-Center compared to that of the South-Islands is about 0,15 log points. The composition effect explains almost 8 log points, that is about the 53% of the total gap.

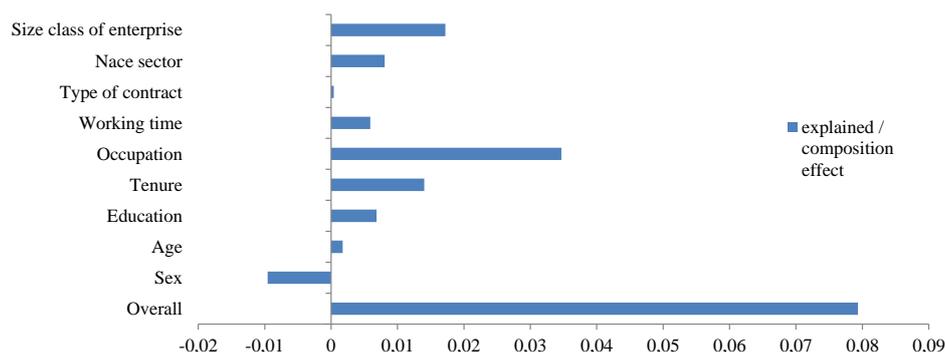
In the detailed decomposition of this effect the largest contribution are given by the occupation and size of enterprise, reflecting respectively the higher share of blue-collar and small enterprises in the South-Islands (Figure 2). A notable role is also played by tenure in the job which is due to the larger presence of short work relationships in the Southern regions. A smaller role is attributed to the educational differences, due probably to the relatively small share of people with higher education associated with a more compressed wage structure along this dimension. The different composition by gender in the two areas, with a larger share of men in

<sup>5</sup> It is based on the estimation of mincerian type wage equation separately for the two different geographical areas (A) that are North-Centre (NC) and South-Islands (SI):  $w_{iA} = \mathbf{x}\boldsymbol{\beta}_{iA} + \varepsilon_{iA}$  (1) where  $w$  is the (log) wage per hour,  $\mathbf{x}$  is a vector of characteristics that influence the wage and  $\varepsilon$  an error term for each work relationship (i). The observed average wage differential between the two areas, under the assumption of the North-Centre wage structure as reference and adding and subtracting the term  $\bar{\mathbf{x}}^{SI}\hat{\boldsymbol{\beta}}^{NC}$ , can then be written as:

$$\bar{w}^{NC} - \bar{w}^{SI} = \bar{\mathbf{x}}^{SI}(\hat{\boldsymbol{\beta}}^{NC} - \hat{\boldsymbol{\beta}}^{SI}) + (\bar{\mathbf{x}}^{NC} - \bar{\mathbf{x}}^{SI})\hat{\boldsymbol{\beta}}^{NC} \quad (2)$$

the South-Islands, lies behind the only detailed composition effect with negative sign.

**Figure 2** – Oaxaca-Blinder decomposition's results: composition effects.



#### 4. Wage dispersion within area

In this paragraph we explore the within area wage dispersion, in order to search for the variables that are mostly associated with it and look for hints of what can explain the difference between the areas.

In doing so we adopt the regression based methodology presented in Fiorio and Jenkins (2007).

The method provides an accounting of the contribution to inequality of the target variable of all the factors that a model uses to explain the measure of income or, in this case, of wages.

The regressions used in the application are the same applied in paragraph 3.

**Table 3** – Regression based decomposition of inequality.

Factor	Geographical area		
	NC	SI	IT
Residual	37.79	43.10	38.72
Sex	2.56	1.68	2.15
Age	5.98	3.79	5.35
Education	4.06	3.60	4.01
Tenure	5.55	8.88	6.32
Occupation	27.78	15.41	26.11
Working Time	1.94	1.39	1.91
Type of contract	0.25	0.02	0.20
Nace Sector	8.45	11.44	8.61
Size of enterprise	5.64	10.69	6.61
Total	100.00	100.00	100.00

First, the inequality decomposition for the South-Islands leaves a much higher share to the residual, despite as we have seen, the distribution for this area is much more compressed than that of the North-Center. The second, most striking difference, is the contribution of the occupation factor. Along this dimension the share to inequality is much higher in the North-Centre than in the South-Islands. As we have seen earlier, the distribution of employment in this second area is much more concentrated on blue collar. This one and the more compressed wage structure along this dimension, emerged also in the OB decomposition, are probably the causes of this feature of the data.

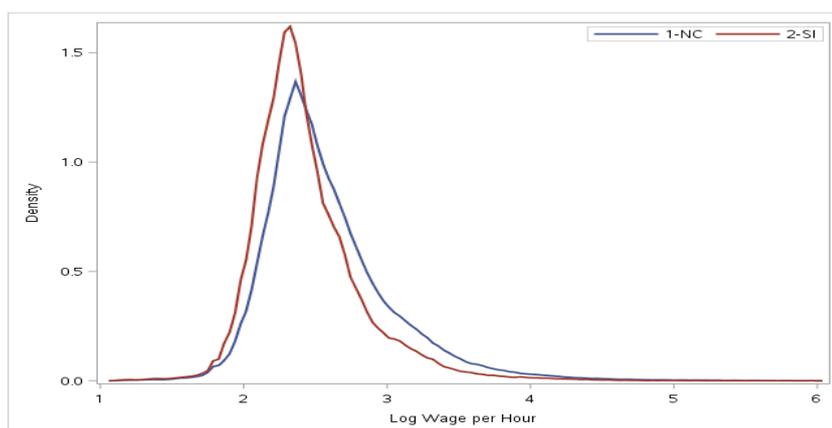
The demographics characteristics, sex, age and education, play a smaller role in the Southern area. It is probably due, in case of education, both to the smaller premium associated to the tertiary education and the more homogeneous distribution of employment by education groups (toward the lower groups).

## 5. Composition effects along the distribution

In paragraph 3 we have seen that composition effects can explain about 53% of the mean difference between North-Center and South-Islands. In this paragraph we try to see what is the contribution of observables along the entire distributions.

Figure 3 compares the density of log wages per hour in the two areas. Summarizing the results seen previously, it shows that the density of the South-Island is clearly shifted to the left and more compressed compared to the density of North-Center which have a much longer right tail.

**Figure 3** – Wage density by geographical area.



To analyze the impact on the wage distribution we adopt the propensity score reweighting procedure by Di Nardo Fortin and Lemieux (1996). The idea of the methodology is to build up a counterfactual wage density by re-weighting the sample of one of the two groups with a factor which accounts for the difference in the composition of employment groups with the other group. In this application we build a counterfactual distribution by reweighting the North Center wages. In this way the counterfactual represents the wage distribution that would have been observed in the North Center if the composition of the observables were similar to that of the South-Islands. More formally, if  $e_{igNC}$  is the employment weight of work relationship  $i$ , in a group defined by observable variables  $g$ , in the NC,  $E_{gNA}$  the sum of weights in group  $g$  in NC and  $E_{NC}$  the sum of the weights in NC, we can write the normalized weight  $\omega_{iNC}$  as the product of the weight of the single work relationship in the group  $g$  and the weight of the group in the area

$$\omega_{iNC} = \frac{e_{igNC} E_{gNC}}{E_{gNC} E_{NC}} \quad (1)$$

an analogous normalized weight can be defined for SI. The counterfactual weight is then defined as:

$$\omega_{iNC}^{CF} = \frac{e_{igNC} E_{gSI}}{E_{gNC} E_{SI}} \quad (2)$$

This would be the weight associated to  $i$  if group  $g$  contributed as much as in the South-Islands. To pass from (1) to (2) is just a matter of multiplying 1 by an appropriate reweighting factor,  $k_{ig}$ :

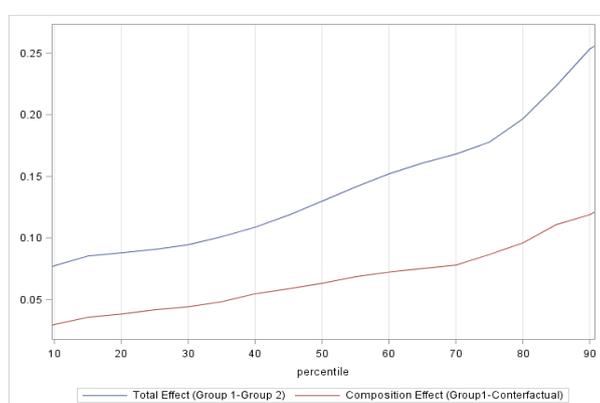
$$\omega_{iNC}^{CF} = \omega_{iNC} k_{ig} \quad (3)$$

$$k_{ig} = \frac{E_{gSI} / E_{SI}}{E_{gNC} / E_{NC}} \quad (4)$$

This reweighting factor can be calculated simply by taking the employment frequency of cells (obtained by crossing the different modalities of observables) in the two areas. However this method can incur in empty cells when the number and the breakdown of observables is large. Another way to calculating the reweighting factor is through a logistic regression with a flexible functional forms as suggested by Di Nardo (2002). In the application the specification includes all the variables included in the OB regression models (the only differences being that age and tenure are in classes, instead of being continuous). The specification contains main effects and two way interactions.

The figure 4 shows along with the differences in the hourly wage for the percentiles of the two distributions (total effect), the difference between the curve of the North-Center and the counterfactual curve, that is the effect due to the composition. The contribution of the composition effect is smaller at the bottom of the distribution, grows in the middle of the distribution and then decrease a little at the top. At the 10th percentile the composition effect explains about the 37.5% of the total effect, at the 40th about the 54.5% while at the top it contributes for the 48%.

**Figure 4** – *Difference between areas by quantile: total effect and composition effect.*



## 6. Conclusions

In this work the wage inequality in non agricultural private sector between and within Italian geographical area has been assessed with the use of a new database in construction at Istat on hourly wages. The paper reassess the differences found in the literature and try to evaluate the composition effect due to the observable characteristics of the employees, the jobs and the employers. With respect to the previous work based on administrative data, in this paper we control also for education and job tenure. It is found that the composition effect explains about half of the difference in mean wages, the rest being due to a wage structure effect. The different structure in terms of contractual occupation and size classes, and job tenure are the core determinants of the composition effect while the differences in the structure of employment in terms of education plays a smaller role due also to the compressed wage structure along this dimension. Along the entire wage

distribution, the composition effect is found to have a larger role in the middle quantiles.

### Ringraziamenti

The authors wish to thank all the group that in this years have worked in one way or another to the construction of the Register on employment and Earnings, and in particular to Stefano de Santis.

### Riferimenti bibliografici

- BLINDER, A.S. 1973. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*. Vol. 8, No. 4, pp. 436–455.
- JOHN DI ISTAT, 2016. Differenziali retributivi nel settore privato. Città: Casa Editrice.
- BOERI T., ICHINO A., MORETTI E. 2016. Divari territoriali e contrattazione: quando l'uguale diventa diseguale. Presentato al Festival della Economia di Trento, 2016.
- BRANDOLINI A., ROSOLIA A. 2016. The euro area wage distribution over the crisis. 34th IARIW General Conference. Dresden. JEL Classification: J31, D33
- FIORIO CARLO, JENKINS STEPHEN. 2007. Regression-based inequality decomposition, United Kingdom Stata Users' Group Meetings 2007, Stata Users Group, <http://EconPapers.repec.org/RePEc:boc:usug07:03>.
- NARDO, NICOLE M. FORTIN, THOMAS LEMIEUX. 1996. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. *Econometrica*, Vol. 64, No. 5, pp. 1001-1044.
- OAXACA R. 1973. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, Vol. 14, No. 3, pp. 693–709.
- PAGANI L. 2017. I differenziali salariali territoriali. *Salari, produttività, disuguaglianze. Verso un nuovo modello contrattuale?* Roma: Il Mulino

## SUMMARY

### **Territorial wage inequality: the effect of employer, employee and job characteristics**

The topic of wage inequality across geographical areas in Italy is investigated with the use of a new source of data, in construction at Istat, the Register of Employment and Earnings (RACLI), an employer employee database mainly based on social security data.

Some advantages of this database over previously available data is the presence of information on education of the worker and length of job relationship (tenure in the job). The analysis assesses the size of the wage dispersion between and within Nuts1 areas in 2014. It is found that the composition of the employment explains about half of the differences in average hourly wage (in log). The largest role in the composition effect is played by the contractual occupation (more concentrated in Mezzogiorno on blue collar workers) and size classes of enterprises (due to a larger prevalence, in this area, of small and very small firms in this area). A notable role is also played by tenure in the job which is due to the larger presence of short work relationship in the Southern regions. A smaller role is attributed to the educational differences, due probably to the relatively small share of people with higher education associated with a more compressed wage structure along this dimension. The composition effect is found to have an impact differentiated along the wage distribution, with an higher contribution on the differences between areas in the middle quantiles.

---

Ciro BALDI, Istat, baldi@istat.it

Sara GIGANTE, Istat, gigante@istat.it

Silvia PACINI, Istat, pacini@istat.it

## **IL CENSIMENTO PERMANENTE: PRIME ANALISI SUGLI OBIETTIVI DI DIFFUSIONE A LIVELLO COMUNALE**

Federico Brogi, Valeria Quondamstefano, Fabrizio Solari

### **1. Introduzione**

La necessità di far fronte alle richieste degli stakeholder interessati ad ottenere informazione statistica ad elevati livelli di qualità, tempestività, disaggregazione territoriale, e i vincoli stringenti in termini di costo e di non disturbo ai rispondenti, hanno spinto l'Istat a ridisegnare i processi produttivi del Censimento della popolazione in modo innovativo. In particolare, il prossimo censimento sarà realizzato sfruttando in modo integrato il complesso delle informazioni disponibili dai diversi registri tematici e dalle indagini campionarie.

Il ridisegno generale del prossimo censimento mira a unire le diverse esigenze per cui è stato progettato: il contenimento dei costi, il miglioramento del livello di qualità complessivo delle statistiche sociali ed, infine, l'ampliamento del dettaglio degli indicatori socio-demografici ed economici prodotti dall'Istituto.

In questo lavoro si vuole mettere in luce come la nuova strategia sia in grado di rispondere agli obiettivi di rilascio dei dati ad un livello di disaggregazione territoriale ampio. Relativamente ad alcuni ipercubi, definiti sulla base di una classificazione incrociata di variabili di natura demografica e variabili relative al mercato del lavoro, al titolo di studio e ai flussi di pendolarismo, sarà analizzato il livello di precisione delle stime a livello comunale.

### **2. Il Censimento permanente della popolazione e delle abitazioni**

Il Censimento permanente della popolazione e delle abitazioni adempie agli obblighi di rilevazione stabiliti dal Regolamento (CE) n. 763/2008 del Parlamento europeo e del Consiglio del 9 luglio 2008, relativo ai censimenti della popolazione e delle abitazioni.

Oltre a fornire indicazioni sulla consistenza numerica e sulle caratteristiche delle abitazioni e degli edifici, produce informazioni sulle principali caratteristiche strutturali, sulle condizioni sociali ed economiche della popolazione a livello nazionale, regionale e locale.

L'obiettivo primario del Censimento permanente è riuscire a mantenere il livello di dettaglio classificatorio garantito tradizionalmente dal censimento decennale - in termini di domini territoriali e di variabili strutturali - per un insieme di variabili core fondamentali (di natura demografica, sociale ed economica) aumentando la frequenza temporale dell'informazione, prodotta a cadenza prevalentemente annuale.

Gli obiettivi conoscitivi specifici del Censimento riguardano sia il conteggio della popolazione totale e delle sue principali componenti, sia la produzione degli ipercubi ottenuti dalla classificazione incrociata delle variabili target.

Ruolo fondamentale assume l'uso dei registri amministrativi. In particolare, il Censimento permanente sfrutterà per le variabili sostituibili il Registro Base degli Individui (RBI) e i registri tematici (lavoro, istruzione, luoghi, etc.), mentre per le altre variabili le informazioni prodotte dal Sistema Integrato del Censimento permanente e delle Indagini Sociali (SICIS).

In riferimento a simili esperienze internazionali relative alla definizione di un disegno di campionamento generale per le indagini sociali, Eurostat (2013) propone un approccio modulare per il disegno integrato delle indagini sociali (si veda, inoltre, Ionnadis et al., 2016).

Inoltre, l'Australian Bureau of Statistics (ABS, 2015) sta progettando un sistema integrato di indagini, chiamato Australian Population Survey, molto simile al SICIS. Occorre sottolineare che l' Australian Population Survey non è progettato per rimpiazzare il censimento.

Lo schema a due componenti in procinto di essere adottato in Istat è assimilabile al censimento basato sui registri studiato dall'Office for National Statistics (ONS) che avrà luogo a partire dal 2023 dopo la tornata censuaria del 2021 (ONS, 2017). In particolare, nel 2021 l'ONS condurrà un censimento tradizionale e, allo stesso tempo, sarà svolta un censimento parallelo basato sulla costruzione di un registro integrato della popolazione utilizzando dati amministrativi e due indagini con caratteristiche simili alle componenti L ed A della strategia italiana.

Dal 2004, l'Istituto di Statistica francese, INSEE, ha ridisegnato il censimento tradizionale, basandosi sul concetto di "rolling census", con lo scopo di produrre annualmente informazioni a livello locale e spalmare i costi censuari su più anni (Durr, 2005). Punto in comune con il Censimento permanente italiano è il sistema di rotazione dei comuni piccoli in un arco temporale pluriennale.

Un'ulteriore esperienza internazionale simile a quella italiana è il rolling census condotto in Israele. In tale caso viene usato un registro integrato che viene aggiustato mediante pesi calcolati per mezzo di un'indagine di copertura. (Pfeffermann, 2015).

### 3. Sistema Integrato del Censimento permanente e delle Indagini Sociali

Il Sistema Integrato del Censimento permanente e delle Indagini Sociali (SICIS) è pensato come una infrastruttura statistica complessa progettata per sfruttare in modo congiunto ed efficiente tutta l'informazione disponibile sui fenomeni di interesse, proveniente dai registri e dalle rilevazioni campionarie. A tal fine l'intero insieme delle indagini sociali condotte attualmente dall'Istat viene ripensato in un'ottica integrata di sistema che incorpora al suo interno, come l'elemento centrale e più importante, il processo del Censimento permanente.

Per descrivere il ruolo del SICIS nel supportare il RBI, è utile classificare le variabili di interesse come totalmente o parzialmente sostituibili o non sostituibili. Sono definite totalmente sostituibili le variabili disponibili per tutte le unità nei registri ad un livello di accuratezza e copertura elevati. Si definiscono parzialmente sostituibili le variabili che sono complete e accurate solo per un sottoinsieme della popolazione obiettivo. Infine, sono dette non sostituibili le variabili non desumibili dai registri. Possono essere stimate esclusivamente attraverso indagini campionarie sfruttando i dati provenienti dal RBI come informazione ausiliaria.

### 4. Master Sample

Il SICIS è basato sul Master Sample, ossia su un campione pianificato allo scopo di supportare il SIR aumentando la quantità di informazione statistica producibile e migliorando, allo stesso tempo, il livello di copertura e di qualità dei dati forniti.

Il Master Sample è strutturato in due differenti componenti campionarie, areale e da lista, denominate A e L. La componente A è basata su un campione di aree e/o di indirizzi, estratti da un archivio integrato. È disegnata per stimare i tassi di sotto-copertura e sovra-copertura del RBI a livello nazionale e sub-nazionale per diversi profili di sotto-popolazioni definite da variabili quali sesso, età e cittadinanza. La componente L è basata su una campione di famiglie estratte dal registro di base degli individui e delle famiglie. È disegnata con lo scopo di ottenere la stima degli aggregati che non possono essere ottenuti usando l'informazione proveniente dai registri.

La prima fase di indagine ha il duplice scopo di correggere per sotto e sovra-copertura i totali di popolazione forniti dal RBI e di colmare i gap informativi del SIR per quanto riguarda gli obiettivi conoscitivi del Censimento permanente. Inoltre, nella prima fase, vengono rilevate alcune informazioni essenziali a

supporto di tutto il sistema delle indagini sociali oggetto della seconda fase di indagine (nei mesi di ottobre e novembre di ciascun anno).

La prima edizione della prima fase del Master Sample (da lista ed areale) si svolgerà nei mesi di ottobre e dicembre 2018 mentre la prima edizione della seconda fase del Master Sample si svolgerà nel 2019.

Sono stati progettati due disegni campionari per la componente L, il primo, detto minimale, prevede circa 500.000 famiglie su circa 2.400 comuni, il secondo, detto ottimale, circa 1.000.000 di famiglie su circa 2.900 comuni. Si è deciso di utilizzare dal 2018 al 2021 disegno ottimale, dal 2021 in poi disegno minimale.

## 5. Studio sperimentale

La proposta di piano di diffusione a livello nazionale sarà articolata su varie aree tematiche e con diversi livelli di dettaglio classificatorio e territoriale.

Ad esempio a livello comunale gli ipercubi riguarderanno:

- struttura della popolazione per sesso, età e cittadinanza;
- migrazioni interne;
- popolazione per grado di istruzione;
- popolazione per condizione professionale o non professionale;
- pendolarismo per motivi di studio o lavoro;
- tipologie di famiglia e di nucleo familiare;

È stato eseguito uno studio sperimentale per analizzare il livello dei coefficienti di variazione legati ai seguenti ipercubi comunali:

- condizione professionale e lavoro: condizione professionale (6) per sesso (2);
- istruzione e formazione: titolo di studio (7) per sesso (2);
- pendolarismo: luogo di destinazione (2) per motivo dello spostamento (2) per sesso (2)

È stato considerato come livello di errore accettabile per la pubblicazione del dato, un CV pari al 33%.

In questa analisi sono riportati esclusivamente i valori dei CV relativi a stime dirette. Si deve quindi tener conto che tali valori possono essere migliorati sensibilmente attraverso l'uso di modelli e/o l'aggregazione temporale e/o geografica.

### 5.1. Condizione professionale e lavoro

Il Registro sull'Occupazione contiene informazioni per la parziale sostituibilità di alcune delle modalità relative alla variabile condizione professionale (occupato, percettore di una o più pensioni per effetto di attività lavorativa precedente o di redditi da capitale, studente). Si sono analizzate tre modalità non sostituibili (*in cerca di occupazione, casalingo/a e in altra condizione*) e una modalità parzialmente sostituibile (*occupato*).

Per ciascun disegno si è valutata una soglia di CV per la pubblicazione dei dati pari al 33%.

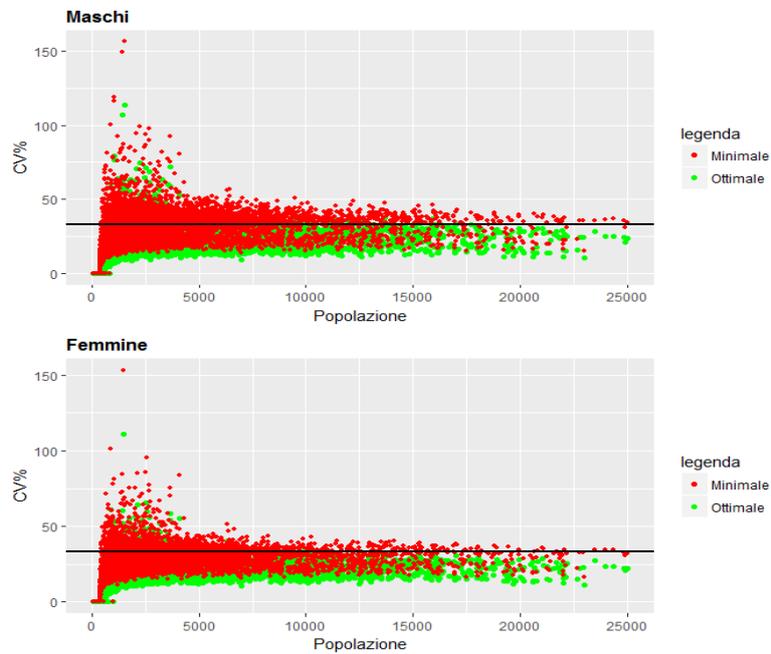
Per la variabile *in cerca di occupazione* modalità maschi (tasso di incidenza sul totale della popolazione a livello nazionale uguale a 2.4%) il numero percentuale di comuni con CV al di sotto della soglia critica del 33% è uguale rispettivamente a 64.6 e 93.5 per il disegno minimale ed ottimale. Per quanto concerne la modalità femmine (valore del parametro a livello nazionale uguale a 2.5%), tali quantità assumono i valori 78.3 e 97.7. Nel caso in cui si cumulassero i campioni per i comuni autorappresentativi, si osserva un leggero miglioramento dei valori sopra indicati (pari circa all'1-2% per entrambe le soglie di CV).

Per la modalità *casalingo/a*, i tassi di incidenza a livello nazionale sono pari a 0.2% per i maschi e 9.4% per le femmine. Per i maschi la percentuale di comuni con CV al di sotto della soglia del 33% è uguale a 5.3 e 13.2 rispettivamente per i disegni minimale ed ottimale. Per le femmine, indipendentemente dal disegno, tutti i comuni mostrano CV inferiori alla soglia del 33%.

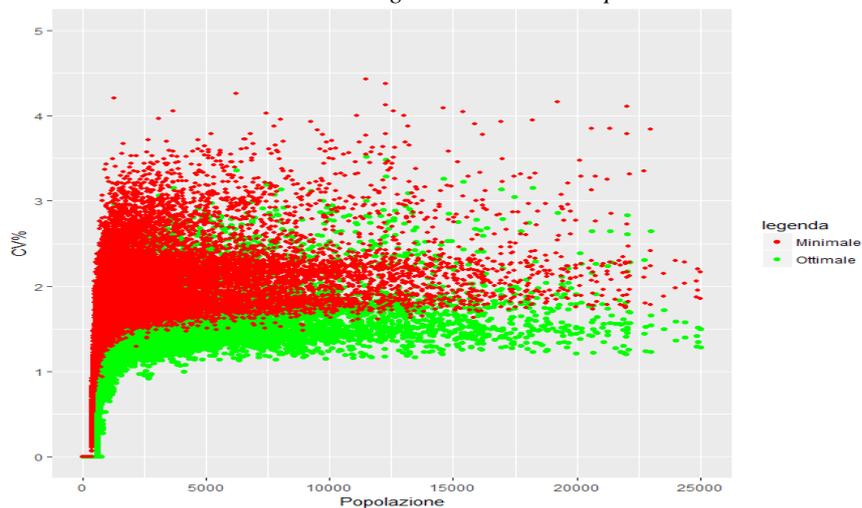
Per quanto riguarda la variabile *in altra condizione* (valori del parametro a livello nazionale pari a 2.7% per i maschi e 2.0% per le femmine), la percentuale di comuni con CV al di sotto della soglia del 33% è uguale a 21.7 e 28.2 per i maschi e 21.5 e 28.0 per le femmine rispettivamente nei disegni minimale ed ottimale.

Per quanto riguarda la variabile parzialmente sostituibile *occupato*, la percentuale di comuni con CV al di sotto della soglia del 33% sia per i maschi sia per le femmine è sempre inferiore al 5% qualsiasi disegno si prenda in considerazione.

**Figura 1** – Distribuzione dei CV: stima (comuni NAR) del numero di persone in cerca di occupazione.



**Figura 2** – Distribuzione dei CV relativi alla stima comunale del numero di occupati secondo i due scenari di disegno del Master Sample.



## 5.2. Istruzione e formazione

La Base informativa longitudinale di microdati da fonte amministrativa sull'Istruzione e sui Titoli di studio (BIT) congiuntamente ai dati relativi al Censimento della popolazione del 2011 fornisce informazioni sulle modalità della variabile "grado di istruzione della popolazione residente di 6 anni e più".

Al momento non è stata effettuata l'integrazione della BIT con la base dati censuaria, così come non è stata eseguita una valutazione delle carenze informative delle variabili contenute nella base dati. In questa analisi, si ipotizza completa sostituibilità delle modalità che si suppone possano avere un grado di copertura molto alto e quindi CV molto bassi, limitando l'analisi al totale di analfabeti, alfabeti privi di titolo di studio, individui in possesso di diploma terziario non universitario del vecchio ordinamento e diplomi di Alta Formazione Artistica, Musicale e coreutica (A.F.A.M.).

Per ciascuno scenario è stata considerata anche in questo caso una soglia di CV per la pubblicazione dei dati inferiore al 33%.

Per la variabile *analfabeti* maschi (tasso di incidenza percentuale a livello nazionale pari a 0.4%) la percentuale di comuni con CV al di sotto della soglia del 33% è pari rispettivamente a 7.5 e 19.5 rispettivamente per il disegno minimale ed ottimale. Le analoghe percentuali per la modalità femmine (valore del parametro a livello nazionale uguale a 0.6%) risultano pari a 16.0 e 32.8. Da notare che per i comuni con CV maggiore del 33% solamente per circa il 3.5% dei casi si hanno celle con numerosità esigua (minore di 5 individui).

Per gli *alfabeti privi di titolo di studio*, sia per i maschi che per le femmine (con tassi di incidenza a livello nazionale uguali rispettivamente a 3.1% e 4.2%), la quasi totalità dei comuni (tranne circa 80 comuni nel disegno minimale) ha CV al di sotto della soglia del 33%. Si evidenzia che per i comuni con CV maggiore del 33% solamente nel 2%-3% dei casi si hanno celle con numerosità esigua (minore di 5 individui), ad eccezione delle femmine nel disegno minimale per le quali tale percentuale è pari a circa il 10%.

Per la modalità maschi della variabile *diploma terziario non universitario del vecchio ordinamento e diplomi A.F.A.M.* (valori del parametro a livello nazionale uguale a 0.1% per i maschi e 0.2% per le femmine), la percentuale di comuni con CV al di sotto della soglia del 33% è pari rispettivamente a 7.1 e 18.2 per disegno minimale ed ottimale; per le femmine tali percentuali assumono i seguenti valori: 11.2 e 27.6. Come nel caso della variabile *analfabeti* circa il 3.5% dei comuni che presentano un valore del CV maggiore del 33% hanno celle con numerosità inferiore a 5 individui.

### 5.3. Pendolarismo

I dati censuari sul pendolarismo si basano sugli spostamenti giornalieri per studio o lavoro che vengono effettuati partendo dal luogo di dimora abituale e rientrando nel medesimo. Al momento i dati amministrativi non consentono di identificare con esattezza tale popolazione di riferimento.

In questa analisi, si sono prese in considerazione le variabili: *motivo dello spostamento* (con le modalità studio e lavoro), *luogo di destinazione* (con le modalità intra comunale ed extra comunale), entrambe parzialmente disponibili nelle fonti amministrative e *sesso* (con le modalità maschi e femmine). I risultati qui riportati, non essendo stato al momento ancora valutato il grado di sostituibilità della variabile *luogo di destinazione*, non tengono conto del fattore deflattivo sul CV derivante dalla sostituibilità parziale della variabile.

Per gli scenari minimale e massimale si è valutata ancora una volta la soglia di CV inferiore al 33% per ciascun incrocio delle tre variabili ai fini della pubblicazione dei dati.

Per ciascuno scenario, senza distinzioni sostanziali per *motivo dello spostamento*, *luogo di destinazione* e *sesso*, la percentuale dei comuni con valori di CV inferiori alla soglia del 33% è sempre superiore al 96% per il disegno minimale e superiore al 99% per il disegno ottimale.

**Tabella 1** – Percentuale dei comuni con valori di CV inferiori alla soglia di pubblicazione.

Incrocio	Disegno minimale		Disegno ottimale	
	Maschi	Femmine	Maschi	Femmine
Studio – Comune	96.4%	98.4%	99.4%	99.6%
Lavoro – Comune	99.8%	98.1%	100.0%	100.0%
Studio – Extra Comune	96.0%	97.6%	99.7%	100.0%
Lavoro – Extra Comune	100.0%	99.5%	100.0%	99.9%

## 6. Conclusioni

Dall'analisi sull'attendibilità delle stime dirette per il piano di diffusione comunale, si evince che, sulla base di quanto sin qui studiato, anche con il disegno ottimale permangono numerose situazioni problematiche e critiche in termini di attendibilità delle stime. Occorre, tuttavia, tenere presente che il Sistema Integrato Censimento ed Indagini Sociali è progettato a supporto del sistema dei registri. L'inadeguatezza del disegno ottimale a rispondere alle numerose situazioni critiche potrà, quindi, trovare una soluzione, parziale o totale, nella misura in cui il sistema dei registri andrà velocemente a regime nel corso dei prossimi anni e saranno

disponibili più informazioni di fonte amministrativa rispetto ad oggi. Pertanto, bisognerà incrementare gli sforzi nella duplice direzione di miglioramento della strategia campionaria - sia in termini di ottimizzazione del disegno di campionamento che delle procedure di stima - e di aumento in termini di quantità e qualità delle informazioni in gran parte sostituibili da registro. Il successo di entrambe queste linee strategiche potrà consentire di riuscire a massimizzare l'informazione producibile dal Censimento.

### **Riferimenti bibliografici**

ABS 2015. Australian Bureau of Statistics announces new population survey, downloadable at: <https://coastalcouncils.org.au/2015/09/13/australian-bureau-of-statistics-announces-new-population-survey/>.

DURR J.-M. 2005. The French new rolling census. *Statistical Journal of the United Nations Economic Commission for Europe*, 22(1), 3-12.

EUROSTAT 2013. Roadmap for the integration of European social surveys, downloadable at:

[https://ec.europa.eu/eurostat/cros/system/files/D12\\_Roadmap.pdf](https://ec.europa.eu/eurostat/cros/system/files/D12_Roadmap.pdf).

IOANNIDIS E., MERKOURIS T., ZHANG L.C., KARLBERG M., PETRAKOS M., REIS F. and STAVROPOULOS P. 2016. On a Modular Approach to the Design of Integrated Social Surveys, *Journal of Official Statistics*, 32(2), 259–286.

ONS 2017. Annual assessment of ONS's progress towards an Administrative Data Census post-2021, downloadable at:

<https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusannualassessments>.

PFEFFERMANN D. 2015. Methodological Issues and Challenges in the Production of Official Statistics, *Journal of Survey Statistics and Methodology*, 3, 425–483.

## SUMMARY

### **Permanent census: first analysis on dissemination of statistics at municipality level**

The need to meet stakeholders demand of information at high levels of quality, timeliness, territorial disaggregation and the tight constraints in terms of cost and statistical burden, pushed the Italian National Statistical Institute to redesign the production processes of the population census in an innovative way. Specifically, the next census will be realized by integrating information from thematic registers and sample surveys.

The new general design of the next census aims to combine the different needs for which it has been planned: cost reduction; improve the quality of social surveys and census statistics; the extension of the detail of social-demographic and economic indicators produced by the Institute.

The aim of this work is to highlight how the new strategy is able to disseminate efficient estimates at a wide level of territorial disaggregation. To this aim, we produced estimates for some hypercubes, defined on the basis of the cross-classification of demographic variables, labour force status variables, educational level and commuting flows at municipality level.

---

Federico BROGI, Istat, fedebrogi@istat.it

Valeria QUONDAMSTEFANO, Istat, quondamstefano@istat.it

Fabrizio SOLARI, Istat, solari@istat.it

## FAMIGLIE, POVERTÀ E RICCHEZZA: ITALIA E SPAGNA A CONFRONTO<sup>1</sup>

Maria Carmela Miccoli, Antonella Biscione

### 1. Introduzione

Nel secondo dopoguerra, in occidente, i cambiamenti nella istituzione “famiglia” sono stati notevoli ed hanno anche coinvolto le relazioni interne ai suoi componenti: è, ad esempio, ormai del tutto desueta l’ideologia del *male breadwinner*, cioè dell’uomo che procaccia le risorse necessarie mentre alla donna compete la cura e la gestione della famiglia. A tale cambiamento nei valori riposti nell’istituzione famiglia ha contribuito il contesto economico generale che da diversi anni appare particolarmente sfavorevole soprattutto per i Paesi dell’Europa mediterranea.

Il presente lavoro focalizza la sua attenzione su due di tali Paesi: L’Italia e la Spagna. L’analisi utilizza un *data base* che consente il confronto strutturale, per i due Paesi e per le macro aree che li compongono, tra le situazioni relative al 2010 e quelle relative al 2013. È con riferimento a tale intervallo temporale che verranno esaminati i cambiamenti strutturali intervenuti nella famiglia considerata sia in relazione alla sua composizione numerica ed alla sua collocazione territoriale, che in relazione al livello di povertà e/o ricchezza.

Il confronto tra le misure di povertà e quelle di ricchezza parte dalla consapevolezza che, negli ultimi anni, i due Paesi hanno dovuto affrontare numerosi problemi di natura finanziaria. Tali problematiche si sono ampiamente accentuate a causa della massiccia competizione economica a livello globale che ha esacerbato il dibattito in ambito scientifico e politico incentrato sull’incremento della forbice tra i ricchi ed i poveri.

Nel presente lavoro, al fine di osservare il grado di povertà e di ricchezza nelle due nazioni oggetto di analisi, sono utilizzate due classi di indici, Più precisamente per la misurazione della povertà si è utilizzata la classe di indici proposta da Foster et al. (1984), mentre e la classe di indicatori sviluppata da Peilch et al. (2006) è

---

<sup>1</sup> Le autrici ringraziano l’anonimo *referee* per i preziosi suggerimenti. M. C. Miccoli ha scritto i paragrafi 2 e 3; A. Biscione i paragrafi 4 e 5; l’Introduzione e le Conclusioni sono frutto del lavoro comune.

stata utilizzata per misurare la ricchezza. È da notare che il nuovo approccio proposto da questi ultimi risulta essere molto sofisticato, in quanto consente di misurare non solo la numerosità della popolazione ricca, ma anche la dimensione della variazione del loro livello di ricchezza (Peilch et al., 2006).

Il presente lavoro si articola come segue: nella prima parte è analizzata la distribuzione dimensionale delle famiglie sui territori italiano e spagnolo; nella seconda il rapporto tra le strutture familiari dei due Paesi e i livelli di occupazione; nella terza sezione sono illustrati gli indicatori di povertà e di ricchezza utilizzati; mentre l'ultima parte è dedicata alla comparazione dei due fenomeni a livello di territori nazionali e di macro-aree interne ai due Paesi.

## 2. La distribuzione dimensionale e territoriale delle famiglie

Iniziando ad esaminare l'ambito strettamente demografico si evidenzia che i cambiamenti hanno visto una netta contrazione del tasso di fecondità totale: ricordiamo, usufruendo dei dati diffusi da Eurostat<sup>2</sup>, che nel quadriennio in analisi in Spagna il tasso di fecondità totale è risultato pari a: 1,37, 1,34, 1,32, 1,22 per cui si evidenzia un trend in sistematica contrazione, il dato dell'Italia anch'essa considerata globalmente, fa rilevare sempre un tasso decrescente, i cui valori nel periodo sono pari a 1,46, 1,44, 1,43, 1,39. Si tratta di risultati leggermente più elevati di quelli emersi per la Spagna, ma sempre estremamente lontani dalla soglia di sostituzione generazionale.

Venendo poi alle unioni coniugali ed alla loro possibile rottura, che si analizza utilizzando i dati sulla divorzialità, si deve preliminarmente notare che le divergenze comportamentali eventualmente rilevabili non possono che risentire dei divergi normativi esistenti tra Paesi. In Italia, infatti, c'è l'obbligo di sostenere due cause: una prima relativa alla separazione a cui segue quella sul divorzio con consequenziale doppio passaggio che incide sia sul tempo occorrente che sui costi derivanti. In Spagna, invece, separazione e divorzio sono considerate due differenti opzioni a seconda delle necessità della coppia che può sia chiedere la sola separazione che chiedere ed ottenere direttamente il divorzio. Ed è anche per questo che il tasso grezzo di divorzialità per la Spagna nel periodo 2010-2012 pari a 2,2 per mille e nel 2013 scende al 2,0 per mille è più che doppio di quello italiano che si attesta allo 0,9 per mille per tutto il periodo in analisi<sup>3</sup>.

---

<sup>2</sup> Tutte le tabelle che seguono nel testo hanno, ove non diversamente indicato, come unica fonte i dati prodotti da Eurostat e presenti nell'Indagine sul reddito e le condizioni di vita per l'intervallo temporale 2010-2013.

<sup>3</sup> Cfr. Eurostat: <http://ec.europa.eu/eurostat/data/database>

Indipendentemente dalle differenze normative tra Paesi, la fragilità coniugale, sommata alla crescente consuetudine alla convivenza come modo alternativo di fare famiglia ed alla presenza sempre più rilevante di figli naturali nati da madri in età avanzata, ci presenta l'immagine di una famiglia in sistematica evoluzione.

Come in molti Paesi occidentali, in Spagna e in Italia si assiste ad un incremento di famiglie costituite da un unico componente o aventi un solo genitore, situazioni queste riconducibili ai differenziali di mortalità esistenti tra i sessi, o alla instabilità coniugale di cui si è appena detto o ancora all'indebolimento del desiderio di formare una famiglia con motivazioni riconducibili a molteplici fattori (ad esempio prolungamento del tempo necessario alla formazione, difficoltà di inserimento nel mondo del lavoro, prolungata presenza nella casa dei genitori utile anche per contenere i costi di gestione di una abitazione, per poi non trascurare anche servizi accessori come aiuto nelle strette faccende domestiche, utilizzo di un genitore per servizi di segreteria, etc).

**Tabella 1 - Percentuale di famiglie in relazione al numero dei componenti, Italia, 2010 e 2013**

Numero componenti	Nord Ovest	Nord Est	Centro	Sud	Isole	Italia
2010						
1	34,48	32,72	30,52	26,36	29,28	31,07
2	28,92	28,57	27,91	22,14	25,27	26,83
3	19,91	19,33	20,52	20,60	19,78	20,05
4	13,14	15,78	17,33	23,38	19,33	17,32
5	3,24	2,41	2,60	6,11	5,19	3,76
6	0,32	1,14	0,76	1,17	1,06	0,83
7+	0,00	0,05	0,37	0,24	0,10	0,15
Totale	100	100	100	100	100	100
2013						
1	35,56	32,40	33,14	26,86	30,77	31,20
2	27,98	27,96	26,30	22,65	23,65	25,70
3	18,52	20,41	19,97	20,64	20,09	20,36
4	14,54	15,04	15,96	21,57	19,40	17,22
5	2,51	3,04	3,38	6,75	4,75	4,26
6	0,64	0,96	1,07	1,13	0,73	0,95
7+	0,25	0,20	0,20	0,39	0,61	0,31
Totale	100	100	100	100	100	100

Le Tabelle 1 e 2 sono state elaborate utilizzando le informazioni desunte dall'indagine Survey EU-Silc condotta dall'Eurostat. In esse sono riportate le composizioni percentuali delle famiglie in relazione alla loro numerosità e alla loro collocazione sul territorio.

**Tabella 2** - Percentuale di famiglie in relazione al numero dei componenti, Spagna, 2010 e 2013

Numero componenti	Nord Ovest	Nord Est	Compr. Madrid	Centro	Est	Sud	Regioni extra	Spagna
2010								
1	25,73	24,17	23,62	24,28	24,49	20,65	22,29	23,62
2	28,39	31,85	29,32	30,83	29,69	28,25	28,31	29,32
3	23,31	19,78	21,52	20,73	21,64	21,98	22,75	21,52
4	16,30	18,47	18,93	18,52	17,44	21,18	17,30	18,93
5	4,11	4,05	4,67	3,99	4,76	6,19	5,88	4,67
6	1,54	1,03	1,19	1,08	1,47	0,84	2,45	1,19
7+	0,62	0,64	0,76	0,58	0,51	0,90	1,03	0,76
Totale	100	100	100	100	100	100	100	100
2013								
1	21,60	23,85	21,25	22,54	20,84	18,41	19,38	20,92
2	30,86	31,19	29,55	30,36	31,50	27,37	29,05	30,05
3	24,21	21,56	22,67	21,02	22,96	23,02	23,17	22,68
4	18,06	18,65	19,69	20,29	17,85	22,64	20,56	19,60
5	3,23	3,65	4,64	4,18	4,65	6,59	6,56	4,83
6	1,51	0,59	1,39	1,21	1,38	0,75	0,76	1,14
7+	0,53	0,51	0,82	0,38	0,83	1,22	0,53	0,78
Totale	100	100	100	100	100	100	100	100

Entrambe le realtà sono state suddivise secondo le classiche ripartizioni territoriali (5 per l'Italia e 7 per la Spagna), oltre al quadro generale riferito all'intera nazione. È subito evidente la più elevata propensione rilevabile in Italia nella presenza di famiglie unipersonali e questo è verificabile sia in ogni raggruppamento che nella Italia nel suo insieme. Nel momento in cui si osservano famiglie composte da una numerosità superiore si ritrova essere la Spagna a farne registrare un valore percentuale più elevato.

### 3. Famiglie e occupazione

Partendo dai valori assoluti della numerosità delle famiglie secondo il numero dei componenti, lo status occupazionale, e la collocazione sul territorio<sup>4</sup>, sempre riferendoci sia all'Italia che alla Spagna, è stata realizzata la seguente tabella a doppia entrata per esaminare la loro collocazione percentuale una prima volta rispetto al totale delle famiglie rilevate, successivamente rispetto al totale dei loro differenti status occupazionali ed infine il peso di ogni status al totale delle famiglie appartenenti ad ognuna delle sette classi considerate.

<sup>4</sup> Giova ricordare che il quadro che viene ad essere qui presentato, per ovvi motivi di spazio, riguarda la totalità delle due nazioni.

**Tabella 3** - Percentuale di famiglie secondo numerosità e stato occupazionale rispetto al totale della numerosità

Stato occupazionale	Numero dei componenti la famiglia							Totale
	1	2	3	4	5	6	7 o più	
Italia, 2010								
Occupati	39,6	32,0	43,4	38,8	33,0	26,2	20,7	37,7
Disoccupati	3,6	3,2	5,2	5,5	7,0	8,2	16,9	4,9
Inattivi	56,8	64,8	51,4	55,7	60,0	65,6	62,4	57,4
Totale	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Italia, 2013								
Occupati	38,6	31,3	44,3	39,8	30,5	29,9	27,0	37,8
Disoccupati	5,4	4,8	7,4	7,2	7,9	7,9	7,5	6,5
Inattivi	56,0	63,9	48,3	53,2	61,6	62,2	65,5	55,7
Totale	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Spagna, 2010								
Occupati	37,3	37,3	43,0	41,0	32,3	31,3	29,2	39,0
Disoccupati	13,3	10,5	13,0	10,8	12,8	13,5	16,67	11,9
Inattivi	49,4	52,2	44,0	48,2	54,9	55,2	54,1	49,1
Totale	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Spagna, 2013								
Occupati	35,7	35,9	40,6	36,7	28,6	22,9	27,8	36,3
Disoccupati	15,0	12,3	16,2	14,7	15,5	15,5	22,8	14,7
Inattivi	49,3	51,8	43,2	48,6	55,9	61,6	49,4	49,0
Totale	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

Da essa emerge che in Italia le famiglie unipersonali risultano avere un peso abbastanza elevato tra gli occupati, mentre abbastanza basso è il livello dell'occupazione complessiva dei componenti delle famiglie "numerose". Non altrettanto è in Spagna, paese comunque caratterizzato da livelli di disoccupazione alquanto più elevati dell'Italia.

#### 4. Indici di povertà e di ricchezza

Pur essendo stati sviluppati in letteratura molteplici indicatori per l'analisi della povertà, nel presente lavoro è utilizzata la classe degli indici FGT proposta da Foster et al. (1984) che può essere espressa nel seguente modo:

$$\varphi_{FGT}(x) = \frac{1}{n} \sum_{i=1}^p \left(1 - \frac{x_i}{\pi}\right)^\alpha \quad (1)$$

dove,  $x = (x_1, x_2, \dots, x_n) \in R_+^n$ , è la distribuzione del reddito,  $n$  rappresenta il numero degli individui o delle famiglie,  $\pi$  rappresenta la soglia di povertà che è pari al 60% del reddito mediano,  $\alpha$  è un parametro di avversione alla povertà ed infine,  $p = \#\{i | x_i < \pi, i = 1, 2, \dots, n\}$  è il numero degli individui poveri. È bene

sottolineare che il parametro  $\alpha$  può assumere valori  $\geq 0$ , aumentando tale valore è possibile ponderare in misura maggiore la popolazione povera.

Inoltre, a seconda del valore che  $\alpha$  assume è possibile ottenere tre indicatori della povertà, se  $\alpha$  è pari a zero, dalla formulazione precedentemente illustrata, si ricava l'incidenza della povertà. Questo indicatore rappresenta una misura grezza della povertà e non prende in considerazione un aspetto rilevante, cioè la distanza in termini di reddito che separa ciascun individuo dalla soglia di povertà. In altri termini, esso non dice di quanto i poveri siano davvero poveri. Infatti, esso non fornisce alcuna indicazione sulla posizione dei poveri rispetto alla linea di povertà (Foster et al, 1984). Il reddito degli individui poveri può essere o molto prossimo alla soglia di povertà o molto distante dalla stessa. Una simile informazione è di fondamentale importanza soprattutto quando si vogliono quantificare i costi monetari delle azioni di lotta contro la povertà.

Di conseguenza, al fine di ottenere ulteriori informazioni circa la situazione in cui versa la popolazione indigente è necessario che  $\alpha$  assumi un valore maggiore dello zero in modo tale da ottenere altri due indicatori che forniscono delle informazioni più dettagliate, vale a dire l'intensità della povertà e la severità della povertà

L'intensità della povertà è l'indice che fornisce informazioni sull'ampiezza della povertà ed è pari al valore medio dei poverty gap in proporzione alla soglia di povertà e ci dice di quanto il reddito dei poveri in percentuale è inferiore alla soglia di povertà mentre la severità della povertà ha il merito di analizzare in modo esaustivo gli individui gravemente colpiti dalla povertà. Inoltre, ponderando in misura maggiore gli scarti che sussistono tra il reddito di coloro che sono estremamente poveri e la soglia di povertà, tale indicatore permette di misurare la severità della povertà detta anche l'intensità della povertà tra i più poveri.

Terminata la rassegna degli indicatori maggiormente utilizzati in letteratura per la misurazione della povertà è necessario soffermarci sugli indicatori per la misurazione della ricchezza i quali hanno avuto finora scarsa attenzione a causa della difficoltà che si incontra nella definizione di una linea di ricchezza. Nel presente lavoro definiamo la linea di ricchezza in modo analogo alla linea di povertà pari al pari al 200% del reddito mediano (tale linea di ricchezza è di default pari al 200%, cfr. Pechl et al, 2006), e  $r$  rappresenta il numero degli individui ricchi. Per poter computare il numero degli individui ricchi è possibile far ricorso alla formula che segue:

$$R_{HC}(x) = \frac{1}{n} \sum_{i=(1)}^n 1_{x_i > p} = \frac{r}{n} \quad (2)$$

Tale definizione presenta delle notevoli somiglianze con l'headcount ratio. Nel

caso in cui si volessero effettuare dei confronti relativi a possibili politiche e confrontare differenti scenari, questo indicatore non soddisfa in alcun modo la definizione di ricchezza. Infatti, se nessuno modificasse il suo status (ricco o non ricco) vale a dire se non ci fosse alcun cambiamento di reddito né tantomeno un trasferimento a carico di una persona ricca non si genererà nessuna variazione di questo indice.

In aggiunta, è bene ricordare che esiste una notevole differenza tra il reddito della popolazione povera e quello della popolazione ricca, infatti, mentre il reddito degli indigenti è compreso tra zero e la soglia di povertà quello degli individui ricchi presenta solo un valore limite che è rappresentato dalla linea di ricchezza. Di conseguenza, risulta necessario standardizzare il reddito delle persone ricche rispetto alla linea di ricchezza, tale operazione consente di evitare un'alta sensibilità dell'indice ai redditi molto elevati. Inoltre, anche per gli indici di ricchezza, come per gli indici di povertà, può essere applicato l'assioma focus il quale afferma che l'indice di povertà non subisce alcuna variazione se il reddito delle persone non povere si modifica e questi individui non modificano il loro status. Una seconda importante differenza che concerne gli indicatori per la misurazione della povertà e quelli relativi alla ricchezza riguarda l'assioma di trasferimento. Con riferimento agli indicatori per la misurazione della povertà, una riduzione del reddito delle persone estremamente povere potrebbe generare un effetto molto più ampio rispetto all'incremento del reddito dell'individuo relativamente più ricco. Ciò dipende dal fatto che riducendo l'utilità marginale, l'indice di ricchezza potrebbe risultare meno sensibile alle variazioni dei redditi più elevati ma sempre delle persone povere.

Pertanto, considerando quanto detto finora, è possibile definire la misura della ricchezza nel modo seguente:

$$R_{\alpha}(X) = \frac{1}{n} \sum_{i=1}^n \left( 1 - \frac{1}{\left(\frac{x_i}{p}\right)_{x_i > p}} \right)^{\alpha} = \frac{1}{n} \sum_{i=1}^n \left( \left( \frac{x_i - p}{x_i} \right)_{+} \right)^{\alpha} \quad (3)$$

Questo indice di ricchezza somiglia alla classe degli indici di povertà FGT. In questa fattispecie, l'indice di ricchezza diminuisce a seguito di un trasferimento di tipo regressivo tra un individuo ricco ed uno ancora più ricco.

## 5. Alcuni risultati empirici

Al fine di valutare il livello di povertà, e poi in un secondo momento quello di ricchezza, nel contesto italiano ed in quello spagnolo, sono stati utilizzati i dati prodotti dall'Eurostat e presenti nell'Indagine sul reddito e le condizioni di vita per

gli anni 2010 e 2013. La variabile utilizzata per osservare entrambi i fenomeni è il reddito equivalente, infatti, la soglia di povertà è pari al 60% del reddito mediano equivalente espressa in termini reali. La tabella 4 presenta i valori relativi alle tre forme di povertà: incidenza, intensità e severità della povertà in relazione alla distribuzione sul territorio italiano.

**Tabella 4 - Distribuzione della povertà territoriale in Italia, 2010-2013**

Macro Regioni	Incidenza della povertà	Intensità della povertà	Severità della povertà	Incidenza della povertà	Intensità della povertà	Severità della povertà
	2010			2013		
N-Ovest	12,74	4,79	3,14	13,88	4,73	3,15
N-Est	10,57	3,24	2,35	15,68	4,34	2,43
Centro	16,43	5,57	3,33	20,77	6,43	3,54
Sud	30,74	11,42	7,42	37,47	13,55	8,10
Isole	32,94	12,03	7,55	40,95	15,22	9,44
Italia	19,01	6,81	4,40	23,43	7,95	4,78

Osservando i dati relativi all'Italia, è possibile notare che nel 2010, a livello nazionale, circa il 19% della popolazione italiana presentava un reddito inferiore alla soglia di povertà. Tra il 2010 ed il 2013, la percentuale di popolazione povera è aumentata in misura notevole passando al 23,43%. Anche le altre forme di povertà hanno subito un incremento, infatti, l'intensità ha registrato un incremento pari all'1,15% mentre la severità ha conosciuto un aumento dello 0,30%. Inoltre, se osserviamo i tre indicatori della povertà nelle macro regioni, sono le Isole a registrare, nel 2010 e nel 2013, le percentuali più elevate, si tratta di un risultato è ben lontano da quello registrato a livello nazionale.

Analizzando, invece, la povertà in relazione alla numerosità delle famiglie, è possibile notare che sono soprattutto le famiglie monoparentali e quelle formate da cinque o più componenti ad essere maggiormente colpite dalla povertà. La situazione non muta se si analizzano due dimensioni della povertà. È opportuno sottolineare che durante il periodo considerato si è registrato una sostanziale incremento del livello di povertà che non ha colpito le famiglie composte da 7 o più componenti le quali presentano una riduzione delle tre forme di povertà.

Invece, per quanto concerne l'analisi del grado di povertà nella penisola iberica, è bene notare che tra il 2010 ed il 2013, la popolazione al di sotto della linea di povertà ha subito un lieve incremento passando dal 19,92% al 20,56%. La macro regione che presenta la porzione maggiore di poveri è quella del sud, mentre per l'anno 2013 è la regione extra a presentare il livello di povertà più elevato.

**Tabella 5. - Povertà tra le famiglie italiane, 2010-2013**

Numero componenti famiglia	Incidenza della povertà	Intensità della povertà	Severità della povertà	Incidenza della povertà	Intensità della povertà	Severità della povertà
1	25,59	9,94	7,13	27,72	10,01	6,53
2	13,42	4,51	2,69	17,81	5,50	3,11
3	14,20	4,58	2,82	18,28	5,83	3,29
4	17,71	5,68	2,99	26,17	8,76	5,01
5	31,16	11,96	7,07	34,18	12,78	7,25
6	34,74	15,13	10,88	39,87	10,19	5,06
7 e più	57,75	25,61	14,95	48,39	22,88	14,15
Totale	19,01	6,81	4,40	18,88	7,95	4,36

Quattro delle sette regioni spagnole presentano un'incidenza della povertà che risulta essere inferiore rispetto a quelle osservata a livello nazionale. Per quanto concerne le altre due forme di povertà, vale a dire l'intensità della povertà e la severità della povertà, presentano valori molto elevati, in alcune regioni, infatti, supera ampiamente i valori che caratterizzano il contesto nazionale.

**Tabella 6. – Distribuzione territoriale della povertà in Spagna, 2010-2013**

Macro Regioni	Incidenza della povertà	Intensità della povertà	Severità della povertà	Incidenza della povertà	Intensità della povertà	Severità della povertà
N-Ovest	16.63	5.85	4.11	16.49	5.49	3.31
N-Est	16,00	5,79	4,8	12,45	4,23	2,29
Com. Madrid	13,84	5,32	4,35	12,94	5,08	3,11
Centro	23,06	6,93	3,78	25,15	8,57	4,93
Est	18,08	7,06	4,43	18,8	7,81	6,59
Sud	26,44	8,45	4,54	28,66	10,75	7,65
Regione Extra	26,13	8,49	4,11	29,84	10,88	7,94
Totale	19,92	6,93	4,35	20,56	7,74	5,48

**Tabella 7 - Povertà tra le famiglie spagnole, 2010-2013**

Numero componenti famiglia	Incidenza della povertà	Intensità della povertà	Severità della povertà	Incidenza della povertà	Intensità della povertà	Severità della povertà
1	28,99	10,66	7,37	28,24	10,39	7,59
2	16,98	5,56	3,41	19,24	6,09	4,75
3	15,29	5,41	3,71	20,37	6,95	4,84
4	19,62	6,49	3,57	25,92	9,01	4,95
5	24,51	9,39	5,90	32,05	12,81	7,23

6	22,93	11,54	7,97	48,17	17,00	10,25
7 +	18,34	6,03	5,51	56,43	18,98	9,94
<i>Totale</i>	20,38	7,16	4,63	24,86	8,67	5,52

Con riferimento alla distribuzione della povertà in relazione ai nuclei familiari, anche in Spagna si riscontra la stessa situazione che caratterizza l'Italia, cioè sono i single e le famiglie più numerose ad essere colpite maggiormente dalla povertà, tale risultato si riscontra sia per l'anno 2010 che per l'anno 2013. Inoltre, sono sempre questi nuclei familiari a presentare valori elevati delle altre due dimensioni della povertà.

Per quanto concerne l'analisi della ricchezza nelle due nazioni è stata stimata una linea di ricchezza pari al 200% del reddito mediano equivalente per ogni area geografica e a livello nazionale (Tabelle 8 e 9, in cui RL sta per *Linea di Ricchezza*).

**Tabella 8.** - *Distribuzione della ricchezza in Italia, 2010-2013*

Macro Regioni	2010				2013			
	RL	HCR	FGTT1	FGTT2	RL	HCR	FGTT1	FGTT2
N-Ovest	36853	9,21	2,29	4,28	38571	7,93	2,10	4,31
N-Est	37711	6,03	1,36	2,64	37848	6,95	1,75	3,49
Centro	34515	9,09	2,16	3,71	34083	8,36	2,03	3,95
Sud	26002	9,53	2,00	3,10	25180	8,56	2,13	4,21
Isole	24778	10,67	2,89	6,62	24778	10,67	2,89	6,63
Italia	33154	8,60	2,01	3,27	33403	8,33	2,10	4,19

**Tabella 9** - *Distribuzione della ricchezza in Spagna, 2010-2013*

Macro Regioni	2010				2013			
	RL	HCR	FGTT1	FGTT2	RL	HCR	FGTT1	FGTT2
Nord-Ovest	27623	7,77	1,94	3,25	27383	8,37	1,95	3,21
Nord-Est	34157	4,95	0,94	1,29	34202	7,83	1,53	2,52
Com. Madrid	34227	7,76	1,62	2,70	35937	9,47	2,04	3,76
Centro	26578	9,45	2,02	3,57	23839	9,94	2,20	3,81
Est	29809	11,08	2,39	4,13	27641	10,31	2,33	4,03
Sud	24816	11,50	2,71	4,40	22766	12,49	3,05	5,36
Reg.Extra	25943	9,35	1,17	1,48	23400	9,30	1,80	2,95
Spagna	28613	9,49	2,03	2,89	27646	10,31	2,33	4,03

Mentre in Italia, a livello nazionale, tra il 2010 e il 2013, il numero dei ricchi si è leggermente ridotto, in Spagna l'indice di ricchezza ha registrato un incremento di circa un punto percentuale. Le isole sono l'area italiana che presenta sia il

maggior numero di ricchi che presentano la quota più elevata di poveri. La regione spagnola con percentuale più elevata di individui ricchi è invece la regione del Sud. Per un'analisi più dettagliata del livello di ricchezza è necessario fare ricorso ad altri due indicatori che consentono di andare oltre il semplice computo di coloro che sono situati, in relazione al loro reddito, al di sopra di un determinato spartiacque. Osservando tali indicatori si evince che, sia in Spagna che in Italia, è aumentato tanto il numero degli individui estremamente ricchi che la quota degli appartenenti alla classe media che presentano un reddito elevato. Infine, con riferimento al contesto italiano, ricordiamo che se è diminuito il numero degli individui che sono al di sopra della linea di ricchezza, è aumentata la dimensione della ricchezza da essi posseduta.

## 6. Conclusioni

Il processo di allargamento delle distanze tra coloro che sono ricchi e coloro che sono poveri che si è registrato in Italia e in Spagna tra il 2010 ed il 2013 è verosimilmente solo un aspetto di un fenomeno che è globale e dal quale non ci si sottrae in nome di un credo efficientista largamente diffuso ed accettato: il credo negli effetti benefici di una concorrenza che si suppone esista, ma i cui benefici evidentemente non si manifestano così come i suoi propugnatori proclamano debba essere.

## Riferimenti bibliografici e fonti dei dati.

- EUROSTAT <http://ec.europa.eu/eurostat/data/database>  
EUROSTAT 2014. *Chiffres clés de l'Europe*, Bruxelles  
EUROSTAT 2015. *Chiffres clés de l'Europe*, Bruxelles  
FOSTER J., GREER J. and THORBECKE E. 1984. A class of decomposable poverty measures, *Econometrica* (3): 761-766.  
MEDEIROS M. 2006. *Poverty, inequality and redistribution: A methodology to define the rich*, United Nations Development Programme, International Poverty Center, Working Paper no. 18, Brasilia.  
PEICHL A., SCHAEFER T. and SCHEICHER C. 2006. *Measuring Richness and Poverty. A micro data application to Germany and the EU-15*, CPE discussion paper No. 06-11, University of Cologne.

### SUMMARY

The aim of this paper is to analyze and compare the family composition with regard to the dimension, the territorial location and the levels of poverty and richness. In Italy and Spain, the two southernmost countries of Mediterranean Europe. In order to quantify the degree of poverty, the class of FGT-indicators proposed by Foster et al (1984) will be used, while, to measure wealth, the wealth index classes developed by Peilch et al (2006) will be considered. The analysis will focus on the period 2010-2013; the data that will be used are part of the official statistics produced by the National Statistics Institutes (ISTAT-INE) and the EU-Silc Survey conducted by Eurostat.

---

Maria Carmela MICCOLI, Università di Salerno; mail to: mmicoli@unisa.it  
Antonella BISCIONE, Università Cattolica “Nostra Signora del Buon Consiglio”,  
Tirana; mail to: a.biscione@unizkm.al

## **LA PRODUZIONE IN PROPRIO DI BENI ALIMENTARI. ASPETTI ECONOMICI E SOCIALI**

Andrea Ciccarelli, Adriana Monte, Gabriella Schoier

### **1. Introduzione**

L'evoluzione degli studi sulla misura del benessere economico sostenibile ha implicato e implica un crescente interesse per l'attività di produzione domestica che le famiglie svolgono producendo beni e servizi che sono destinati alla soddisfazione dei loro bisogni. Queste tematiche sono oggetto di analisi da parte di diversi ambiti di ricerca in quanto collegate alla sicurezza alimentare, al benessere, alla salute, alla cura del territorio e alla tutela del paesaggio, ma anche ad aspetti economici relativi alla produzione domestica ed alla sua valutazione, alla tutela e salvaguardia delle produzioni tradizionali. In questo contesto rientrano la coltivazione di un orto, di un terreno agricolo e l'abitudine di allevare animali da cortile per produrre alimenti destinati prevalentemente al consumo familiare.

Da un'indagine sui consumi alimentari in Friuli Venezia Giulia, svolta nel 2010 dall'IRTEF (Istituto per la Ricerca sulle Tecniche Educative e Formative di Udine) è risultato che il 47% circa degli intervistati consuma alimenti prodotti in proprio dalla coltivazione dell'orto. Questo ha portato lo stesso Istituto ad approfondire l'analisi e ad ideare, progettare e realizzare un'indagine sugli aspetti connessi alla coltivazione degli orti in Friuli Venezia Giulia nel 2016<sup>1</sup>. A partire dalle risultanze della suddetta indagine, in questo lavoro verranno utilizzate le informazioni raccolte per cercare di valutare le eventuali ricadute economiche e gli aspetti sociali collegati alla produzione da parte delle famiglie di beni alimentari derivanti da attività agricole non professionali o, più precisamente, da attività agricole svolte da persone diverse da agricoltori, utilizzando terreni agricoli non facenti parte della SAU<sup>2</sup> o terreni edificabili e giardini/cortili di abitazioni.

---

<sup>1</sup> Si ringrazia l'IRTEF di Udine per aver consentito l'utilizzo delle informazioni raccolte con l'indagine pilota "Orti non professionali in Friuli Venezia Giulia", svolta con il contributo economico di ERSAs, Agenzia regionale per lo sviluppo rurale Friuli Venezia Giulia; CAFCS.p.A.; L.T.A. Acque; IRIS Acqua; il patrocinio di INU, Istituto Nazionale di Urbanistica F.V.G.; Comune di Castelnovo del Friuli; media partner RAI F.V.G.; quotidiano Messaggero Veneto.

<sup>2</sup> Con Superficie Agricola Utilizzabile (SAU) "si indica l'insieme dei terreni investiti a seminativi, orti familiari, prati permanenti e pascoli, coltivazioni legnose agrarie e castagneti da frutto. Essa costituisce la superficie

## 2. La coltivazione non professionale: alcune problematiche

Le attività dell'agricoltura, della zootecnia, della silvicoltura sono quelle che utilizzano risorse di origine vegetale e animale per la produzione di prodotti agricoli e di prodotti animali, e comprendono anche la loro trasformazione (ad esempio l'uva in vino) quando questa viene svolta trasformando in prevalenza i propri prodotti<sup>3</sup>. Tali attività possono essere finalizzate alla produzione per il mercato e per l'autoconsumo da parte di unità produttive, che siano famiglie di agricoltori o aziende agricole, ma possono essere svolte anche da famiglie i cui componenti svolgono attività lavorative in settori diversi dall'agricoltura.

Possiamo considerare in questo contesto la coltivazione di tipo non professionale come quella attività svolta dalle famiglie consumatrici non agricole; essa si caratterizza per il fatto che viene svolta per soddisfare prevalentemente i bisogni delle famiglie stesse, le quali producono in proprio una parte o tutti i prodotti di agricoltura, zootecnia e silvicoltura di cui hanno bisogno. Lo svolgimento di tale attività da parte di una famiglia implica vari aspetti, alcuni marginali, ma che richiamano importanti problematiche. Qui di seguito richiamiamo alcune di esse.

In primo luogo l'utilizzo di suolo del quale la famiglia stessa dispone: indipendentemente dal titolo di godimento del terreno utilizzato, si tratta comunque di terreni considerati non agricoli (terreni edificabili in attesa di essere utilizzati o terreni su cui è costruito un edificio) o di terreni agricoli che non rientrano nel calcolo della superficie agricola utilizzabile (SAU) in quanto appartenenti a privati (ma spesso precedentemente appartenuti a un'azienda agricola – o ad un agricoltore – che ha cessato la loro attività). Conoscere a quanto ammonta tale superficie produttiva può essere utile anche per spiegare, almeno in parte, la continua diminuzione della SAU rilevata ai vari Censimenti dell'agricoltura<sup>4</sup>.

Altro aspetto connesso all'utilizzo del suolo è la cura del paesaggio, che appare un problema di sempre maggior interesse, collegato all'utilizzo di terreni non più appartenenti ad aziende agricole o ad agricoltori da parte delle famiglie.

Per quanto riguarda alcuni degli aspetti economici connessi alla coltivazione non professionale degli orti, appare interessante il tentativo di valutare quanto peso possa avere la produzione derivante da tale attività rispetto a quella derivante da attività svolte da unità economiche produttive in una determinata area, ad esempio una regione, come anche comprendere quanta parte dei consumi delle famiglie è

---

investita ed effettivamente utilizzata in coltivazioni propriamente agricole. È esclusa la superficie investita a funghi in grotte, sotterranei o appositi edifici”, si veda <http://www.istat.it/it/strumenti/definizioni-e-classificazioni/glossario>.

<sup>3</sup> Si veda Istat (2009), *Classificazione delle attività economiche - Ateco 2007*, Collana Metodi e Norme n. 40, disponibile all'indirizzo internet: <http://www.istat.it/it/strumenti/definizioni-e-classificazioni/ateco-2007>.

<sup>4</sup> Si veda al riguardo Spigola, 2012.

soddisfatta da questi beni prodotti in proprio.

Anche gli aspetti sociali rivestono un importante ruolo. La crescente consapevolezza del concetto di “benessere sostenibile” e della cura della salute propria e dei propri cari nella duplice accezione di salute fisica collegata alla alimentazione e al movimento, ma anche di salute emotiva e della necessaria esistenza e cura di relazioni sociali.

### 2.1. *Aspetti economici*

Il *System of National Accounts* (SNA 2008) indica tra i tipi principali di beni e servizi prodotti e consumati all'interno della medesima famiglia «il cibo o altri beni agricoli prodotti per il proprio consumo finale dagli agricoltori, includendo gli agricoltori di sussistenza o altri per i quali la produzione agricola è solo un'attività secondaria o anche del tempo libero»<sup>5</sup>. Nella stima degli aggregati di contabilità nazionale vengono considerati però solo gli autoconsumi dei produttori agricoli come facenti parte della produzione<sup>6</sup>, mentre le restanti famiglie vengono ritenute consumatrici e quindi viene rilevato il loro autoconsumo nell'indagine sulla spesa delle famiglie. In tale indagine l'Istat rileva «le spese sostenute dalle famiglie residenti per acquisire beni e servizi destinati al consumo familiare o per effettuare regali a persone esterne alla famiglia», rileva anche gli autoconsumi e in tale definizione vengono compresi anche «i beni provenienti dal proprio orto o dalla propria azienda agricola direttamente consumati dalla famiglia (autoconsumi) o regalati»<sup>7</sup>. Tali autoconsumi vengono rilevati in termini di spesa che si sarebbe sostenuta se tali beni fossero stati acquistati sul mercato, indipendentemente dal fatto che siano stati prodotti dalla famiglia stessa o ricevuti in regalo<sup>8</sup>.

Lo stesso SNA 2008 evidenzia come «in linea di principio, tutti i frutti e i vegetali coltivati per il proprio uso personale dalla famiglia con piccoli appezzamenti o con ampi orti dovrebbero essere inclusi entro i confini della produzione, anche nei paesi sviluppati. In pratica è improbabile che venga apprezzato lo sforzo di produrre stime senza che gli ammontari coinvolti siano significativamente grandi»<sup>9</sup>, tuttavia in alcune economie locali (ad esempio a livello regionale) la possibilità di stimare tali produzioni può essere utile proprio nell'ottica di misurarne la portata sia rispetto alla produzione agricola rilevata, sia

<sup>5</sup> Si veda *System of National Accounts 2008*, United Nations, New York, 2009, pag 186.

<sup>6</sup> Si veda Istat, *Metodologie di Stima degli aggregati di contabilità nazionale a prezzi correnti*, Metodi e Norme N. 21, 2004, pag. 72.

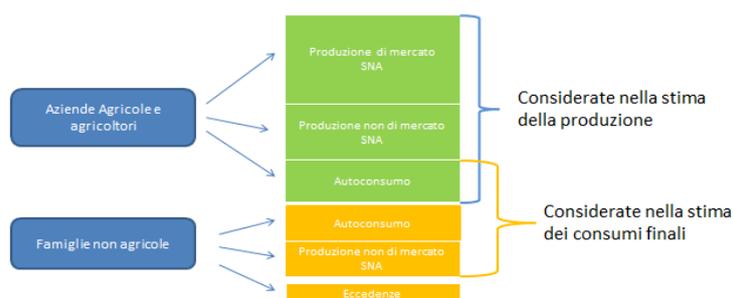
<sup>7</sup> Si veda Nota Metodologica in <http://www.istat.it/it/archivio/164313>.

<sup>8</sup> Nell'ottica inoltre della costruzione del conto satellite delle famiglie la stima degli autoconsumi dei beni prodotti in proprio dalle famiglie, affrontata da un approccio di output (rivelazione indagine sulla spesa delle famiglie) o di input (possibile stima attraverso indagini sull'uso del tempo) è esaurientemente approfondita in Montella, 2012.

<sup>9</sup> Si veda *System of National Accounts 2008*, United Nations, New York, 2009, pag. 466.

rispetto al consumo effettuato. Ciò che infatti sfugge è a quanto ammonta la produzione derivante da agricoltura non professionale e quanta parte di tale produzione serve per i bisogni della singola famiglia e quanta invece viene ceduta alle altre famiglie. Tutto questo infatti porterebbe a stimare in che misura parte del consumo locale di tali beni possa essere soddisfatto anche con quella parte di produzione da coltivazioni non professionali che dovesse rimanere non utilizzata (si veda anche la Figura 1).

**Figura 1** – Produzione agricola.



Vi è infatti un crescente interesse per la produzione “locale” come parte del sistema globale di produzione di cibo. In questa ottica l’orto domestico potrebbe significativamente contribuire ad un futuro alimentare sostenibile<sup>10</sup>.

## 2.2. Aspetti sociali

Un ulteriore aspetto che va tenuto in adeguata considerazione riguarda i riflessi di natura sociale che, più o meno direttamente, inducono le famiglie ad occuparsi della coltivazione di un orto, e che possiamo generalmente ricondurre alle tematiche del benessere e della sostenibilità.

Nelle economie più avanzate, infatti, la spinta alla coltivazione di generi alimentari ha raramente come fine primario il sostentamento della famiglia, sebbene anche nel nostro paese simili esigenze non siano poi così lontane nel tempo (si pensi all’esempio degli *orti di guerra* sperimentati anche nelle grandi città durante la seconda guerra mondiale). L’origine di tale usanza, del resto, può essere ricondotta al periodo dell’industrializzazione e delle grandi migrazioni verso le aree urbane, durante le quali le famiglie venute dalle campagne, generalmente

<sup>10</sup> Gosh, 2014, sulla base delle caratteristiche morfologiche del terreno, dell’apporto calorico delle coltivazioni, e in base al tipo di dieta standard prova a fornire una stima quantità sufficiente di ortofrutta per persona all’anno individuando come dimensione di base dell’area coltivabile una superficie media di 60 m<sup>2</sup>.

emarginate dal punto di vista economico e sociale, utilizzavano i cosiddetti orti dei poveri per mitigare la situazione di povertà e malnutrizione (Tei et. al., 2010).

Oggi, piuttosto, la coltivazione di un orto può essere più correttamente fatta ricondurre alla necessità di ricostruire in qualche modo il contatto con la natura (generalmente perso da quanti vivono nelle aree urbane, e che vedono nella coltivazione dell'orto un'occasione per diminuire lo stress fisico e psichico), alla volontà di accedere a cibi biologici e naturali (dei quali siano riconoscibili le caratteristiche organolettiche oltre a quelle di tracciabilità/sicurezza alimentare) o anche al tentativo di educare alla sostenibilità (si veda il caso degli *orti scolastici*) fino, addirittura, all'utilizzo dell'ortocultura per fini terapeutici (ad esempio, nell'assistenza agli anziani o a persone con diverse abilità – Peano et al., 2014).

Non si dimentichi, poi, che un numero sempre crescente di famiglie che si dedica a tali attività agricole potrebbe avere anche un consistente impatto sul territorio circostante e, quindi, sulla sostenibilità ambientale futura; come ben sottolineato dall'OCSE, infatti, «l'agricoltura può modificare il paesaggio, contribuire alla gestione sostenibile delle risorse, alla preservazione della biodiversità, a mantenere la vitalità economica e sociale delle aree rurali» (OCSE, 2001). In tal senso, e in talune località, la dimensione del fenomeno potrebbe addirittura suggerire di impostare piani di progettazione e pianificazione del territorio (Tei et. al., 2010).

### **3. Un'indagine sulla coltivazione non professionale degli orti in Friuli Venezia Giulia**

Di seguito vengono riportate alcune evidenze ricavate da un'indagine sulla coltivazione non professionale degli orti in Friuli Venezia Giulia (FVG) svolta nel 2016 dall'Istituto per la Ricerca sulle Tecniche Educative e Formative (IRTEF) di Udine. In una precedente indagine sulle abitudini alimentari condotta dallo stesso Istituto nel 2010 su un campione di consumatori del FVG, risultava che il 46,5% degli intervistati consumava anche beni alimentari prodotti in proprio dalla coltivazione dell'orto, pur non svolgendo un'attività agricola professionale.

L'indagine del 2016 è stata quindi un'indagine in profondità svolta sugli aspetti collegati alla coltivazione dell'orto. Il campione osservato è costituito da 1007 coltivatori non professionali, una parte dei quali, circa un terzo, appartiene a quel 46,5% indicato sopra, mentre gli altri sono stati selezionati nel modo seguente:

- a. selezione dei comuni rientrati nell'indagine sui consumi del 2010;
- b. utilizzo software cartografici/satellitari (*google earth* e cartotecnica del NICE - FVG) per la creazione di una "mappatura" dei possibili appezzamenti "non professionali" (non esiste un "censimento" dei suddetti appezzamenti)

c. selezione casuale delle aree individuate in ciascun comune in base alla distribuzione della popolazione tra comune capoluogo e frazioni.

La superficie dei terreni utilizzati è di varia dimensione, ma il 45% ha una ampiezza che non supera i 150 mq., con una superficie mediana di 200 mq. Vi sono oltre che piccoli orti con meno di 20 mq (9,5%), anche appezzamenti di terreno con superfici superiori a 2500 mq (9,9%): si deve infatti considerare che molti appezzamenti sono stati ereditati da una precedente attività di tipo agricolo e che sono stati considerati anche superfici coltivate a bosco. Si tratta per il 96% di terreni di proprietà, ereditati (38,1%) o acquistati e nell'87% dei casi sono inclusi nel giardino dell'abitazione o sono adiacenti all'abitazione stessa.

### 3.1. Le principali attività agricole svolte

La produzione ottenuta deriva da svariate attività svolte con carattere non professionale. La Tabella 1 mette in evidenza come la coltivazione di ortaggi sia svolta dal 95,7% del campione e sia spesso accompagnata dalla coltivazione di alberi da frutta. Ben presente è anche la coltivazione sia di uva da tavola che uva da vino che viene in gran parte vinificata in proprio. Oltre alla possibile stima dell'autoconsumo familiare dei prodotti, si pone l'attenzione sulla eccedenza di produzione rispetto ai bisogni della famiglia e sull'utilizzo di tale eccedenza, che spesso viene regalata a conoscenti o amici, mentre raramente gli intervistati dichiarano di cederla dietro compenso.

**Tabella 1 – Principali attività agricole svolte.**

	%	Producono in quantità uguale o superiore ai loro consumi	Regalano più dell'80% dell'eccedenza a conoscenti e/o amici
Ortive	95,7%	30,5%	92,5%
Vitigni di uva da vino	11,0%		
<i>Oltre il 60% dell'uva destinata a vinificazione in proprio</i>	7,1% (64,9%)	26,4%	94,7%
Vitigni di uva da tavola	21,2%	81,0% <i>consuma in famiglia più dell'80% dell'uva</i>	
Alberi da frutta	63,5%	20,6%	90,7%
Piante aromatiche, nostrane, particolarità locali	9,5%	17,7% <i>se si considera anche coloro che non sanno</i>	
Piante forestali e/o boschi	12,4%	81,0% <i>impiega in famiglia più dell'80% del legname</i>	
Allevamento animali	22,9%	20,0% <i>(carni)</i>	91,7%
Apicoltura	0,7%	57,1%	

Fonte: Elaborazioni proprie su dati indagine IRTEF.

## 4. Le motivazioni e il benessere soggettivo

La considerazione della rilevanza degli aspetti sociali e della consapevolezza del benessere individuale e collettivo connessi alla coltivazione dell'orto da parte delle famiglie (evidenziata in Spigola, 2012), viene confermata dall'indagine IRTEF per il Friuli Venezia Giulia. Per il 64,3% degli intervistati infatti, la motivazione principale è da ricondursi proprio al desiderio di benessere individuale o della propria famiglia (il produrre alimenti genuini, lo stare all'aria aperta, il fare esercizio fisico e ridurre lo stress), ma anche gli aspetti socio-culturali come il mantenimento di una tradizione familiare, la conservazione di piante e colture della tradizione locale e la valorizzazione di un terreno sono indicate dal 29,5% dei rispondenti (Tabella 2). Non si lavora l'orto per un risparmio economico ed anche le eccedenze di produzione vengono infatti regalate (Tabella 1), non vendute.

Nell'indagine è stato chiesto di indicare anche una seconda motivazione oltre alla principale; dalla considerazione congiunta di entrambe si possono individuare quattro diversi atteggiamenti degli intervistati nei confronti del benessere quali motivazioni per la coltivazione dell'orto, che possono essere considerate come indicatori della maggiore o minore consapevolezza individuale delle ricadute delle scelte di benessere e salute sulla propria quotidianità (Tabella 3). L'atteggiamento riportato non dipende da condizioni soggettive dell'intervistato, come sesso, condizione professionale età e livello di istruzione, anche se le persone più anziane e con un minor livello di istruzione mostrano un atteggiamento di minor consapevolezza. Neppure la presenza di bambini (sotto i 13 anni) in famiglia modifica la situazione.

**Tabella 2 – Motivazione principale per la coltivazione non professionale dell'orto.**

Motivazione principale	
Produrre per la famiglia alimenti genuini, sicuri e salubri	34.0%
Piacere di stare all'aria aperta e in un ambiente "agricolo"	23.5%
Per mantenere una tradizione di famiglia ( <i>attuale/di origine</i> )	22.9%
È uno stile di vita che consente di ridurre lo stress e di fare esercizio fisico	6.9%
Per valorizzare un terreno che altrimenti sarebbe rimasto incolto	6.2%
Per risparmiare e contenere la spesa alimentare della famiglia	2.9%
Per conservare piante antiche e colture della tradizione locale	0.4%
Altro	3.3%

Fonte: Elaborazioni proprie su dati indagine IRTEF.

**Tabella 3 – Condizione degli intervistati rispetto al benessere e salute.**

Livello di consapevolezza	%
<u>Piena consapevolezza</u> – scelta di benessere come motivo principale e secondario	40.9%
<u>Consapevolezza</u> - scelta di benessere come motivo principale (assieme ad un motivo secondario diverso)	23.4%
<u>Parziale consapevolezza</u> – scelta di benessere solo come motivo secondario	21.2%
<u>Mancanza di consapevolezza</u> – scelta diversa da benessere sia come motivo	14.5%

---

principale che secondario

Fonte: Elaborazioni proprie su dati indagine IRTEF.

---

## 5. Un'analisi descrittiva sulle tipologie di coltivatori non professionali

Un'ulteriore analisi si è posta l'obiettivo di identificare eventuali profili diversi di coltivatori non professionali. A tale scopo sono state considerate le seguenti variabili: livello di consapevolezza (Tabella 3), presenza in famiglia di ragazzi sotto i 13 anni, titolo di studio, condizione professionale, anni di coltivazione dell'orto, età, numero componenti della famiglia. Le variabili sopra indicate sono sia quantitative che qualitative, per cui si è scelta come metodologia di analisi la *Two Step Cluster Analysis* (Bacher et al., 2004, Ming-Y et al., 2010) con inserimento casuale dei casi oggetto di analisi per ovviare al problema che l'analisi stessa ha di dipendenza dall'ordine.

### 5.1. Metodologia di analisi

In letteratura sono stati sviluppati diversi algoritmi di *clustering* per raggruppare i dati in gruppi caratterizzati da un'elevata omogeneità interna e da un'elevata eterogeneità tra i cluster, tuttavia la maggior parte dei metodi permette di ottenere buoni risultati se applicati a variabili numeriche o categoriali, ma la maggior parte di essi non dà risultati apprezzabili su variabili categoriali e numeriche rilevate contemporaneamente sui dati. La *Two Step Cluster Analysis* presenta una soluzione a questo problema (Bacher et al., 2004, Ming-Y et al., 2010). In particolare la *Two Step Cluster Analysis* (sviluppata da Chiu et al., 2001) utilizzando il linguaggio di programmazione SPSS consta di due step:

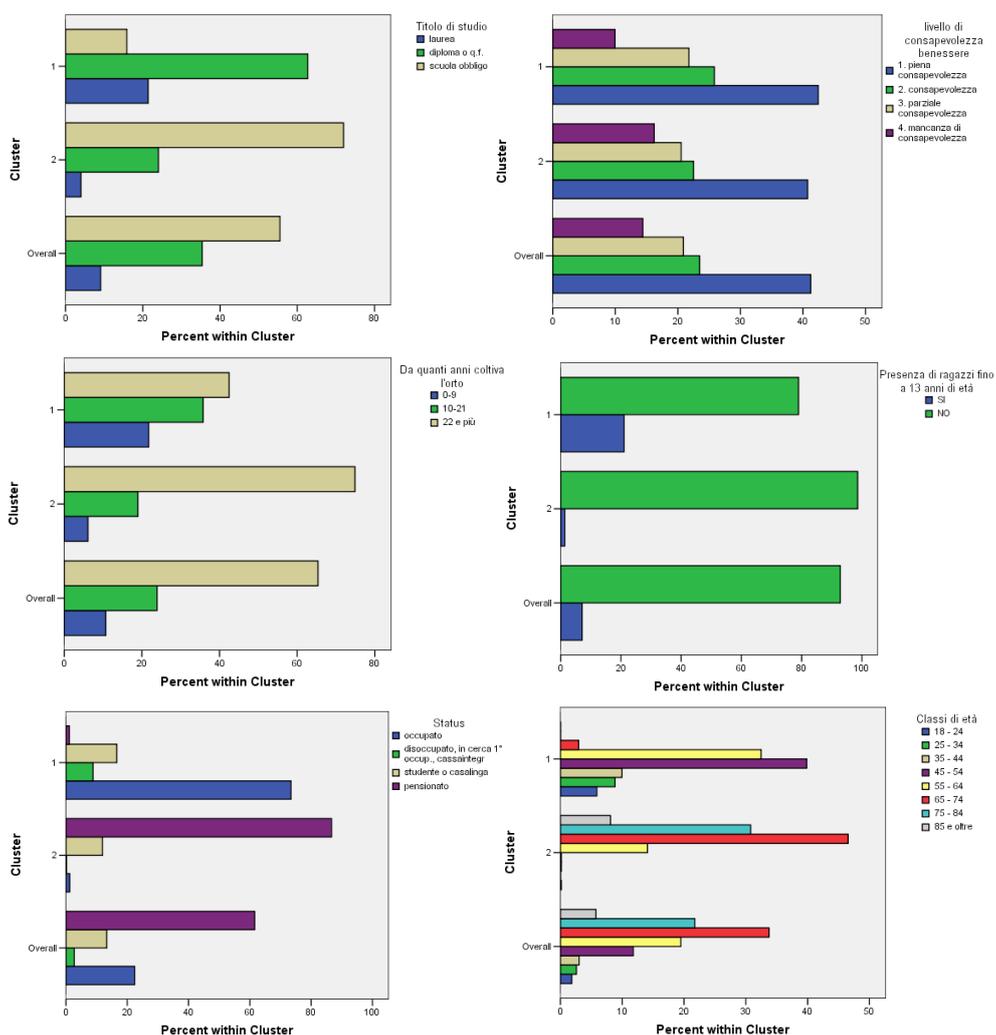
*Step 1 – Pre-clustering dei casi:* si utilizza un approccio sequenziale per quanto riguarda il pre-clustering dei casi. L'obiettivo è, a partire dalla matrice dei dati iniziale calcolare una nuova matrice di dati con meno casi da essere applicata nel passo successivo; Al fine di raggiungere questo scopo, i nuovi pre-cluster calcolati e le loro caratteristiche vengono utilizzate come nuovi casi. I pre-cluster sono definiti come regioni dense nello spazio delle variabili considerate, in questo modo il metodo può essere inserito all'interno degli algoritmi *density based*.

Si osservi che i risultati possono dipendere dall'ordine di ingresso dei casi, quindi è consigliabile utilizzare l'ordine casuale.

*Step 2 – Raggruppamento dei casi:* viene applicata una tecnica gerarchica basata sul modello (vale a dire una *hierarchical clustering model based methodology*). Simile alle tecniche gerarchiche agglomerative, i pre-cluster vengono fusi *step-by-step* finché, se non si dà una regola di stop, tutti i cluster sono uniti in un unico cluster. A differenza delle tecniche gerarchiche agglomerative

classiche, viene utilizzato un modello statistico sottostante. Il modello presuppone che le variabili continue siano distribuite all'interno dei cluster come variabili indipendenti e normalmente distribuite e le variabili categoriali siano distribuite all'interno di cluster come variabili multinomiali tra loro indipendenti.

**Figura 2 – Rappresentazione grafica delle distribuzioni delle variabili nei cluster.**



A seconda del tipo di variabili oggetto dell'analisi si utilizzano due tipi di distanza: distanza euclidea nel caso di variabili quantitative e *a log-likelihood distance* in caso variabili sia quantitative che qualitative.

## 5.2. *Principali risultati*

L'applicazione ha consentito di individuare due cluster ben definiti che suggeriscono l'esistenza di due tipologie diverse di coltivatori non professionali (si veda Figura 2). Il primo gruppo (Cluster 1 – formato da un campione di 271 soggetti) è caratterizzato da coltivatori in prevalenza giovani o di media età, con un titolo di studio medio alto, in prevalenza occupati, che lavorano l'orto anche da non molti anni, con una numerosità familiare più elevata (3,4 componenti in media) e con la presenza di ragazzi fino a 13 anni di età. L'altro gruppo (Cluster 2 – costituito da 653 coltivatori)<sup>11</sup> è costituito in prevalenza da coltivatori anziani, con un basso titolo di studio, non appartenenti alle forze lavoro, che lavorano da molti anni l'orto, con una bassa numerosità familiare (2,1 componenti in media).

Le (poche) domande utilizzabili nell'analisi non consentono una caratterizzazione estremamente fine dei due gruppi, ma emergono in ogni caso alcuni spunti interessanti: nel gruppo dei più giovani appare maggiore la spinta data dalla possibilità di produrre alimenti genuini e dalla necessità di combattere lo stress; nel cluster degli anziani le motivazioni sembrano arrivare soprattutto dalla possibilità di proseguire una tradizione familiare e dal piacere di stare all'aria aperta e in un ambiente agricolo. Il gruppo degli anziani dedica un tempo mediamente superiore alla cura dell'orto (ma questo è in parte dovuto anche alla maggiore disponibilità di tempo in assoluto, collegata allo status di pensionato), mentre appare marginale in entrambi i gruppi la motivazione legata al risparmio economico (ancorché superiore per il gruppo di anziani). La consapevolezza del benessere derivante dalla coltivazione dell'orto, infine, si mostra con intensità lievemente superiore negli appartenenti al Cluster 1 rispetto a quelli del Cluster 2.

## 6. Conclusioni

Negli ultimi anni, la coltivazione non professionale degli orti ha attirato l'attenzione di studiosi di varie discipline, potendosi questa legare a numerosi aspetti del vivere quale quelli relativi al benessere, alla qualità del cibo, alla sicurezza alimentare, alla sostenibilità ambientale e all'impatto sul territorio.

Non tutti gli elementi sopra citati, tuttavia, sono stati ancora adeguatamente investigati: sembrerebbero sufficientemente chiari i riflessi benefici in termini di

---

<sup>11</sup> Dall'analisi sono stati esclusi 83 casi, in quanto le risposte alle domande considerate erano parzialmente (o totalmente) mancanti.

attività fisica e diminuzione dello stress (Peano et al., 2010), piuttosto che di sostenibilità ambientale (Gosh, 2014) e impatto territoriale (OCSE, 2010), ancorché le associazioni evidenziate paiono investire soprattutto la sfera concettuale, mancando solidi tentativi di misurazione di tali effetti; ancora molta strada, invece, va fatta in relazione alla stima dell'impatto che tali attività possono avere a livello economico: infatti, pur se la coltivazione dell'orto rimane confinata sostanzialmente nell'autoconsumo (come altre attività similmente svolte nella famiglia), l'idea di poterla osservare in forma aggregata all'interno di un conto satellite del sistema di contabilità nazionale consentirebbe la valutazione (anche quantitativa) di un fenomeno il cui impatto potrebbe essere anche consistente.

In ogni caso, da un punto di vista micro-economico se l'analisi rimane a livello familiare, l'indagine effettuata in Friuli Venezia Giulia individuerebbe nella ricerca del benessere e nella possibilità di produrre cibi genuini e sicuri la spinta alla gestione dell'orto, relegando a stimoli solamente marginali le eventuali motivazioni di natura prettamente economica. In particolare, dall'analisi risulterebbero due tipologie ben distinte di utilizzatori: da un lato persone avanti con gli anni, presumibilmente da sempre legati alla terra, che cercano di portare avanti le vecchie tradizioni di famiglia; dall'altra un gruppo di individui generalmente più giovani, con titolo di studio mediamente elevato, la cui spinta alla coltivazione è data soprattutto da una ricerca di benessere e nelle cui famiglie, spesso, sembrerebbe emergere la presenza di ragazzi (fino a 13 anni) ai quali, presumibilmente, almeno una parte della produzione è in qualche modo "dedicata".

Ne emerge, in sostanza, un quadro complesso, ancorché non esaustivo, dei comportamenti delle famiglie, la cui incidenza sulla sostenibilità, sul territorio e sull'economia potrebbe essere ben più consistente di quanto fino ad ora ipotizzato.

### Riferimenti bibliografici

- BACHER J., K. WENZING M., VOGLER M. 2014. SPSS Two Cluster – A First Evaluation, *Universitat Erlangen-Nurnberg*, pp. 1-20, [www.statisticalinnovations.com/products/twostep.pdf](http://www.statisticalinnovations.com/products/twostep.pdf) cited July.
- CHIU T., FANG D., CHEN J., WANG Y., JERIS C. 2001. A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2001*, pp. 263–268.
- GOSH S. 2014. Measuring sustainability performance of local food production in home gardens, *Local Environment*, Vol. 19, No. 1, pp. 33-55.

- MING-YI S., JAR-WEN J., LIEN-FU L. 2010. A Two-Step Method for Clustering Mixed Categorical and Numeric Data, *Tamkang Journal of Science and Engineering*, Vol. 13, No. 1, pp. 11-19.
- MONTELLA M. 2012. La produzione domestica: il valore aggiunto generato dalle famiglie. In Dipartimento di Economia Pubblica (Ed/Eds) *Working paper*, No. 155, Università di Roma "La Sapienza".
- OCSE. 2001. *Multifunctionality. Towards an Analytical Framework*, OECD publications, Paris.
- PEANO C., GIRGENTI V., MASCETTI E., SOTTILE F. 2014. Educare alla sostenibilità: il caso degli orti scolastici, in *Campo-Territorio: rivista de geografia agraria*, Vol. 9, N. 19, p. 419-450.
- TEI F., GIAQUINTO G. 2010. Origini, diffusione e ruolo multifunzionale dell'orticoltura urbana amatoriale. *Italus Hortus* 17 (1), pp. 59-73.

## SUMMARY

### Self-production of foodstuffs. Economic and social aspects

This paper aims at investigating the typologies of non-professional economic activities linked to agriculture, to understand which are the main features of these activities in Friuli-Venezia Giulia (FVG) and the relationship between such features and those of households/individuals carrying out them. These issues are arousing growing interest in several research areas since they are connected to food security, wellbeing, health, taking care of land and landscape protection, as well as to economic aspects related to domestic production and its evaluation, the protection and preservation of traditional products. In this context, there is the cultivation of a vegetable garden, an agricultural land, and the habit of raising barnyard animals to produce food primarily intended for household consumption. This paper highlights some social and economic aspects linked to the cultivation of vegetable gardens, using data collected in an in-depth survey carried out in Friuli-Venezia Giulia by IRTEF (Institute for Research on Educational and Training Techniques in Udine). We obtain a complex, though not exhaustive, picture of household behaviours, whose impact on sustainability, territory and economy might be far more sizeable than what so far supposed.

---

Andrea CICCARELLI, Università di Teramo, [aciccarelli@unite.it](mailto:aciccarelli@unite.it)  
Adriana MONTE, Università di Trieste, [adriana.monte@deams.units.it](mailto:adriana.monte@deams.units.it)  
Gabriella SCHOIER, Università di Trieste, [gabriella.schoier@deams.units.it](mailto:gabriella.schoier@deams.units.it)

## **LE SCELTE DI CONSUMO DELLE FAMIGLIE ITALIANE<sup>1</sup>**

Andrea Cutillo, Eleonora Meli

### **1. Introduzione**

La disuguaglianza e il benessere economico delle famiglie sono solitamente misurati tramite indicatori di reddito e ricchezza; tuttavia, questi misurano la capacità di spesa delle famiglie, capacità che consente di calibrare il profilo dei consumi lungo il ciclo di vita e di soddisfare le proprie necessità, gusti e stili di vita. La valutazione dei consumi, integra all'analisi sulla distribuzione del reddito un'analisi su ciò che il reddito permette di acquistare. In questo modo, ci si riferisce a un modello in cui il benessere economico deriva dal consumo, il quale a sua volta dipende da reddito e ricchezza. Le spese per consumo rappresentano quindi il valore monetario dei beni e servizi che le famiglie acquisiscono e coglie le scelte di allocazione delle risorse per ottenere uno standard di vita che più si avvicini a quello desiderato. Nell'allocazione delle spese sulle diverse voci di consumo alcune spese risultano indifferibili per conseguire uno standard di vita accettabile, altre sono invece maggiormente soggette a scelte allocative.

In questo lavoro si analizzano le spese per consumo delle famiglie disaggregate nei gruppi sociali individuati nel Rapporto Annuale sulla situazione del Paese (Istat, 2017).

Il lavoro presenta, inoltre, un'analisi delle scelte allocative dei gruppi sociali: scorporando dalle spese per consumo quelle che risultano indifferibili per conseguire uno standard di vita minimamente accettabile. Posto che le spese indifferibili sono quelle definite dalla soglia di povertà assoluta (Istat, 2009), che rappresenta la spesa minima necessaria per acquisire i beni e servizi inseriti nel paniere di povertà assoluta. Le spese residue, di conseguenza, sono maggiormente imputabili a scelte soggettive nell'allocazione delle risorse. Analizzando la componente residua delle spese si possono tracciare dei profili di "scelta" di consumo alla luce delle caratteristiche dei gruppi sociali di appartenenza.

---

<sup>1</sup> L'intero articolo è frutto di lavoro comune; tuttavia, ad A. Cutillo possono attribuirsi i paragrafi 3 e 4; ad E. Meli i paragrafi 1, 2 e 5.

## 2. Costruzione e caratteristiche dei gruppi sociali

Nello studio delle scelte di consumo si è utilizzata la classificazione in gruppi sociali adottata dall'Istat nel Rapporto Annuale 2017. La lettura delle decisioni allocative delle famiglie in chiave di stratificazione sociale consente una caratterizzazione allo stesso tempo più articolata e più semplice: più articolata perché le dimensioni che entrano in gioco nell'interpretazione sono tutte quelle utilizzate per la creazione dei gruppi; più semplice perché analizzando le spese delle famiglie per gruppi omogenei, l'interpretazione risulta già semplificata da quelle variabili che sono entrate nella classificazione. La stratificazione delle famiglie è stata effettuata in base ad alcune caratteristiche che danno conto della multidimensionalità dell'appartenenza sociale, tale per cui nella costruzione dei gruppi si è tenuto conto sia di aspetti di natura economica (reddito, condizione occupazionale), sia culturale (titolo di studio), sia socio-demografica (cittadinanza, dimensione della famiglia, tipologia del comune di residenza). I nove gruppi individuati, sono stati creati tramite una tecnica di segmentazione che ha permesso di tralasciare qualsiasi ipotesi sulla distribuzione delle variabili di interesse, a priori, suddividendo le famiglie in funzione del loro benessere economico. I nove gruppi, risultano, così, omogenei al loro interno per livello di reddito familiare equivalente<sup>2</sup>, e ciascun gruppo è connotato da una specifica combinazione di modalità delle altre variabili considerate. Tale operazione di suddivisione statistica delle famiglie in gruppi sociali si affianca alla classica analisi della società basata sulle classi sociali più volte utilizzate dall'Istat e derivanti dalla classificazione di Schizzerotto (1988). I gruppi, così individuati, non si differenziano soltanto in base alle possibilità economiche e al contesto territoriale di appartenenza, ma anche rispetto a un insieme di caratteristiche sociali, economiche e demografiche, che risultano essere molto influenti sulle necessità e sulle scelte dei consumatori. Nell'analisi per livelli di spesa, i gruppi orientano i consumi in linea con le loro condizioni reddituali, ma la presenza di elementi di tipo culturale e sociale fa sì che le scelte allocative possano essere lette anche in funzione di queste ulteriori caratteristiche.

Di seguito verranno esposte le caratteristiche dei nove gruppi attraverso i quali sono state declinate le scelte di consumo.

Il primo gruppo sono le **famiglie a basso reddito con stranieri**, composto integralmente da famiglie in cui è presente almeno uno straniero. Presenta le peggiori condizioni economiche definite in base al reddito, con uno svantaggio di circa il 40% rispetto alla media. Sono 1,8 milioni di famiglie (il 7,1% del totale) e

---

<sup>2</sup> I gruppi sono stati individuati tramite l'indagine Eu-Silc, che rileva i redditi delle famiglie, e successivamente ricostruiti sull'Indagine sulle Spese per Consumo delle Famiglie tramite la combinazione delle variabili utilizzate, comuni tra le due indagini.

4,7 milioni di individui (il 7,8% del totale). Le famiglie con stranieri risiedono più spesso delle altre nei grandi centri urbani e sono presenti in larga misura nel Centro-nord (oltre quattro casi su cinque). Sono famiglie la cui persona di riferimento è occupata in oltre l'80% dei casi. Nonostante gli occupati siano prevalentemente in posizioni non qualificate, nella metà dei casi la persona di riferimento possiede un diploma di scuola secondaria superiore, e uno su dieci ha un titolo universitario. Rispetto al titolo di godimento dell'abitazione, è più frequente che vivano in affitto (66,3% rispetto ad una media nazionale del 19,0%) e hanno la percentuale sensibilmente più bassa di abitazioni di proprietà (16,3% contro il 68,4 di media). Stessa situazione di svantaggio economico lo vivono le **famiglie a basso reddito di soli italiani** (1,9 milioni di famiglie per 8,3 milioni di individui). Hanno un reddito medio equivalente inferiore del 30% rispetto alla media nazionale. Sono famiglie numerose, mediamente di 4,3 componenti, e in oltre il 90% dei casi si tratta di coppie con figli. Il 54,0% delle persone di riferimento ha ottenuto solamente la licenza di scuola media inferiore. Queste famiglie risiedono meno spesso delle altre nei grandi centri urbani (9,3%, la percentuale più bassa tra tutti i gruppi) e più frequentemente al Sud e nelle Isole. Questo gruppo dispone del 5,3% del reddito equivalente complessivo nazionale, denotando una delle differenze negative più ampie tra quota di famiglie (7,5%) e quota di reddito. Le **famiglie tradizionali della provincia** sono meno di un milione di famiglie (il 3,3% del totale) e raccolgono 3,6 milioni di individui (il 6,0% del totale degli individui), con un reddito equivalente medio inferiore del 25% rispetto alla media nazionale. Il principale percettore è occupato nel 70,6% dei casi; se la persona di riferimento non è occupata proviene comunque da una lunga esperienza sul mercato del lavoro, essendo pensionata da lavoro. In oltre l'80% delle persone di riferimento il titolo più alto conseguito è la licenza media e nei restanti la licenza elementare. È il gruppo a maggiore connotazione di genere, essendo la persona di riferimento un uomo in nove casi su dieci. Si tratta di famiglie numerose (in media 4,3 componenti) di tipo prevalentemente tradizionale: l'88,6% sono coppie con figli e nell'8,2% dei casi rappresentano la famiglia tradizionale a più generazioni. Questo modello di famiglia si trova con maggiore frequenza nel Mezzogiorno (46,2%) e nei comuni al di sotto di 50mila abitanti. Sono proprietari dell'abitazione in tre casi su quattro. È uno dei gruppi a minore benessere economico, in particolare in termini monetari: queste famiglie dispongono di una quota del reddito equivalente nazionale inferiore al loro peso sul totale delle famiglie (2,5 contro 3,3%). Le **anziane sole** e dei **giovani disoccupati** comprendono 3,5 milioni di famiglie (il 13,8% del totale), per un totale di 5,4 milioni di persone (l'8,9% degli individui). È un gruppo composto da famiglie di soli italiani di modeste dimensioni (fino a tre componenti), in cui sono presenti, oltre a famiglie in cui la persona di riferimento è inattiva (nell'88,7% dei casi), una

minoranza di famiglie in cui la persona di riferimento è disoccupata (il restante 11,3%). Quando la persona di riferimento è inattiva, sono prevalentemente donne in età avanzata (in media 68,4 anni) il cui reddito proviene prevalentemente da pensioni diverse da quelle da lavoro; la componente dei disoccupati è invece a prevalenza maschile e di età più giovane (43,1 anni in media). Il gruppo ha un reddito equivalente inferiore del 19,3% rispetto alla media nazionale. La disomogeneità rilevata rispetto al genere e alla condizione professionale emerge anche rispetto alla distribuzione per titolo di studio della persona di riferimento, essendo questa in buona misura guidata dalle esperienze sociali e dalle normative vigenti vissute dalle differenti generazioni: gli inattivi hanno prevalentemente un basso titolo di studio (elementari) e i disoccupati presentano invece il diploma di secondaria superiore come classe modale. Queste famiglie si concentrano con maggiore frequenza al Sud e nelle Isole. Le **famiglie degli operai in pensione** hanno un reddito equivalente medio non particolarmente distante dal valore medio nazionale, scontando una differenza di circa il 7%, e possono quindi definirsi come appartenenti alle famiglie a reddito medio. È il gruppo più corposo, con 5,8 milioni di famiglie (pari al 22,7% del totale delle famiglie) e 10,5 milioni di individui (17,3% del totale), composto in grande prevalenza da famiglie in cui il principale percettore è ritirato dal lavoro (oltre otto su dieci). Oltre i tre quarti dei pensionati da lavoro di queste famiglie erano operai o lavoratori in proprio e il loro titolo di studio arriva alla licenza di scuola media inferiore. Sono famiglie più presenti nei comuni fino a 10mila abitanti che sono riuscite e vivono in abitazioni di proprietà in più del 75% dei casi. La quota di reddito posseduta da queste famiglie (21,2% del totale) è pressoché analoga al loro peso sul totale delle famiglie. I **giovani blue-collar** sono 2,9 milioni di famiglie (l'11,3% del totale) e 6,2 milioni di individui (10,2% del totale). La persona di riferimento è operaio a tempo indeterminato in tre casi su quattro e lavoratore atipico (dipendente con contratto a termine o lavoratore indipendente con contratto di collaborazione) nei restanti casi. La persona di riferimento ha la licenza media nel 46,3% dei casi, mentre nel 40,9% ha un titolo di scuola secondaria superiore. Quanti hanno un titolo universitario sono in larghissima prevalenza in posizione di lavoratore atipico, suggerendo quindi come i contratti a termine e di collaborazione hanno sul mercato del lavoro una resa reddituale scarsa, tanto da accomunare questo sottoinsieme di lavoratori alla maggioranza operaia che definisce il gruppo. Sono proprietari dell'abitazione di residenza in circa il 60% dei casi. Le **famiglie di impiegati** hanno esclusivamente persone di riferimento occupate, nel 77,4% impiegati e nel restante 22,6% lavoratori in proprio. Hanno un titolo di studio elevato, di scuola superiore (70,9%) o universitario (23,6%), con quota elevata anche di titoli post laurea (5,5%). È un gruppo molto caratterizzato: vi si colloca la quasi totalità degli impiegati e circa la metà dei lavoratori in proprio. È un gruppo consistente (12,2 milioni di persone,

20,1% del totale; e 4,6 milioni di famiglie, il 17,8% del totale), il loro reddito equivalente ha un vantaggio di circa il 14% rispetto alla media. In quattro casi su dieci la persona di riferimento è una donna. Queste famiglie si trovano prevalentemente nel Centro-Nord e nei centri e nelle periferie delle aree metropolitane. Sono proprietari dell'abitazione di residenza in tre casi su quattro. Le **pensioni d'argento** sono un gruppo ad alto reddito, con un vantaggio di circa il 32% rispetto al reddito equivalente medio. Si tratta di 2,4 milioni di famiglie (9,3% del totale), che comprendono 5,2 milioni di individui, pari all'8,6% del totale, in cui il principale percettore di reddito è in più di due terzi dei casi ritirato dal lavoro. In oltre otto casi su dieci le famiglie hanno la casa di proprietà e sono geograficamente collocate nel Centro-Nord, nei centri e nelle periferie delle aree metropolitane. Le famiglie dei pensionati d'argento dispongono di una quota del reddito equivalente nazionale superiore al loro peso sul totale delle famiglie (12,3 contro 9,3%). La **classe dirigente** comprende le famiglie a maggiore reddito equivalente, con un vantaggio di quasi il 70% rispetto alla media. Sono 1,8 milioni di famiglie (il 7,2% del totale) e 4,6 milioni di individui (7,5% del totale). Nel gruppo c'è una non trascurabile quota di famiglie con stranieri, pari al 5,5% del totale. Sono famiglie la cui persona di riferimento ha sempre titolo universitario (in un caso su quattro ha anche conseguito un titolo post laurea). In oltre l'80% dei casi l'abitazione è di proprietà. Rispetto alla distribuzione nazionale, sono famiglie relativamente più presenti nel Centro (26,9%) e risiedono nei grandi centri delle aree metropolitane (30,3%) e nei comuni con più di 50 mila abitanti (23,2%). Detengono il 12,2% del reddito totale. È il gruppo in cui si osserva la differenza positiva più ampia tra quota di reddito e quota di famiglie (5,0 punti percentuali).

### 3. La spesa nei gruppi sociali

Sulla base delle caratteristiche descritte, si comprende come il livello e la composizione della spesa per consumi si differenzino molto tra i diversi gruppi sociali, in base non solo alle differenti possibilità economiche, ma anche in base ad alcune caratteristiche sociali, economiche, demografiche e territoriali, che risultano molto influenti sulle necessità e sulle scelte dei consumatori.

I valori di spesa seguono in maniera abbastanza lineare le differenti condizioni reddituali: quando è migliore la situazione economica si spende di più e viceversa. A loro volta, i livelli di spesa hanno un notevole impatto sulla composizione dei consumi. Coerentemente con la legge di Engel, tra le famiglie con disponibilità più contenute è più alta la quota delle spese destinate al soddisfacimento dei bisogni primari, come ad esempio quelle per alimentari, abitazione, mobili, articoli e servizi per la casa. Viceversa, tra le famiglie con possibilità economiche migliori

sale l'incidenza di spese importanti dal punto di vista dell'inclusione e della partecipazione sociale, ma non strettamente essenziali, quali quelle per servizi ricreativi, spettacoli e cultura e quelle per servizi ricettivi e di ristorazione.

Le famiglie della classe dirigente, che presentano il livello di benessere economico più elevato, spendono 3.810 euro al mese in media. Il gruppo presenta la quota di spesa alimentare più bassa di tutte, il 13,6% del totale, e le più elevate per ricreazione, spettacoli e cultura (6,6%) e per servizi ricettivi e di ristorazione (6,4%). All'estremo opposto in termini di benessere economico, la spesa delle famiglie a basso reddito con stranieri, pari in media a 1.697 euro, si concentra su beni e servizi essenziali: il 21,0% è destinato alla spesa alimentare e il 36,8% all'abitazione. Le anziane sole e giovani disoccupati spendono sui livelli del gruppo degli stranieri a basso reddito (1.700 euro). Le spese alimentari pesano intorno al 20% del loro bilancio, così come per le famiglie di italiani a basso reddito, per le famiglie degli operai in pensione e per quelle tradizionali della provincia. L'ampiezza familiare aumenta la spesa, portando le famiglie tradizionali della provincia e le famiglie di italiani a basso reddito, con oltre 4 componenti in media, ad avere una spesa elevata in valore assoluto: 3.183 euro in media le prime e 2.848 le seconde. Questi due gruppi sociali presentano profili di spesa molto simili tra loro, anche a causa di una distribuzione territoriale con ampie aree di sovrapposizione. La spesa sanitaria, largamente incompressibile, è molto legata all'età dei componenti della famiglia, la quota di questa spesa assume infatti i valori più alti nei gruppi sociali in cui la presenza di anziani è maggiore, arrivando a pesare per il 5,9% tra gli operai in pensione, il 5,1% tra i pensionati d'argento e il 4,6% tra le anziane sole e i giovani disoccupati. Un'età media più giovane determina quote più elevate di spesa per abbigliamento e calzature: le famiglie di impiegati spendono il 5,7% del budget per questa voce, mentre le famiglie di anziani hanno quote più basse (il 3,6% tra le famiglie degli operai in pensione e tra le anziane sole, con una spesa media di 73 e 55 euro). Quanto detto trova eccezione tra i pensionati d'argento, le cui migliori condizioni economiche e il cui stile di vita fa sì che tale quota arrivi al 4,2% del totale (130 euro mensili). Anche la spesa per servizi ricreativi, ricettivi e di ristorazione è legata alla fase di vita familiare, così come al diverso livello di istruzione dei componenti delle famiglie. Le quote più elevate di spesa, e di valori assoluti, si osservano, oltre che nella classe dirigente, tra le famiglie degli impiegati (390 euro, il 12,8% del totale della spesa), i giovani blue-collar, pur disponendo di minori possibilità economiche rispetto agli impiegati, riservano a queste voci ben l'11,5%. I gruppi sociali caratterizzati da maggiore mobilità, legata sia all'attività lavorativa che a quella formativa, presentano anche quote di spesa per trasporti elevate: superano la quota del 13% si trovano le famiglie a basso reddito di soli italiani e le famiglie della provincia, entrambi gruppi giovani, numerosi, e con una maggiore incidenza di occupati.

**Tabella 1** - Spesa media mensile delle famiglie per gruppo sociale e ripartizione - Anno 2015 (valori in euro e composizione percentuale).

GRUPPI SOCIALI	Spesa media mensile (100per cento)	Capitolo di spesa (% sul totale di spesa senza beni e servizi essenziali)*							
		Alimentari	Abbigliamento	Abitazione, elettricità, gas	Salute	Trasporti	Spettacoli e cultura	Alberghi e ristoranti	Altri capitoli <sup>o</sup>
Famiglie a basso reddito con stranieri	1.697	21,0	5,2	36,8	2,8	11,3	3,3	3,5	16,1
Famiglie a basso reddito di soli italiani	2.848	20,9	5,2	29,6	4,1	13,1	5,9	4,4	16,9
Famiglie tradizionali della provincia	3.183	20,9	5,3	29,7	4,1	13,2	5,1	4,4	17,3
Anziane sole e giovani disoccupati	1.700	19,1	3,3	43,4	4,6	7,0	3,1	2,9	16,7
Famiglie degli operai in pensione	2.021	19,6	3,6	40,5	5,9	8,1	3,8	3,1	15,4
Giovani blue-collar	2.396	16,6	4,6	33,3	3,7	12,5	5,6	5,9	17,7
Famiglie di impiegati	3.011	15,6	5,7	32,8	3,9	12,3	6,1	6,7	16,8
Pensioni d'argento	3.098	16,2	4,2	37,8	5,1	10,2	5,2	5,2	15,9
Classe dirigente	3.810	13,6	5,0	37,2	4,0	10,3	6,6	6,4	16,9
<b>Totale famiglie</b>	<b>2.499</b>	<b>17,7</b>	<b>4,6</b>	<b>36,1</b>	<b>4,5</b>	<b>10,6</b>	<b>5,1</b>	<b>4,9</b>	<b>16,6</b>

\* I nomi esatti dei capitoli di spesa presenti in testata, semplificati per semplicità di lettura, sono: prodotti alimentari e bevande analcoliche; abbigliamento e calzature; abitazione, acqua, elettricità, gas e altri combustibili; servizi sanitari e spese per la salute; trasporti; ricreazione, spettacoli e cultura; servizi ricettivi e di ristorazione.

<sup>o</sup> Sono compresi i capitoli: bevande alcoliche e tabacchi; mobili, articoli e servizi per la casa; comunicazioni; istruzione; altri beni e servizi

All'opposto, i gruppi che hanno minori necessità di spostamento: gli operai in pensione e le anziane sole e i giovani disoccupati. Rispetto alle diverse necessità di mobilità legate al ciclo di vita, fanno eccezione nuovamente le famiglie dei pensionati d'argento, il cui principale percettore è uscito dal mercato del lavoro, ma dinamico e propenso alla partecipazione culturale: presentano infatti valori superiori alla media di spese dedicate a servizi ricreativi, spettacoli e cultura e a servizi ricettivi e di ristorazione (10,4% in totale contro il 10,0% di media nazionale).

#### 4. Le scelte allocative scorporando i consumi indifferibili

Posto che nelle scelte di consumo alcune spese sono indifferibili per conseguire uno standard di vita minimamente accettabile; altre, invece, possono considerarsi maggiormente imputabili a scelte soggettive nell'allocazione delle risorse economiche. Per scorporare le spese indifferibili, in questo lavoro si utilizza la metodologia del paniere di povertà assoluta (Istat, 2009). Questo rappresenta proprio l'insieme dei beni e servizi che vengono considerati essenziali per una determinata famiglia per conseguire uno standard di vita minimamente accettabile. Nello scorporo delle spese, la componente alimentare è stata decurtata dal capitolo alimentare, la componente abitativa dal capitolo abitazione e la componente residuale da tutti gli altri capitoli, in quota parte proporzionale ai livelli dei capitoli stessi. Calcolando la soglia di povertà assoluta media all'interno dei diversi gruppi sociali, questa raggiunge il valore massimo in termini assoluti per le famiglie tradizionali della provincia italiana (1.307 euro), che sono quelle con più componenti, in media, ed il valore minimo per il gruppo delle anziane sole e giovani disoccupati (701 euro) che, all'opposto, è il gruppo con famiglie di minore numerosità, in media. In termini relativi, il peso percentuale segue esattamente la classificazione fatta sui dati di reddito: pesa infatti il 48,8% del totale della spesa per il gruppo delle famiglie con stranieri a basso reddito, il 42% circa per famiglie a basso reddito di soli italiani, famiglie tradizionali della provincia e anziane sole e giovani disoccupati e assume valori via via decrescenti fino al 33% delle pensioni d'argento e al 27% della classe dirigente. Complessivamente, rispetto al totale delle spese al netto dei beni e servizi essenziali, si ridimensiona in maniera consistente il peso dell'abitazione e, soprattutto, dei beni alimentari. Entrambi i capitoli di spesa, peraltro, perdono in larga misura la regolarità empirica della loro riduzione relativa al crescere della spesa complessiva; infatti, per i capitoli considerati i valori più bassi si registrano per i due gruppi a minore benessere economico, le famiglie di stranieri a basso reddito (rispettivamente, 3,3% e 25,8%) e le famiglie a basso reddito di soli italiani (4,1% e 24,7%). Sono, inoltre, le famiglie più anziane a spendere di più in beni alimentari (con un massimo di 10,9% per le famiglie degli operai in pensione). Si rileva un aumento relativo del peso delle spese per abbigliamento, particolarmente importante per le famiglie a minore benessere economico, in parte per la maggiore numerosità media di queste e in parte per un'età media dei componenti più bassa. Rispetto a questo ultimo aspetto, per i gruppi più giovani aumenta il peso relativo delle spese per comunicazione. Infine, le spese per trasporti sono più rilevanti in termini relativi tra i gruppi ad alta intensità di lavoro e tra i gruppi a minore benessere economico. Per la cultura e per le spese per servizi ricreativi sono le famiglie di impiegati a dedicare la maggiore quota di spesa (16,8% del totale), con valori elevati, sopra il 15%, anche

per la classe dirigente, i giovani blue-collar e le famiglie a basso reddito di soli italiani. I valori più bassi si trovano in due gruppi di anziani a medio benessere economico, le anziane sole e i giovani disoccupati (8,7%) e le famiglie degli operai in pensione (9,5%). Tra le famiglie del gruppo delle pensioni d'argento la spesa per cultura e servizi ricreativi si assesta al 13,3% del totale del budget di spesa, nonostante, per età del maggiore percettore di reddito, queste famiglie dovrebbero assumere comportamenti più simili a quelli delle anziane sole e degli operai in pensione.

**Tabella 2** – Spesa media mensile delle famiglie per gruppo sociale e ripartizione senza beni e servizi essenziali - Anno 2015 (valori in euro e composizione percentuale).

GRUPPI SOCIALI	Spesa senza essenziali in euro (100 per cento)	Percentuale sul totale della spesa (%)	Capitolo di spesa (% sul totale di spesa senza beni e servizi essenziali)*							
			Alimentari	Abbigliamento	Abitazione, elettricità, gas	Salute	Trasporti	Spettacoli e cultura	Alberghi e ristoranti	Altri capitoli <sup>o</sup>
Famiglie a basso reddito con stranieri	869	51,2	3,3	9,4	25,8	4,5	19,2	5,5	6,5	25,7
Famiglie a basso reddito di soli italiani	1.653	58,0	4,1	7,3	24,7	5,9	19,2	8,6	6,9	23,3
Famiglie tradizionali della provincia	1.876	58,9	5,7	7,3	26	5,7	18,4	7,3	6,4	23
Anziane sole e giovani disoccupati	999	58,8	8,5	4,5	40,0	6,0	10	4,3	4,4	22,2
Famiglie degli operai in pensione	1.225	60,6	10,9	5,0	35,6	7,7	11,1	5,1	4,4	20,2
Giovani blue-collar	1.487	62,1	5,6	6,5	26,4	5,1	16,9	7,7	8,1	23,7
Famiglie di impiegati	1.910	63,4	4,4	7,6	28,1	5,2	16,2	8,1	8,7	21,8
Pensioni d'argento	2.076	67,0	8,4	5,2	34,4	6,3	13,1	6,6	6,7	19,5
Classe dirigente	2.783	73,0	6,2	6,1	34,5	4,9	12,3	7,8	7,8	20,5
<b>Totale famiglie</b>	<b>1.577</b>	<b>63,1</b>	<b>6,9</b>	<b>6,3</b>	<b>31,5</b>	<b>6,0</b>	<b>14,3</b>	<b>6,8</b>	<b>6,8</b>	<b>20,5</b>

\* I nomi esatti dei capitoli di spesa presenti in testata, semplificati per semplicità di lettura, sono: prodotti alimentari e bevande analcoliche; abbigliamento e calzature; abitazione, acqua, elettricità, gas e altri combustibili; servizi sanitari e spese per la salute; trasporti; ricreazione, spettacoli e cultura; servizi ricettivi e di ristorazione.

<sup>o</sup> Sono compresi i capitoli: bevande alcoliche e tabacchi; mobili, articoli e servizi per la casa; comunicazioni; istruzione; altri beni e servizi

## 5. Conclusioni

Nell'analisi delle scelte allocative per capitoli di spesa, tolte le spese che vengono considerate incomprimibili per mantenere un adeguato stile di vita, emerge come le famiglie a basso reddito con stranieri siano quelle che effettuano scelte di consumo più distribuite tra i capitoli di spesa considerati, superando quindi la media nazionale per trasporti, altri servizi, abbigliamento, comunicazioni, alcolici e tabacchi. A differenziare le spese sono anche le famiglie di impiegati che scelgono, più delle famiglie appartenenti ad altri gruppi, di collocare le proprie risorse su trasporti, alberghi e ristoranti, spettacoli, cultura e abbigliamento; di contro le famiglie della classe dirigente, che hanno livelli di spesa più alti si contraddistinguono per destinare la quota di spesa "accessoria" sull'abitazione più che su altri capitoli di spesa. Sembra dunque possibile affermare che le preferenze in termini di beni a cui destinare le quote "eccedenti" di spesa disponibile cambino sensibilmente sulla base del gruppo sociale di appartenenza. Tale lettura è possibile in virtù del fatto che questi gruppi sono stati costruiti non solo sulla componente reddituale, ma anche sulle dimensioni culturali e sociali che hanno evidenti effetti differenziali nelle scelte allocative.

In particolar modo risultano caratterizzanti le spese per spettacoli, cultura, alberghi e ristoranti; spese che possono, in qualche modo, sottolineare la componente culturale, ludica e di partecipazione culturale delle famiglie. A destinare la quota maggioritaria di spesa "accessoria" a questi capitoli sono le famiglie di impiegati, che pur non avendo un reddito particolarmente elevato, effettuano questo tipo di consumi più di altre famiglie, anche con redditi e titoli di studio più elevati; ciò è interpretabile in funzione della giovane età delle persone di riferimento delle famiglie in questo gruppo, della condizione occupazionale, che vede tutti occupati e anche della fase del ciclo di vita, per cui le famiglie appartenenti a questo gruppo sono più spesso coppie senza figli.

A seguire, sempre considerando questo tipo di consumi, sono le famiglie appartenenti ai giovani blue-collar, che condividono molte caratteristiche socio demografiche con le famiglie di impiegati, ma hanno situazione reddituale e occupazionale peggiore, compensata però da una quota più alta di titoli di studio elevati. Nelle caratterizzazione dei consumi culturali e sociali spicca il gruppo della classe dirigente, che condivide marginalmente le caratteristiche socio demografiche dei gruppi più giovani, ma comprendendo famiglie in cui la persona di riferimento è almeno laureata, indirizza i consumi verso queste forme di partecipazione culturale.

**Riferimenti bibliografici**

- BOURDIEU P. 1984. *Distinction: A social critique of the judgement of taste*. Harvard University Press.
- KRUGMAN P. 2012. *End this Depression Now!* New York: W. W. Norton & Co.
- ISTAT. 2017. *Rapporto Annuale sulla situazione del Paese 2017*. Roma: Istat.
- ISTAT. 2009. *La misura della povertà assoluta*. Collana Metodi e Norme n.39. Roma: Istat.
- SCHIZZEROTTO A. 1988. *Classi sociali e società contemporanea*. Milano: Franco Angeli.

**SUMMARY****Consumption choices by social groups**

This article analyzes expenditure choices for consumption by social groups. Income plays an important role in the definition of expenditure choices, however in this work we add other household characteristics deriving from the construction of Istat social groups (2017), that fit with the idea of personal choices also based on preferences. Analyzing social groups consumption by subtracting the so called “essential quota” (the necessary quota in order to reach acceptable living conditions), it has been possible to describe the real choices in the way households decide to spend their money in the different expenditure categories. Age, occupational conditions, education and household size have large effect on expenditure choices for consumption. Overall, in recreational and culture expenditure choices.



## **ENVIRONMENTAL EDUCATION: A WEB MINING ANALYSIS**

Domenica Fioredistella Iezzi, Francesco Zarelli

### **1. Introduction**

According to the Belgrade Charter (1975), “the goal of the environmental education is to develop a world population that is aware of, and concerned about, the environment and its associated problems, and which has the knowledge, skills, attitudes, motivations and commitment to work individually and collectively toward solutions to current problems, and the prevention on the new ones”.

EPA (Environmental Protection Agency) of United States clarifies that “environmental education is a process that allows individuals to explore environmental issues, engage in problem solving, and take action to improve the environment. As a result, individuals develop a deeper understanding of environmental issues and have the skills to make informed and responsible decisions”<sup>1</sup>. In recent years, attention of government about environmental is growing, and also demand from the green jobs market has increased.

In this paper, we analyse the educational projects proposed from the Italian National parks for schools: common themes of theoretical and practical lessons, multimedia audio-visual equipment, themes of working group of students.

We use text clustering and social network analysis conjointly to classify the content of the educational project of parks. The paper is structured as follows: in section 2, we describe the corpus and the methods; in section 3, we illustrate the main results, and in section 4, we delineate the conclusions and the future steps.

### **2. Educational projects in the Italian parks**

We build a corpus composed of recent environmental education projects retrieved from [www.parks.it](http://www.parks.it) web site. These projects explanations regarding 15 Italian natural marine/mountain protected areas, for the most part belonging to the year 2016.. In Table 1, we show some lexical measures of the corpus. We can see

---

<sup>1</sup> See the website <https://www.epa.gov/education/what-environmental-education>.

that The projects have very different patterns: some extremely articulated (composed on average of about 1200 token, e.g. Cinque Terre Park, on very synthetic projects with 123 token (Capo Carbonara park).

In the first step of text mining process, we retrieved from the website park, the projects, and apply exploratory methods to discover some topics set in the environmental teaching in the parks (Iezzi et al., 2014; Zarelli, 2016).

**Table 1** – *Natural areas and respective number of environmental education projects, total occurrences and number of graphical forms.*

Protected Area	Number of texts (projects)	Occurrences	Number of forms
N. P. Abruzzo, Lazio e Molise	13	5,653	1,069
N. P. Majella	76	8,673	1,640
N. P. Appennino Tosco – Emiliano	17	10,693	1,993
N. P. Arcipelago Toscano	4	1,498	541
N. P. Cilento, Vallo di Diano e Alburni	177	18,793	2,178
N. P. Cinque Terre	28	34,274	4,149
N. P. Circeo	3	633	295
N. P. Dolomiti Bellunesi	39	3,106	961
N. P. F. Casentinesi M. Falterona Campigna	32	5,916	1,414
N. P. Maddalena	7	1,032	427
N. P. Valgrande	31	4,200	1,151
R. P. Migliarino San Rossore	72	9,500	1,973
M. P. A. Capo Carbonara	3	323	191
M. P. A. Egadi	3	1,087	338
M. P. A. Miramare	75	12,725	1,812
M. P. A. Torre Cerrano	11	1,212	451
Total	591		

*Note N. P. = National Park; R. P. = Regional Park; M. P. A. = Marine Protected Area.*

To explore the corpus, we apply several methods from French 'Lexicometrie', and French 'Data Analysis' schools. It includes specificity analysis, multi-dimensional exploration (Factorial analysis), and text clustering methods. We use the software IRAMUTEQ (Loubère et al, 2013) to classify contents of the projects, based on principal components obtained from a lexical correspondence analysis (forms x natural protected areas). Moreover, to map the most relevant concepts we have used several dichotomous illustrative variables: pre-primary school, primary education, lower secondary school, upper secondary school, teachers and groups.

To improve the interpretation of the data we consider the dendrogram by ALCESTE method (Reinert, 2003) on forms with decreasing rankings by chi square. We consider the graph of every cluster using Similarities Analysis too, by Fruchterman – Reingold algorithm based on co-occurrence (Fruchterman et al.,



Figure 2 – First factorial plan. Axes 1 and 2 on lexical correspondence analysis on the active and the illustrative variables.

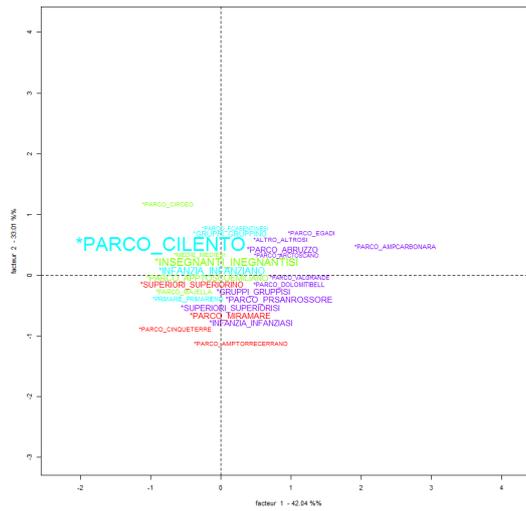
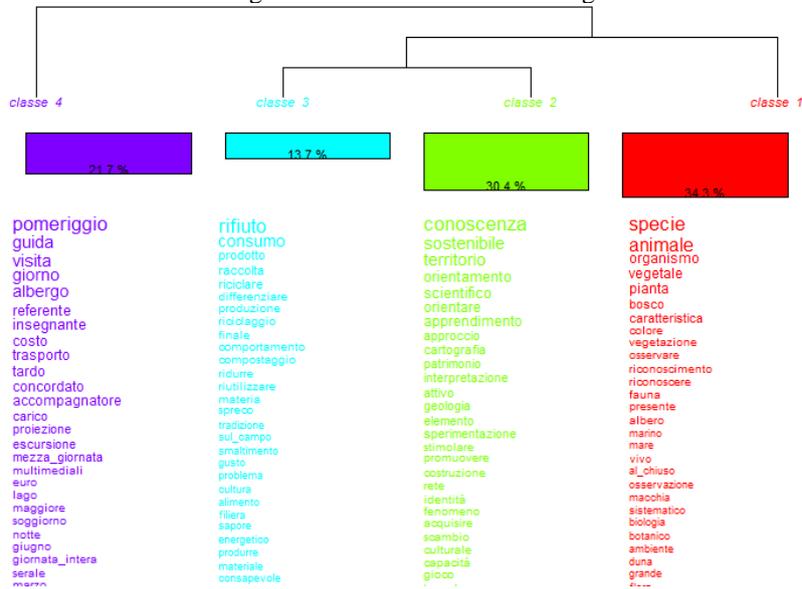


Figure 3 – Hierarchical clustering ALCESTE method – Dendrograms.

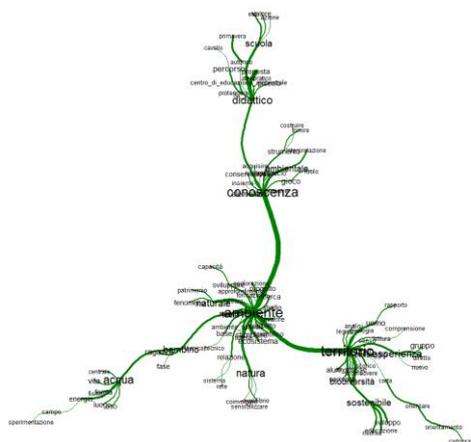




segments, we can find to discover, didactic trip hike, guided tour, outdoor, biodiversity, magnification, principal features, classroom session encounter, team effort, didactic sheet.

The cluster number two (green) concerns students of secondary school, teachers and groups. It includes 735 segments, 30.4% of examined total. The graphic forms regard in particular outdoor activities and learning to express oneself through direct experience, self-evaluation in places that hold many surprises, moving with confidence within the wildlife. In addition, the development of competence in the teaching environmental education, participation and evaluation methodologies and techniques.

**Figure 5** – Fruchterman-Reingold graph for the cluster 2.

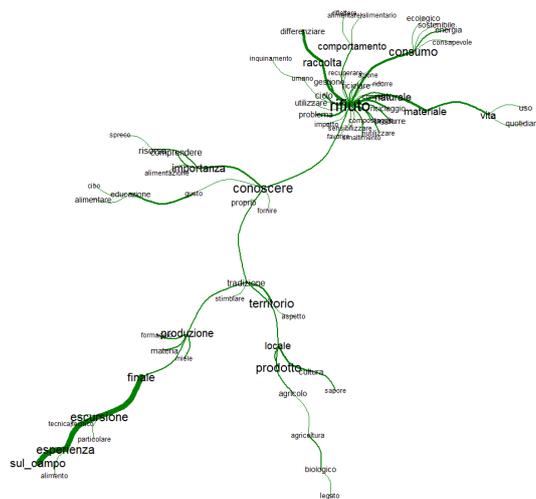


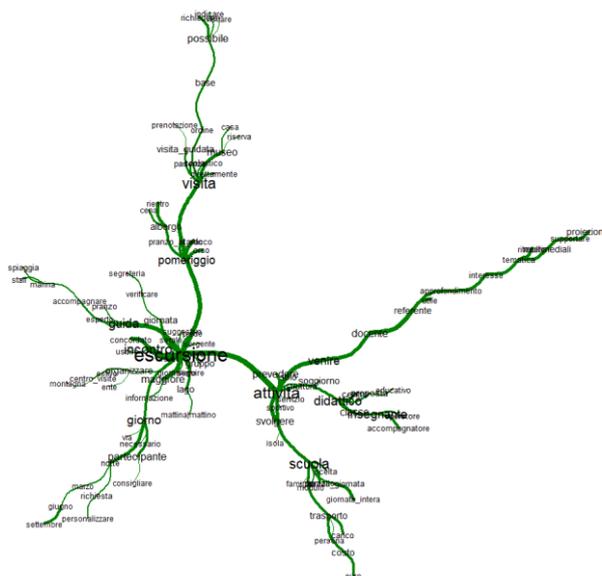
The parks that characterize this cluster are Appennino Tosco-Emiliano, Majella and Circeo. Among the specific graphic shapes of such group, we notice: orienteering, cartography, interpretation, sustainable, to build, to test, identity, compass, creativity, to experience, environmental educational centre. The repeated segments are final hike, final fieldwork and final practical feasibility, sustainable development, throughout the year, sustainable tourism, environmental interpretation, detailed study and sense of belonging. We can present among the text segments (Elementary Context Units E.C.U.) that are typical of the cluster, the following examples: “discovering natural and historical values, diversity and identity, sense of belonging to the territory where the same students live. The stays also propose, through the day-to-day practice in the natural life, experiences and human life in the nature, stiles based on sustainability and the respect.” “The method of adventure of exploring facilitates the discovery, the synesthetic

approach of knowledge, the entirety of languages method, data-gathering and information; communication techniques related to environmental interpretation; laboratories of sensory analysis.”

The projects in cluster 3 (sky) concern in particular students of lower secondary school. It includes 330 segments, 13.65% of examined total. Educational activities based on direct experiences to develop a sustainable culture. The training aims to favour the knowledge of waste management, renewable energies, and sustainable nutrition, fascinating them with new discoveries, to recognize opportunities from the problems. Among the specific graphic forms of this group, emerge waste, on-site, resource, to diversify, to reuse, agricultural, food, nutritional, compost, renewable, cycle. Among the repeated segments, we can mention understanding the relevance, techniques, production of cheese wheel, honey, local traditions, tastes and culture, sustainable nutrition, production consume and waste, separate collection, biodiversity, natural cycles, knowing the territory. Among the text segments (Elementary Context Units E.C.U.) characteristic of the cluster ”learning from the nature to recycle the waste”, “the motivation to recycle, to reuse”.

**Figure 5** – Fruchterman-Reingold graph for the cluster 3.



**Figure 6** – Fruchterman-Reingold graph for the cluster 4.

In the cluster number four (violet), it includes 524 segments, 21.67% of examined total. It consider all the level of school and concerns the organization of the educational activities, times, logistics, administration. Among the specific graphic forms of such group we can find All-day, afternoon, visit, stay, hotel, guide, teachers, personalize, multimedia, showing, transport, cost, euro, administrative office. Different activities are free, for example the Region; the Public Administrations pay the costs. The names of didactic farms, environmental organizations, and environmental educational cooperatives are often in this group. It is important the collaboration of the parks staff.

#### 4. Conclusions

The environmental education works towards multidisciplinary experiences on-site. It mainly points towards students from resident communities. It is a way to construct a common “ecological alphabet and language” (Bonifanti, Frabboni, Guerra, Sorlini, 1993). Families may be involved in the process, accompanying the children in the farms, museums, park visitor’s centres; these kind of arrivals, in particular toward internal natural areas, could be another important opportunity for sustainable tourism. The outdoor activities often proposed throughout the year,

contribute to consider the possibility of seasonally adjusted visits, admiring natural beauties in different periods. Like Ministry for the Environment highlights, the real key actors are Regions, local Administrations and different Organizations (Ministry for the Environment, 2013). Indeed the data confirm the strong cooperation between parks, environmental organisations, school, museums, University and local Administrations. Moreover, it emerges clearly the synergy between public administrations and farms, cooperatives in different educational areas of expertise: green economy signals, they probably might be even greater if students from non-resident communities were included in the environmental projects. The analysis seems to confirm the principles evolved in the time: the environmental education is sustainable development education. In particular, the environmental training seems to respect the following distinctive eight educational itineraries proposed in the Environmental Education Guidelines of Italian Ministry of the Environment and Ministry of Education, Universities and Research in 2014.

They are: *Safeguard of water and the sea* - pre-primary and primary education; *Safeguard of biodiversity: Flora and Fauna* - pre-primary school and primary education; *Sustainable nutrition* pre-primary school, primary education, and secondary school; *Waste management* - pre-primary school, primary education and lower secondary school; *Safeguard of biodiversity: ecosystem* - secondary school; *Green economy: green jobs and green talent* - upper secondary school; *Sustainable town: pollution, soil consumption, and waste* - upper secondary school; *Climatic changing adjustment* - upper secondary school. Moreover, children learn to know the natural and socio-economic history of their territory, asking questions about the environmental problems, identifying possible sustainable solutions. In this way, students become now awareness of their personal irreplaceable role to protect the life, the biodiversity.

## References

- BARIL, E. and GARNIER, B. IRaMuteQ 0.7 alpha 2 Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. [http://www.iramuteq.org/documentation/fichiers/Pas%20a%20Pas%20IRAMUTEQ\\_0.7alpha2.pdf](http://www.iramuteq.org/documentation/fichiers/Pas%20a%20Pas%20IRAMUTEQ_0.7alpha2.pdf).
- BONIFANTI, P., FRABBONI, F., GUERRA, L., SORLINI C. 1993. Manuale di educazione ambientale. Roma-Bari. Laterza.
- BRAGGION, T.V., CHELIDONIO, G. 2005. *Esplorare per conoscere*. Roma: Carocci Faber.
- FRUCHTERMAN, T. M.J. and REINGOLD E. M. 1991. Graph Drawing by Force-directed Placement. SOFTWARE—PRACTICE AND EXPERIENCE.

- VOL. 21(1 1). 1129-1164.  
<http://citeseer.ist.psu.edu/viewdoc/download?doi=10.1.1.13.8444&rep=rep1&type=pdf>.
- IEZZI D.F. 2012. A New Method For Adapting The K-Means Algorithm To Text Mining. *Statistica Applicata -The Italian Journal of Applied Statistics*. 22(1). 65-76. [http://sa-ijas.stat.unipd.it/sites/sa-ijas.stat.unipd.it/files/22\(1\)\\_Iezzi.pdf](http://sa-ijas.stat.unipd.it/sites/sa-ijas.stat.unipd.it/files/22(1)_Iezzi.pdf).
- IEZZI D. F. 2015. Il web content: una nuova sfida per la Statistica sociale. In A. Mussino, editor, *Non tutto, ma di tutto sulla Statistica sociale*. Edizioni Nuova Cultura. Roma: 167-184.
- IEZZI D.F. MASTRANGELO M. 2014. Fuzzy c-means for web mining: The Italian tourist forum case. In VICARI D., OKADA A., RAGOZINI G., WEIHS C. (Eds). *Analysis and Modeling of Complex Data in Behavioral and Social Sciences*. Springer. Berlin. 153-160.
- IEZZI D. F., ZARELLI, F. 2015. What tourists say about the Italian national parks: a web mining analysis. *Rivista Italiana di Economia Demografia e Statistica*. LXIX(3).73-82.
- LOUBÈRE, L. and RATINAUD, P. 2013. Documentation IRaMuTeQ 0.6 alpha 3 version 0.1. [http://www.iramuteq.org/documentation/fichiers/documentation\\_iramuteq\\_21\\_12\\_2013.pdf](http://www.iramuteq.org/documentation/fichiers/documentation_iramuteq_21_12_2013.pdf).
- MARCHAND P. and RATINAUD P. 2012. L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). In *Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles*. JADT 2012. Liège. 687-699. <http://lexicometrica.univ-paris3.fr/jadt/jadt2012/Communications/Marchand,%20Pascal%20et%20al.%20-%20L'analyse%20de%20similitude%20appliquee%20aux%20corpus%20textuels.pdf>.
- MINISTRY OF THE ENVIRONMENT AND MINISTRY OF EDUCATION, UNIVERSITIES AND RESEARCH. 2014. Environmental education guidelines. [http://www.minambiente.it/sites/default/files/archivio/allegati/LINEE\\_GUIDA.pdf](http://www.minambiente.it/sites/default/files/archivio/allegati/LINEE_GUIDA.pdf)
- MINISTRY FOR THE ENVIRONMENT, LAND AND SEA. 2013. Sixth National communication under the UN Framework convention on climate change. [https://unfccc.int/files/national\\_reports/annex\\_i\\_natcom/submitted\\_natcom/application/pdf/ita\\_nc6\\_resubmission.pdf](https://unfccc.int/files/national_reports/annex_i_natcom/submitted_natcom/application/pdf/ita_nc6_resubmission.pdf)
- MINISTERO DELL'AMBIENTE, MINISTERO DELLA PUBBLICA ISTRUZIONE. 2000. "L'educazione in cammino, contributi dei Gruppi di Lavoro Istruttori per la Conferenza Nazionale di Educazione Ambientale.
- REINERT M. 2003. Le rôle de la répétition dans la représentation du sens et son approche statistique dans la méthode Alceste (The function of repetition in the

- representation of meaning and its statistical approach in the Alceste method). *Semiotica*. 147 (1-4). 389-420.
- SMITH P. R. 1987. "Outdoor Education and its Educational Objectives", *Geography*. Vol. 72 No. 3
- ZARELLI F. 2016. I parchi nazionali in Italia nel periodo della crisi. Doctoral thesis. Department of Social and Economic Sciences. Sapienza University of Rome. XXVII cycle. Tutor Iezzi, F.D. <http://hdl.handle.net/11573/926657>

## **SUMMARY**

### **ENVIRONMENTAL EDUCATION: A WEB MINING ANALYSIS**

Outdoor Education and its Educational Objectives points towards encouragements of the students discovering their personal contribution to save the biodiversity. Big data offer the opportunity to analyse the different features of the environmental projects. Through text mining we study the projects of natural areas for the schools, trying to note the differences and similarities.



## THE EQUITABLE AND SUSTAINABLE WELL-BEING AT LOCAL LEVEL: A FIRST ATTEMPT OF TIME SERIES AGGREGATION<sup>1</sup>

Mariateresa Ciommi, Chiara Gigliarano, Stefania Taralli, Francesco M. Chelli

### 1. Introduction

The Istat project aimed at measuring well-being at local level<sup>2</sup> has recently made available time series data on a collection of BES indicators also at NUTS3 level. This recent novelty provides the motivation for studying appropriate methodologies to analyze such complex information. In particular, synthesizing individual indicators into a single index, especially if in a time perspective, by constructing a composite indicator, is an important challenge (Nardo, 2008). Indeed, the choice of the aggregation method is very important and delicate, as the composite indicator is required to be sensitive to territorial disparities, able of capturing transformations over time and, at the same time, robust with respect of the variability and over-dispersion that characterize many BES indicators at NUTS3 level (Taralli *et al.*, 2015)

In this paper, we implement a latent variable (factor) model with the aim to construct a well-being composite indicator for the 110 Italian Provinces based on a selection of eight BES indicators over the period of time between year 2004 and year 2015.

Factor models are widely used in the construction of composite indicators. For instance, Chelli *et al.* (2015) applied a factor analysis to reduce the high number of indicators involved in the definition of the BES at local level. Based on the result of the factor analysis computed among indicators of each domain, the authors then

---

<sup>1</sup> Disclaimer: This paper is a collective work, realized under the collaboration agreement between Istat and Polytechnic University of Marche for the analysis of the well-being at territorial level. In any case, the opinions expressed do not necessarily reflect those of the Italian National Institute of Statistics.

<sup>2</sup> The data used in this paper were produced by Istat under the project “Equitable and Sustainable Well-being Measures at local level”, started as a continuation of the “Provinces’ BES” project, that was promoted by CUSPI (Coordination of Statistical Offices of the Italian Provinces) and realized under the Istat’s methodological and technical coordination.

constructed a composite indicator as linear combination of the estimated factor scores with weights based on the Gini index of concentration.

The rest of the paper is organized as follow. Section 2 is devoted to the description of the data used in our analysis. Section 3 describes the methodology. Section 4 and Section 5 are devoted to the discussion of the results and conclusions, respectively.

## 2. Data

This study is based on the data extracted from the 2016 release of Istat's Provinces' BES dataset, which contains time series of all the indicators included in the Provinces' BES project. Following Rijpma (2017) we have chosen one leading indicator for each of the eight well-being domains to be considered in our analysis. Table 1 shows the selected indicators and the related domains.

**Table 1** – *The selected measures of equitable and sustainable well-being (Bes indicators).*

Domain	Indicator	Code	Unit
Work	Non-participation rate (15-74)	LAB	Percentage
Education	People with lower secondary education or less (25-64)	EDU	Percentage
Health	Life expectancy at birth	LE	Years
Economic well-being	Household disposable income	INC	Euro
Politics and Institutions	Young people in local government ( $\leq 39$ )	POL	Percentage
Security	Homicide rate	HOM	For 100,000 inhabitants
Quality of services	Separate collection of municipal waste	WAST	Percentage
Environment	Air pollution	PM10	Total days per year

*Source: Provinces' BES dataset.*

For a detailed description of the indicators involved in our analysis, see the Appendix of this paper.

The dataset contains 110 observations corresponding to the 110 Italian Provinces listed in the NUTS 2013 classification<sup>3</sup> and a complete time series over

<sup>3</sup> The current NUTS 2013 classification is valid from 1 January 2015, and for Italy at the NUTS3 level it includes 110 territorial units, coinciding with the 110 provinces that existed in Italy at the reference date. During 2016, following the reform of the Local Authorities implemented by the Italian Government, some Provinces have become Metropolitan cities, while some others have been suppressed or modified due to regional laws. Since these changes have not yet been transposed into the statistical classification, in this paper, the term Provinces refers to the 110 units accounted in NUTS3, so including the new Metropolitan cities and the Provinces that no longer exist.

the years between 2004 and 2015 for each of the eight indicators. Missing data are about 2.5% of total observations. This percentage is mostly due to the new Provinces that were established in the year 2010.

**Table 2 – Summary statistics (2004-2015)**

Indicator	Mean	Standard Deviation	Min	Max
Per capita GDP	24339	6419	12800	51200
Non-participation rate (15-74)	17.89	10.83	3.70	48.45
People with lower secondary education or less (25-64)	47.00	7.73	27.30	69.34
Life expectancy at birth	81.58	0.92	78.40	84.10
Household disposable income	39016	6909	25132	59609
Young people in local government ( $\leq 39$ )	30.85	5.28	5.00	58.00
Homicide rate	0.99	2.56	0.00	83.14
Separate collection of municipal waste	34.18	20.14	0.00	87.40
Air pollution	58.05	45.93	-1	321.00

Source: Our elaboration on Eurostat database and Provinces' BES dataset.

A further indicator was included in the analysis, which is the Per capita Gross domestic product (PCGDP) at current market prices, measured in Purchasing Power Standard (PPS), provided by the Eurostat's Regional economic accounts<sup>4</sup>.

Table 2 reports summary statistics of the indicators considered over 2004-2015 period, while Figure 1 shows their trends.

Figure 1 clearly shows that the series have very different trends, which reflect the divergences among the territorial units, and the different nature of the phenomena.

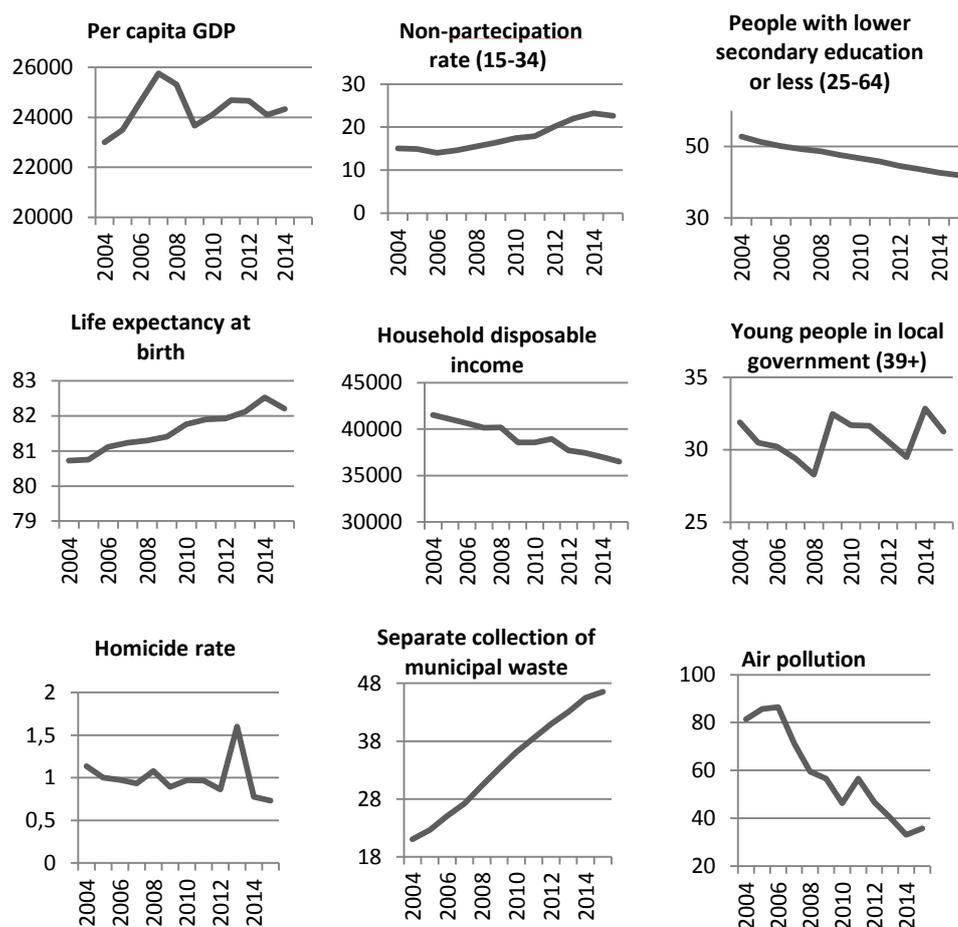
The indicator with the highest temporal variability, computed through the coefficient of variation (data not shown), is Homicide rate (CV=2.6), followed by Air pollution (CV=0.8), Separate collection of municipal waste (CV=0.6), and Non-participation rate (CV=0.6), while the remaining indicators show a CV below the threshold of 0.2.

Looking at the trends, we note a relevant improvement over time in the values of Separate collection of municipal waste, which passes from 21.0% in 2004 to 46.5% in the year 2015 (an overall increase of 121.4%). Time-variation is smaller, but still substantial for Air pollution, which shows an overall decrease of 55.9% and for Non-participation rate, with an increase over time of 49.7%. Generational replacement contributes to an overall reduction of 20.5 per cent points for the indicator People in working-age with lower secondary education or less. In the

<sup>4</sup>Available at: <http://ec.europa.eu/eurostat/web/regions/data/database>

same period the Household disposable income decreases by 12.1%, while the Per capita GDP increases by 9.1%.

**Figure 1** – Trends of the eight BES indicators and of the Per capita GDP - average of NUTS3 data (2004-2015).



Source: Our elaboration on Eurostat database and Provinces' BES dataset

Notice the outlier value of the Homicide rate in correspondence to the year 2013, due to the large number of migrants who died in a shipwreck near the Lampedusa Island and were considered as victims of a crime of massacre. Overall, the global Homicide rate marks a steady and consistent decrease (by 35.4%) over the period 2004-2015, while Life expectancy at birth has a low rate steady growth,

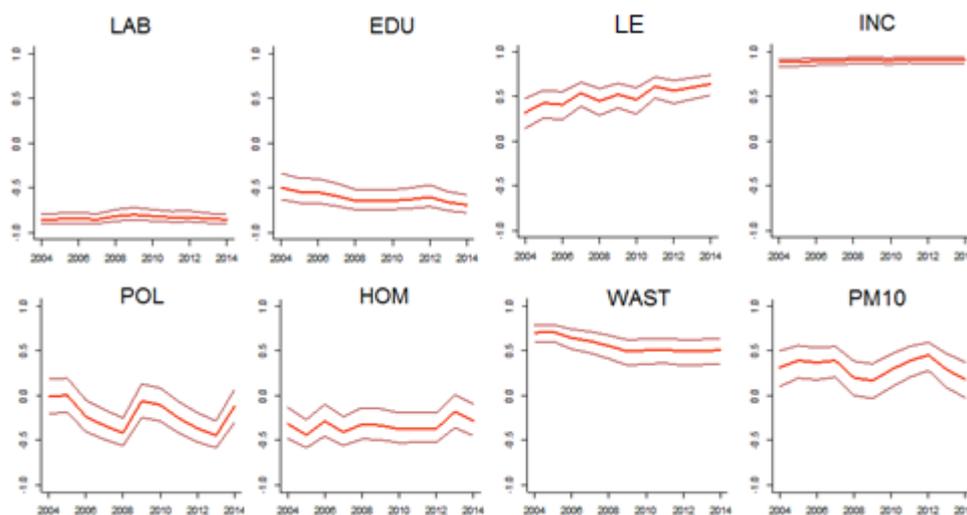
except for the 2015 breakdown due to exceptional number of deaths in that year in Italy. The percentage of young people in local government shows a cyclical trend, with high peaks at the years of municipal elections. The underlying trend is moderate growth (from 32% in 2004 to 33% in 2015).

**Table 3** – Correlations among BES indicators (year 2015)

20015	LAB	EDU	LE	INC	POL	HOM	WAST	PM10
LAB	1							
EDU	0.61	1						
LE	-0.58	-0.45	1					
INC	-0.70	-0.51	0.63	1				
POL	0.27	0.15	0.12	-0.12	1			
HOM	0.32	0.14	-0.35	-0.32	0.05	1		
WAST	-0.56	-0.27	0.32	0.45	-0.20	-0.13	1	
PM10	0.10	0.04	-0.13	-0.05	-0.04	0.08	0.24	1

Source: Our elaboration on Eurostat database and Provinces' BES dataset

**Figure 2** – Cross-sectional correlation coefficients and 95 per cent confidence intervals between BES indicators and PCGDP.



Source: Our elaboration on Eurostat database and Provinces' BES dataset

Looking now at the correlations among the eight BES indicators (see Table 3) we note that the strongest positive correlations are between the domains Work and

Education (with correlation coefficient equal to 0.49 in 2014 and to 0.61 in 2015), while the domains Work and Quality of services show the strongest negative correlation coefficient both in 2004 (-0.75) and in 2015 (-0.56).

The correlation coefficients between each of the BES indicators and the Per capita GDP tend to remain constant over time, with 95% confidence intervals that are rather tight. The strongest correlations are between PCGDP and Household disposable income, as expected (it ranges from 0.85 in 2004 to 0.90 in 2013) (Figure 2).

### 3. Methodology

In the literature, several methods have been proposed for constructing composite indicators; for a review we refer, among others, to Nardo et al (2008).

Here we adopt a data-driven approach that implies that the aggregation step is essentially based on objective, rather than subjective, criteria.

To the best of our knowledge, this is the first attempt to construct a well-being composite indicator using data for Italian Provinces in a time series perspective (2004-2015). Here we estimate a composite indicator through a latent variable (factor) model (hereafter, LVM), following the methodology introduced in Rijpma (2016).

The main advantage of this model is that it allows for imputation of missing data. This is an important issue, especially when working with time series data and even more at territorial level. In fact, on one hand, with historical data missing values are very frequent, and in particular at a high territorial detail time-series can be affected by administrative change, but, on the other hand, the way in which imputation is done influences the results. According to LMV, data are imputed using the covariance among indicators.

Another advantage of LVM refers to the weighing procedure that is a necessary step to construct a composite indicator, based on multiple series. The LVM assigns higher weights to highly correlated indicators and, as stressed by Foster et al. (2013), this procedure increases the robustness of the rankings.

However, the method is not without disadvantages. Firstly, we know that the correlation between variables does not necessarily capture the right trade-offs between the indicators. In addition, the model assumes that the underlying well-being concept is reflected in the correlation structure, and not that each indicator measures a distinct and unique part of well-being. Concerning this, it has been argued that applying a reflective measurement model to Bes is a misspecification of the measurement model itself as the Bes is a formative construct (Istat, 2015). However, in the Bes the multicollinearity, which is a highly undesirable property in

formative models, is not fully solved, as several indicators show substantial correlations. Finally, the model assumes that the observed and latent variables are continuous and this is a further limitation.

We consider a factor model for the observed data  $y_{ij}$ , where  $i$  is the province and  $j$  the indicator. The model is defined as

$$y_{ij} \sim N(\beta_{0j} + \beta_{1j}\chi_i, \omega_j^2) \quad \text{and} \quad \chi_i \sim N(0,1)$$

where  $N$  denotes the normal distribution,  $\chi_i$  is the unobserved and latent variable for the province  $i$ ,  $\beta_{1j}$  is a parameter reflecting how well the observed indicator  $j$  differentiates between units (Provinces),  $\beta_{0j}$  is an intercept and  $\omega_j^2$  represents the variance of indicator  $j$ . The latent variable (which corresponds to the well-being composite indicator) is assumed to be standardized, with mean equal to zero and standard deviation equal to one. This is not a problem, since the composite indicator has no natural unit of measurement.

The model is estimated in a Bayesian framework. This approach presents several advantages, since it works very well with the multilevel structure of our data and, more important, it may account for similarities among provinces in the same region. Bayesian multilevel models take this information into account through the priors. Rather than considering the same standard normal distribution for all provinces ( $\chi_i \sim N(0,1)$ ), we may follow Høyland et al. (2012) and allow as a prior a distribution of the composite indicator that varies across the  $m$  macroareas (or macro-regions); formally

$$\chi_i \sim N(\alpha_{m[i]}, 1) \quad \text{and} \quad \alpha_{m[i]} \sim N(0,1).$$

The use of priors might introduce a degree of arbitrariness and subjectivity. In fact, prior distributions usually reflect the researcher's beliefs about the model, before looking at the distribution of the real data. Thus, to reduce this subjective belief we use uninformative priors, assuming that: (i) the loadings follow a normal distribution,  $N(0, 10^{-7})$ , (ii) the variance is modelled through a Gamma distribution,  $G(0.01, 0.01)$ , and (iii) the group-specific error term follows a normal distribution,  $N(0, \sigma^2)$  with variance uniformly,  $\sigma^2 \sim U(0, 100)$ .

The model is estimated through Gibbs sampling and implemented in JAGS<sup>5</sup>. The Bayesian multilevel latent variable model has two main advantages: first of all, the weights for the aggregation procedure, which are based on correlations, are chosen so that they differentiate between provinces as best as possible; moreover,

---

<sup>5</sup> JAGS is a program for analysis of Bayesian hierarchical models using Markov Chain Monte Carlo (MCMC) simulation. See Plummer (2003) for more details. We are grateful to Rijpma who kindly provided us with the code.

the model accounts for sources of uncertainty, such as measurement error in a composite indicator or missing data.

#### 4. Results

As a first result, we are interested in analysing the distributions of the factor loadings, synthesized in Table 4, which show how much each BES indicator contributes to the composite indicator.

**Table 4**– Summary of the factor loadings' distribution for each BES indicator

Indicator	Mean	5-th percentile	50-th percentile	95-th percentile
Non-participation rate (15-74)	-0.967	-0.981	-0.967	-0.952
People with lower secondary education or less (25-64)	0.829	0.801	0.829	0.856
Life expectancy at birth	0.282	0.235	0.282	0.327
Household disposable income	-0.445	-0.488	-0.445	-0.402
Young people in local government (39+)	0.331	0.276	0.332	0.385
Homicide rate	-0.286	-0.331	-0.286	-0.240
Separate collection of municipal waste	-0.225	-0.272	-0.225	-0.179
Air pollution	0.470	0.430	0.470	0.509

Source: Our elaboration on Provinces' BES dataset

The first two indicators listed in the table are the main contributors to the composite indicator. In particular, the Non-participation rate and the People with lower secondary education or less are the indicators that affect more the well-being composite indicator, the former with a strongly negative effect and the latter with a strong positive impact. Third, in decreasing order of importance, is the Air pollution, with a mean loading of 0.470, followed by the negative loading of the Household disposable income (-0.445 on average).

Looking at the temporal trend of the composite indicators by macro-area (Figure 3) Italy seems to be divided into two areas of well-being: on the one hand, the South and the Islands, with consistently negative composite indicator's values (that is below the Italian average), indicating a structural and persistent disadvantage; on the other hand, the North and the Centre, always taking positive values (above the Italian average).

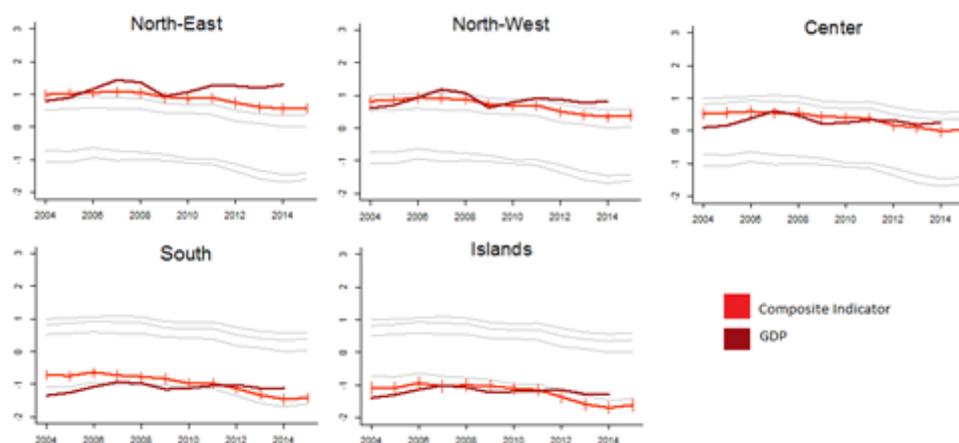
The highest levels of the composite index (red line) are registered, in particular, in the North-East of Italy (NE). The average value of the composite indicator of NE, indeed, remains above the ones of all the other Italian macro-areas (grey lines) during the whole period. In particular, until year 2008 the value of the composite indicator for the NE stays around an average of 1.0-1.1. Starting from 2008 a

slightly decreasing trend leads the BES indicator for NE to levels between 0.4 and 0.6 in the years 2014-15 (with an overall reduction of 44%).

The composite index for the North-West (NW) shows for each year values that are slightly lower than NE but with very similar trends: it remains stable around average values of 0.8-0.9 until 2008, followed by a gradual decrease, more pronounced than NE, from year 2008 to 2015 (with an overall decrease of 56.2%).

The composite indicator for the Center of Italy is fairly stable around an average value of 0.56 until the year 2008. Starting from 2008 it sharply declines, reaching the value of 0.03 in 2015, which corresponds to an overall decrease of 95% compared to 2008. It seems, therefore, that the 2008 economic crisis has affected the well-being of this area much harder than in the North of Italy. The loss for the Center, in relative terms, is similar to the one suffered by the Islands (-61.5%), which were already in a more disadvantageous condition. However, South is the most penalized and deprived macro-area in Italy: in 2014-2015 the well-being level in the South of Italy lies between -1.25 and -1.62; this corresponds to a negative variation, compared to year 2008, equal to -85%.

**Figure 3** – *Standardized Per capita GDP, estimates and 90% Confidence Interval of the BES composite indicator, by macro-area (years 2004-2015).*

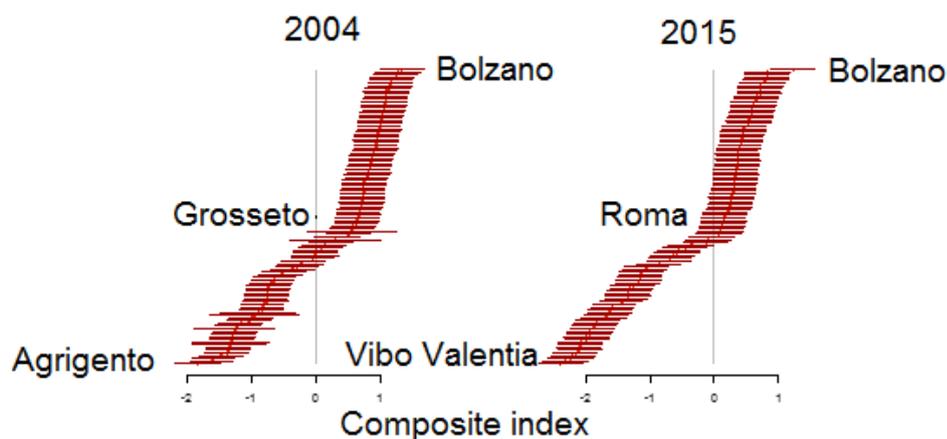


Source: Our elaboration on Eurostat database and Provinces' BES dataset

Figure 3 also compares the trends of the BES composite indicator (red line) and of the Per capita GDP (dark red line) for each macro-area. It can be noted that the two series are generally rather close. In NE, starting from the year 2009, there is a clear-cut separation between PCGDP, which tends to grow, and BES, which shows a decreasing trend. At different points in the time-line, but still after 2008, this reversal of the relationship between BES and PCGDP occurs in all areas of Italy.

Figure 4 depicts the estimates of the composite indicator for all Italian Province in 2004 and 2015. For each Province, the graphic displays the mean value of the composite indicator and its confidence interval at 90%. Looking at the value of the composite index we can evaluate, for each period, the trend and position of the BES levels for each province, with respect to the general mean (grey vertical bar, equal to 0). The form of the right-hand curve compared to the left-hand shows that over the last few years the distances between best and worst performers deepened both in general and compared to the average. It emerges that territorial disparities in 2005 affect mainly the lower part of the distribution, which corresponds to those provinces that were already disadvantaged in the years before the crisis. Bolzano shows the best performance, whereas, although with some exceptions, the South and the Island Provinces are ranked in the last positions, both in 2004 and 2005.

**Figure 4** – Composite indicator and its (90%) confidence interval for 2004 and 2015 for all Provinces.



Source: Our elaboration on Eurostat database and Provinces' BES dataset

## 5. Concluding Remarks

A latent variable model has been adopted in order to construct a composite indicator aimed at representing the trend of well-being for the Italian Provinces over the years 2004-2015. The composite indicator proposed may be considered a useful tool for summarizing the well-being trends, since it allows enhancing differences among macro-areas and their trends also during the years of the recent financial crisis.

Future research will consider (i) including additional well-being indicators, in order to cover all the well-being dimensions and (ii) aggregating the analysis also at regional level (and not only at macro-regional level) to provide a more in-depth analysis of territorial disparities, inequalities and divergences across the Italian territories.

## Appendix

Here we provide a detailed definition of the BES indicators selected for our analysis.

- Non-participation rate (15-74) is defined as the ratio: (*Unemployed persons aged 15-74 years plus inactive persons - who have not looked for a job in the past 4 weeks but willing to work*) / (*Labour force aged 15-74 years plus inactive persons*).
- People with lower secondary education or less (25-64): *Percentage of people aged 25-64 years who have completed at most lower secondary education (ISCED level not above 3a, 3b or 3c).*
- Life expectancy at birth: *Average number of years that a child born in a given calendar year can expect to live if exposed during his whole life to the risks of death observed in the same year at different ages.*
- Household disposable income: *Ratio of the annual household disposable income over the total number of household members.*
- Young people in local government ( $\leq 39$ ): *Percentage of young people (aged 39 years or less) elected or in charge in municipal councils over the total number of people elected or in charge.*
- Homicide rate: *Number of homicides per 100 thousand inhabitants.*
- Separate collection of municipal waste: *Percentage of municipal waste collected separately over the total collection of municipal waste.*
- Air pollution: *Number of days in a year during which the level of PM10 exceeds the limit of 50g/m<sup>3</sup>.*

## References

- CHELLI F. M., CIOMMI M., EMILI A., GIGLIARANO C., TARALLI S. 2016. Assessing the Equitable and Sustainable Well-Being of the Italian Provinces. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 24(1), pp. 39-62.
- FOSTER J.E., MCGILLIVRAY M., SETH S. 2013. Composite Indices: Rank Robustness, Statistical Association, and Redundancy, *Econometric Reviews*, Vol. 32, pp. 35-56.

- HØYLAND, B., MOENE K., WILLUMSEN F. 2012, The Tyranny of International Index Rankings, *Journal of Development Economics*, Vol. 97, pp 1-14.
- ISTAT (2015). Il Benessere Equo e sostenibile in Italia, Roma, Istat.
- NARDO M., SAISANA M., SALTELLI A., TARANTOLA S. (2008). Handbook on constructing composite indicators: Methodology and user guide. OECD Publishing.
- PLUMMER, M., 2003, "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling," DSC, (Disability Services Commissioner) 2003 Working Papers, [www.r-project.org/nosvn/conferences/DSC-2003/Drafts/Plummer.pdf](http://www.r-project.org/nosvn/conferences/DSC-2003/Drafts/Plummer.pdf).
- RJPMA A. 2016. What can't money buy? Wellbeing and GDP since 1820. *Working paper* No. 0078. <http://www.cgeh.nl/working-paper-series/>
- TARALLI S., CAPOGROSSI C., PERRI, G. 2015. Measuring Equitable and Sustainable Well-being (Bes) for policy-making at local level (NUTS3), RIEDS, vol. LXIX n. 3, pp. 97-106.

## SUMMARY

### **The Equitable and Sustainable Well-Being at local level: a first attempt of time series aggregation**

The on-going economic crisis has reinforced the feeling that macroeconomic indicators, namely economic indicators at country level, do not provide a correct picture of the living conditions in a territory. In fact, individual and local characteristics also influence the well-being of individuals and, within the same country, territories can vary at a large extent. Thus, the analysis of well-being at local level is crucial.

Here, we analyse the time series of a selection of indicators that constitute the Equitable and Sustainable Well-being (BES) at local level (NUTS3). After providing an overview of the temporal trends in the selected well-being indicators, we construct a composite index of well-being for groups of regions (NUTS1) applying a latent variable model estimated in a Bayesian framework.

---

Mariateresa CIOMMI, Università Politecnica delle Marche, [m.ciommi@univpm.it](mailto:m.ciommi@univpm.it)  
 Chiara GIGLIARANO, Università dell'Insubria, [chiara.gigliarano@uninsubria.it](mailto:chiara.gigliarano@uninsubria.it)  
 Stefania TARALLI, Istat, [taralli@istat.it](mailto:taralli@istat.it)  
 Francesco M. CHELLI, Università Politecnica delle Marche, [f.chelli@univpm.it](mailto:f.chelli@univpm.it)

## THE CONSUMPTION OF FISH IN OLD AGE: A STATISTICAL ANALYSIS ON THE DATASET “MULTISCOPO”

Francesco Bozzo, Stefania Girone, Vincenzo Fucilli,  
Alessandro Petrontino, Ruggiero Sardaro.

### 1. Food Consumption in Italy

The unfavorable and widespread economic situation has led to a general reduction in food consumption over the last few years, and although a timid recovery (+0.3% on an annual basis) was recorded in 2015, in the first half of 2016, household expenditures on food met another slight downturn. This decline is the reflection of a still very uncertain market where the moderate inflation has partially concealed, for some types of product, the downward trend of the quantities purchased<sup>1</sup>. According to the latest available data, drawn up by Ismea on the results of the *Nielsen Panel* (“sales at distribution” and “household expenditures”), the highest percentages in terms of national domestic expenditures of food products are those relating to cereal-base foods or derived from cereals and milk foods (14.7%); followed by vegetables (10.8%), meats net of cold cuts (10.2%), fruits (8.8%) and fish products (7.4%) as shown in Figure 1A. As for the percentage variations between the first-half of 2016 and the first-half of 2015, the worst reductions in consumption among foodstuff are those of meat (-6.1%), various kinds of cold cut (-5.6%), milk and derivatives (-3.4%), vegetables (-1.3%) and fresh eggs (-0.2%) as shown in Figure 1B. The sales of oils and vegetable fats remain stable, while those derived from cereals (+0.3%), fruit (+1.1%) and fish products (+2% in the first-half of 2016, +4.3% in 2015 compared to 2014), among which the latter are especially good for fresh and defrosted products (+4%), while the consumption of preserves is slightly increased (+0.1%) and of smoked foods is decreased (-0.3%).

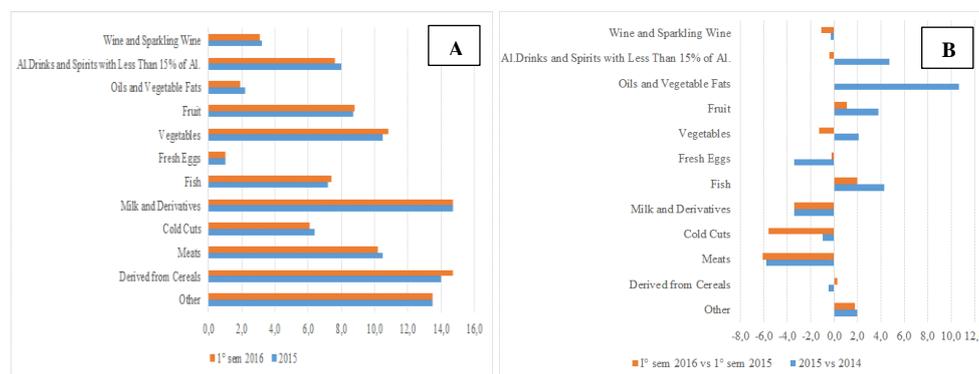
Among alcoholic and non-alcoholic beverages, the wine and sparkling wine sector is contracted at a total of 1.1% (it decreased by 0.3% in 2015 compared to 2014), while the percentage of the other beverages (non-alcoholic drinks and spirits with less than 15% of alcohol) decreased by 0.4% (they increased by 4.7% in 2015 compared to 2014). At present, therefore, the national average of food and drink spending is equal to 441.50 euros per month, a slight recovery compared to 436.06 euros in 2014 (+1.2%). The share of food expenditure in percentage terms is 17.7% of the total expenditure incurred by Italian families. At territorial level, significant structural heterogeneities are linked to income levels, prices and spending behaviors, with northern values higher than those of central Italy, and especially

---

<sup>1</sup> Ismea, 2016.

those of the South and the Islands<sup>2</sup>. As for the types of food consumed, there is some positive health boost with significant shares for people who eat more and more fiber, cereal, vegetable and fruit foods. The shares of fish consumers are equally significant, which are likely to grow, unlike meat and salami consumers who are declining.

**Figure 1** – Share (A) and Percentage Variation (B) in food purchasing.



Source: *Ismea*, 2016.

So, in the light of the emerging issues of food consumption in Italian families, it was considered that a specific study on the behavior of consumers of fish products was particularly useful in order to trace a profile as closely as possible to “the perfect consumer” of fish. Identifying the consumer’s point of view is of primary importance for the production systems as well as for the success of future markets. The framework of fish consumption that will emerge will make it possible to build a market segmentation map on which to correctly place the cues collected on topics and results emerging from the survey carried out. This will allow you in economic-commercial terms to think about possible marketing strategies that industry operators may adopt to make the fish product more attractive to the target consumer identified through the applied statistical model. Not a few demographic, economic, geographical, health, food, social and housing thematic dimensions, analyzed in the multivariate analysis, will be described in this paper. Among these, the one on eating habits and food combinations is particularly interesting, especially as a result of the national and international interest in health and nutrition: in fact, it is widely demonstrated how much the lifestyle and individual/social behaviors influence their nutritional choices and how much they consequently act in an incisive way to avoid the development of many pathological conditions to control their evolution or, on the contrary, to cause their onset. In this regard and with particular reference to fish, it is well known the variety of health components contained in it: from omega-3 to phosphorus, from digestibility to low cholesterol, etc. Thus, using a *Logit Model*, how some explanatory variables, specifically identified to describe the seven dimensions mentioned above, are more

<sup>2</sup> Istat, 2016.

or less significantly related to the frequency of fish consumption will be evaluated. In this sense, the work could, among other things, be a useful value added to the current cognitive framework inherent in the benefits of taking a healthy diet based on Mediterranean style in which fish has always covered a prominent place and the strategic perspectives for public health workers.

## **2. The reference scientific material**

Scientific research does not have extensive literature on fish consumption in the strict sense but, it is equally possible to identify some important scientific contributions<sup>3</sup> that have highlighted the positive or negative effects that certain determinants have on “frequency of consumption”. Some researchers, for example, have demonstrated how important the role of health information is, which would result in a great influence on fish consumptions. In fact, these increase with the strengthening of the belief that fish is good for health, especially in old and weaker people, and therefore more exposed to illness. Nutrition research has identified a group of foods that, following their benefits on degenerative chronic problems, should be included in the diet. One of these is the group of fish products for which the presence of high protein value, low saturated fat, and the presence of vitamins and minerals have been recognized, thus representing the largest source of vitamin D. This opinion about fish as a healthy food has, however, recently been undermined by lesser-known information on safety risk for the potential negative impact on the health for chemical contamination in wild fish. The psychological impact due to the information about safety risk depends primarily on consumer confidence, source of information and information content. In this regard, other researchers have focused on the importance of communication and information sources, on the gap between consumer perception and scientific evidence or differences between the consumption of wild and farmed fish. Specifically, studies by Verbeke et al. have shown that the perception of wild fish is linked to the concern for environmental sustainability, whereas bad knowledge is associated with the idea that farming fish has less intrinsic quality. Studies on communicative effectiveness and information processing have shown that adverse messages or negative pressure on food safety issues can heavily affect consumption decision-making in consumers.

## **3. Statistical material and methods of analysis**

With the aim of investigating food-eating habits in fisheries, it was thought that the use of microdata was necessary to trace a more detailed profile of the “perfect consumer” of fish. In this regard, the consumption information in our country is

---

<sup>3</sup> Banerjee, Ray, 2017; Gaviglio et al., 2009; Gharagozlian, S. et al, 2017; Suárez, Astoviza, 2010; Verbeke et al., 2007b, 2007c, 2008.

available from the second half of the sixties of the last century: ISTAT survey on “Consumption of Italian Families” has, in fact, allowed you to outline a reference framework between the old and new millennium. However, among the traditional statistical sources, Istat has been conducting a survey on “Aspects of Daily Life” for more than twenty years (1993), which has as its main objective to provide, using harmonized definitions and methods, comparable data, both transversal and longitudinal, for the analysis of the daily lifestyles<sup>4</sup>, the welfare and the quality of life of families and individuals, in structural, economic, health, food, geographic, housing and social terms, etc. This is a scientific research that is part of an integrated system of social surveys, the “*Multiscopo sulle famiglie*”, which is conducted annually, whose information *core* is essentially focused on topics related to the daily living habits of an individual and allows you to understand how people live and how satisfied they are with their conditions, the economic situation, the area in which they live, the public utility management that should contribute to improving the quality of life. Thus, the analysis of the results of such a survey has been able to identify the most frequent fish consumers in Italy more than could be done using the specific consumption survey, which would have provided more detailed information in terms of buying fish (fresh, frozen, canned, etc.) but less on other aspects related to fish consumption, concerning the private sphere. Education, work, family life and relationship, leisure, health, lifestyle, etc. are all aspects investigated from a Multi-scope point of view in which the objectivity of behaviors and the subjectivity of expectations, motivations, and judgments contribute to defining the information framework.

The multi-scope reference population, ie the set of statistical unit being investigated, is made up of families<sup>5</sup> residing in Italy and their members. Therefore, the members of cohabitation are excluded<sup>6</sup> Families were randomly extracted from the municipal registry according to a sampling strategy<sup>7</sup> designed to construct a statistically representative sample of the resident population in Italy.

**Table 1** - Reference sampling composition, 2015.

Sample	Municipalities			Families			Individuals	
	Sample Effect.	Sample Theor.	Univ.	Sample Effect.	Sample Theor.	Univ.*	Sample Effect.	Univ.*
	836	849	8.092	19.160	24.855	25.302	44.226	60.493

\* Multiscope Survey Estimate “Aspects of Daily Life”. Data in thousands.

Source: Istat, 2017 (*Aspetti metodologici dell'indagine “Aspetti della vita quotidiana. The reference period is 2015*)

<sup>4</sup> The “lifestyle” associated with the “family nucleus” concept has a strongly indicative role in the behavior of family members, whose actions can be observed in their specificity and singularity as in the network of relationships that are making up within each family. In particular, it is an essential pillar of socio-economic coexistence, both in its relational and consumption-related dimensions.

<sup>5</sup> The family is understood as the unit as a whole, that is, a group of people who are cohabiting and bound by marriage, kinship, affinity, adoption, protection or affective ties.

<sup>6</sup> The survey period is mainly 12 months ahead of the interview, although for some questions the reference is at the time of the interview.

<sup>7</sup> As for the complex sampling design, refer to the Appendix “Multi-scope Sampling Strategy and Accuracy Level of Results” published online by ISTAT.

The actual sample net of *missing* is therefore a result of 19,160 families distributed in 836 Italian municipalities of different demographic amplitude, for a total of 45,226 individuals (Table 1). This is a sample of subjects (Table 2): males (48.02%) and at the age of 60 or older (29.9%), married couples (65.18%), people who graduated from high school (32.61%) and an average household size of 2-3 members.

**Table 2** - Descriptive sample analysis, 2015.

Variable (ref. % higher)	N.	Frequency	Percentage	Min.	Max.
Male (Yes: 1)	45.226	21.718	48,02	0	1
Age (>=60 years)	45.226	13.513	29,9	0	>=75
Married (Yes: 1)	45.226	29.477	65,18	0	1
Level of schooling (Graduated from High School: 4)	45.226	14.011	32,61	0	5
Household size (0-3)	45.226	28.136	62,21	0	12
Job Position (chiefs of labourers, subordinate and similar labourers, novice workmen and homeworkers on behalf of companies: 1)	45.226	12.102	39,6	1	4
Ateco (Industry: 2)	45.226	4.937	16,15	1	4
Geographical Context (South: 4)	45.226	13.351	29,52	1	6
Metropolitan Areas (metropolitan areas and in municipalities >10,000 inhabitants: 1)	45.226	29.232	64,64	0	1
Sport Activity (no: 0)	45.226	33.687	76,55	0	1
BMI >= 18 years (normal weight: 2)	45.226	19.419	51,32	1	4
Health (good: 4)	45.226	22.091	48,85	1	5
Main Meal (lunch: 1)	45.226	3.0875	68,27	0	1
Frequency of Pasta, Rice and Bread Consumption (once a day: 4)	45.226	23.143	52,93	1	5
Frequency of Fish Consumption (least once a week: 3)	45.226	24.494	56,21	1	5
Frequency of Meat and Cold Cuts Consumption (less than once a week: 2)	45.226	21.415	48,93	1	5
Frequency of Vegetables and Fruits Consumption (least once a week: 3)	45.226	16.401	37,48	1	5
Confidence in the State Institutions (no trust: 0)	45.226	9.276	23,92	0	11
Economic Resources (adequate: 3)	45.226	25.143	55,99	1	4
Social Life (go out less than 7 times in the last 12 months: 0)	45.226	40.954	90,55	0	1
Environment (quite satisfied: 3)	45.226	23.416	59,83	1	4

Source: own processing from Istat data, 2017 (Methodological aspects of the survey "Aspects of everyday life. The reference period is 2015).

As regards the job position, among the major percentages there are those of chiefs of labourers, subordinate and similar labourers, novice workmen and homeworkers on behalf of companies (39.6%) and directors, employees (39.6%).

Among the economic activities under the Ateco Code, industry is the leader, excluding the construction sector, (16.15%). With reference to the geographical context of the respondents' residence, the South is the leader (29.52%), followed by the North-West (21.81%) and the North-East (20.31%), while below 20% there are the respondents from Central Italy and the Islands. Nearly 30,000, more than 60% in percentage terms, there are the respondents resident in metropolitan areas and in municipalities with more than 10,000 inhabitants. They are also individuals who do little sport in their free time: in fact, only 23.45% do one or more sports on a continuous basis, nevertheless they are normal weight individuals (51.52%), and they feel they are in good health (48, 85%). The same eating habits reveal a tendency to "healthy eating": a) more than 80% indicated, at least once a day, a consumption of pasta, rice and bread; b) almost 50% indicated, at least once a day, a consumption of vegetables and fruits; c) about 70% indicated a consumption of meat and cold cuts equal to a frequency ranging between "never" and "less than once a week"; d) about 60% indicated a consumption of fish "at least once a week".

With reference, then, to the last 12 months and taking into account the needs of all family members, 56% respondents consider adequate the overall economic resources of the family, however, equally significant is the percentage of 37% of those who consider that the same resources are scarce. These are individuals with a mediocre social life (90.5% of those attended theaters, cinemas, museums, concerts, etc., less than 7 times in the last 12 months) and, with confidence in the state institutions (Parliament, Government, Political Parties, Judicial System, etc.) for the 60% of the responses ranging between 1 and 4 on a scale from 1 (no trust) to 9 (full confidence). The same percentage is detectable in terms of satisfaction with the environmental situation: 59.83% said that they were quite satisfied with the environment (air, water, etc.) in which they live (Table 2).

### 3.1. *The results of a multivariate statistical approach*

The last available *Multiscopo* dataset in 2015 has resulted in a database of 706 variables, including a dichotomous dependent variable that is related to more or less common fish consumption<sup>8</sup>. Therefore, 88 variables were selected through the Cramér's V index, considering only correlations with the dependent variable above 3% and significant at least 10%. Subsequently, the absence of multicollinearity between predictors (V index values between 1% and 17%) was verified. Then a *Logit Model* with a type of selection criteria of stepwise variable and with 10% input and output threshold was applied. In the reference model 27 variables (Table 3) were correlated with the dependent variable considered positively or negatively, more or less significant, and each of which can be related to specific themes that

<sup>8</sup> The fish-related variable is structured in the dataset in 5 modes: 1. Never; 2. Less than once a week; 3. A few times a week; 4. Once a day; 5. More than once a day. This variable, assumed as dependent, was considered dichotomous (NO / YES) by combining the first two modes in: NO = "Less than once a week", and the last 3 modes in: YES = "At least once a week".

can be summarized in aspects related to the demographic, economic, geographical, health, food, social, personal (mood: satisfaction and trust) and housing contexts.

**Table 3** - *Logit Model about the average frequency of fish consumption in Italy.*

Parameter (Code)	Estimate	Err. Stand.	T-Test
Intercept	2,860	0,368	60,246 ***
Family (NCOMP)	-0,026	0,016	2,853 *
Age (ETA')	0,031	0,011	8,750 ***
Celibate and Nubilated (SI/NO)	-0,084	0,048	3,134 *
Schooling (ISTRUZIONE)	0,065	0,026	6,245 **
Job (ATECO)	0,049	0,020	5,663 **
Family Economic Resources (RISEC)	0,128	0,033	15,170 ***
Geographical Context (RIPMf)	0,092	0,015	38,400 ***
Health (SALUTE)	0,047	0,028	2,931 *
Weight Control Frequency (CPESO)	0,053	0,017	9,450 ***
Main Meal at dinner (CENA)	-0,097	0,039	6,154 **
Amount of Wine Consumption (VINO)	0,031	0,015	4,024 **
Frequency of Bread and Pasta Consumption (PANPAS)	-0,132	0,024	29,770 ***
Frequency of Meat and Cold Cuts Consumption	0,444	0,028	256,135 ***
Frequency of Milk and Cheeses Consumption	-0,053	0,023	5,411 **
Frequency of Fruits and Vegetables Consumption	0,356	0,021	294,052 ***
Frequency of Legumes Consumption (LEGUMI)	0,370	0,026	210,549 ***
The use of olive oil for cooking or seasoning	0,385	0,093	17,185 ***
Attention to the right amount of salt (QTSALE)	0,148	0,023	41,289 ***
The use of iodized salt (IODINE)	0,063	0,037	2,857 *
Amount of Mineral Water Consumption (ACMIN)	0,065	0,013	25,649 ***
Frequency of Dining Out (CENFUO)	0,105	0,020	27,709 ***
Frequency to See Friends in Free Time (AMICI)	0,033	0,016	4,430 **
Frequency to Go to Theaters, Cinemas, Concerts, etc.	0,067	0,032	4,443 **
Read Newspapers (LQUOT)	0,028	0,013	4,771 **
Satisfaction with the environmental situation	-0,072	0,026	7,704 ***
Confidence in the State Institutions	0,041	0,007	30,631 ***
Home with Private Garden (GARDEN)	-0,141	0,038	13,778 ***

\*\*\*: sign. 1%; \*\*: sign. 5%; \*: sign. 10%.

Source: ref. Table 2.

Specifically, with regard to the first of the contexts, the demographic one, it follows that: 1) family composition is inversely related to fish consumption, in fact, the tendency to eat fish frequently is more a habit for a couple than in a family, as the more the number of components increases, the less the fish consumption frequency is; 2) age is correlated positively with our dependent variable, in the sense that the higher the age is, the higher the frequency of fish consumption is; 3) civil status is negatively correlated, and therefore married people eat more fish than those who are single people; 4) the degree of education is correlated so that the more it increases, the more the fish consumption rises. The demographic-structural aspects of the population reference sample have certainly made it possible to identify among the most frequent fish consumers those who are elderly couples with smaller families because no longer with dependent children who both at the tender age and adolescent age / juvenile generally make food choices more linked

to commercial foods (sometimes “junk food”) than foods with the most delicate flavors like fish. Education is presumably linked to the standard of living: those who said that they had a higher level of education were also the ones who carried out more profitable economic activities and were more satisfied with their economic resources. All these elements have a significantly positive impact on fish consumption frequency.

The analysis of economic aspects is based on “The work carried out (ATECO, 2007)” and “Assessment of Household Economic Resources”. Thus, there is a greater increase in fish consumption among those who said that they carried out activities more in the tertiary sector rather than in the primary and, therefore, more in the sector of services than in agriculture and / or industry. These were individuals who the more adequate and / or optimistic their total family economic resources were considered, the more fish consumption was reported.

As regards the geographical area, the tendency towards the consumption of fish products increases among those living in the South and in the Islands, decreasing among those in the North and central Italy. In this regard, the calculation of the Cramér's V index (Table 4) has enabled us to identify from a territorial point of view the regions in which there would be a significant, positive and greater correlation with the average fish consumption frequency: residents in the regions of central Italy (Tuscany, Umbria, Marche, Lazio and Abruzzo) said that they consumed fish more frequently; the same can be said for those who resided in the regions of southern Italy (Molise, Campania, Basilicata, Calabria and Sicily) except for Apulia and Sardinia; whereas, for the residents of regions in northern Italy, none excluded, a negative correlation with fish consumption is implied.

With reference to the “Health” context, analyzed in terms of chronic illnesses (diabetes, hypertension, cirrhosis, osteoporosis, tumor, etc.) contracted by the interviewee, it is that: the more you are in good health and pay attention to your body weight, the more you eat fish frequently.

As regards eating habits, there is a negative correlation between frequent fish consumption and choice of dinner as a main meal: the tendency to consume fish increase at lunch time as a main meal, decreasing when the choice of the main meal is dinner. Moreover, a higher amount of wine consumption, a more significant consumption frequency of meat or cold cut, legumes, fruit and vegetables, olive oil for cooking or seasoning would positively affect fish consumption; on the contrary, a more frequent use of pasta, bread and rice as well as milk and cheeses would negatively affect fish consumption.

In the first case, the intake of proteins from animal source foods such as meat and legumes together with vitamins, fibers and mineral salts from fruit and vegetables source foods creates a compensatory effect on the nutritional values contained in fish, in the sense that the subjects use it to complete the diet.

Adding pasta, bread and rice as a source of carbohydrate as well as calcium contained in milk and cheeses to your diet, on the contrary, creates a substitute effect on fish consumption, a symptom of a diet that is more related to a vegetarian style, in which vegetable foods are more preferred than animal foods.

Furthermore, the greater amount of mineral water consumption could be justified by its high mineral content, which would bring a number of health benefits to which the interview subjects are particularly attentive: mineral water lowers, for example, blood pressure, cholesterol, the risk of kidney stones due to magnesium and potassium contents, and improves bone density, skin, muscle and digestion thanks to calcium content.

**Table 4** - Analysis of correlation level related to the average consumption ("At least once a week") of fish in Italy by using Cramér's V index.

Regions	Correlation Cramér's V index ( $\varphi_c$ )	Sign
Piedmont	-0,0171	***
Val D'Aosta	-0,0145	***
Lombardy	-0,0147	***
Trento	-0,0886	***
Veneto	-0,0328	***
Friuli Venezia Giulia	-0,0210	***
Liguria	-0,0093	***
Emilia Romagna	-0,0278	***
Tuscany	0,0117	**
Umbria	0,0052	
Marche	0,0466	***
Lazio	0,0316	***
Abruzzo	0,0076	
Molise	0,0144	***
Campania	0,0707	***
Puglia	-0,0033	
Basilicata	0,0064	
Calabria	0,0293	***
Sicily	0,0474	***
Sardinia	-0,0310	***
Garden	-0,0281	***
Terrace	0,0169	***

\*\*\*: sign. 1%; \*\*: sign. 5%; \*: sign. 10%.

Source: as reported in Table 2.

Attention, then, to the right amount of salt and the choice of iodized salt are aspects that would point out a profile that is clearly outlined, that is of a healthy fish consumer who likes to eat healthy. It is also a consumer who goes out to dinner and essentially has a social life: this is demonstrated by the positive correlation, although with more or less strong significance between frequency of fish consumption and frequency of dining out, to see friends in free time and to go to theaters, cinemas, concerts, etc.

These are individuals who are in mature age and so have more free time to read newspapers and are also confident in the state institutions (Parliament, Government, Judicial system, Political Parties, etc.). However, more frequent fish consumption is also positively correlated with dissatisfaction with the environmental situation (air, traffic, noise, crime, etc.) of the place where you live, almost as if you were looking for "redemption" from the discontent of living in an unhealthy habitat through the intake of healthy food.

Finally, observing housing aspect, it turned out that those who said that they had a home with private garden consumed fish less frequently than those who, differently, said that they had a terrace home (Table 4). This leads to the conclusion that, in the presence of such spaces at homes, consumers prefer the garden for meat grills and the terrace in case of fish meals.

#### 4. Conclusions

In this paper, the proposed analytical approach has made it possible to draw the average profile of Italian fish consumers, which would, in summary, coincide with subjects in old age, couples and no longer with dependent children, and resident in the South and in the Islands more than in the North and central Italy, who are healthy and eating healthy diet, with a level of education and a modest standard living and moderately satisfied on economic life expectancy of their economic resources. Today's elders have undoubtedly higher levels of education and economic well-being compared to those of the last century, along with clearly healthier lifestyles. This has resulted in huge gains not only on average life but also on the quality of survival. Health, especially in the older age groups who are more exposed to chronic-degenerative pathologies, is one of the key factors to lead a quality life. The new generation of elderly people shows better health conditions than in the past: in fact, the share of functional limitations and that of those who say to be in bad or very bad health is lower. Food choices oriented towards "healthy eating" significantly have a better effect on the quality of life in the elderly. Fish, in our particular case, for its high nutraceutical content is among the healthy foods and it is no coincidence that its higher consumption is strongly associated to preferences that are linked to better lifestyles. Those who, on the contrary, said that they were eating more fish were also those who had lower levels of education, more unfavorable economic conditions and bad health, perceived and effective<sup>9</sup>. As regards the structural aspects, all factors being equal such as age, gender, territory and level of education, it is confirmed that the elderly high fish consumers are the couples in an "empty nest". This is also a question of physically, culturally and socially active elderly consumers: living actively the late stage of life in turn triggers a virtuous circle that counteracts social isolation and bad health, understood in the widest sense<sup>10</sup>.

So, to summarize, the emerging information framework undoubtedly provides useful economic and commercial ideas, as it allows you to think of possible marketing strategies that can attract more fishermen or consumers. Being this, for example, perceived as genuine and delicate food, the trust in the seller from who purchases are made is paramount. In fact, far from relying on casual sellers, the

---

<sup>9</sup> One need only to think of those who had exhausting jobs, had little or no precaution, or had often to give up on health care due to limited economic availability. These are today's elders who more than others reveal bad health and, thus, lifestyles.

<sup>10</sup> Istat, 2016. *Op.cit. note 14*.

elderly consumer tends to rely on a well-known and trusted seller. Certainly, the fish market of a modern fish marketing chain creates a feeling of confidence, but different from that of “*ad personam*”, in the market or small shop to which the elderly consumer is usually addressed, and as he is usually educated and healthy, he will pay particular attention to the information / certifications of the purchased product. That being said, one of the potential commercial strategic plans designed specifically toward target consumers should aim at enhancing the nutraceutical aspects of fish: the same labels of packed fish products should be able to improve the health aspects, those on origin, timing, quality, etc., with simplicity and immediacy. In this regard, France recently (2017) adopted the “Nutri-Score”<sup>11</sup>, a new food labeling system based on the color combination that goes from green (A-grade: better nutritional quality) to red (E-grade: worse nutritional quality) to indicate the levels of fat, salt and sugar contained in foodstuff. A new and, in some ways, innovative labeling system, such as the one used by the French, useful to elderly consumers with frequent and sometimes difficult reading difficulties, would lead consumers to more healthy choices and encourage food companies to reformulate their products, focusing more on their nutritional quality. The effects of good communication and marketers, about those health factors that are heavily related to fish consumption, would benefit not only traders but also the ministry of health that wants to promote better healthy eating and lifestyle.

## References

- BANERJEE G., RAY A.K. 2017. The advancement of probiotics research and its application in fish farming industries, *Research in Veterinary Science*, Volume 115, pp. 66-77.
- GAVIGLIO A., PIRANI A. 2009. La pesca sostenibile nella percezione del consumatore, in TREVISAN G. (a cura di). *La nuova PCP per il Mediterraneo. Strumenti innovativi di gestione sostenibile e comportamenti responsabili*, Franco Angeli, Milano.
- GHARAGOZLIAN S., HANSEN R., HAUGEN M., JOHANSEN O., SEIERSTAD S.L., SELJEFLOT I., ARNESEN H. 2017. Changes in dietary pattern when including 700 g of salmon per week to patients with atherosclerotic heart disease, *Clinical Nutrition ESPEN*, Volume 19, pp. 38-44.
- ISMEA 2016. Consumi alimentari. I consumi domestici delle famiglie italiane. Report Consumi, nr. 3/16, agosto 2016.
- ISTAT 2016. Spese per consumi delle famiglie, *Statistiche Report*, 7 luglio 2016.
- ISTAT 2017. Aspetti metodologici dell'indagine “Aspetti della vita quotidiana”, Periodo di riferimento: anno 2015.
- ISTAT. *Annuario Statistiche della caccia e della pesca*, Collana Annuari, Roma, anni vari.
- OMS/Europe, <http://www.euro.who.int/en/home>.

<sup>11</sup> OMS/Europe, <http://www.euro.who.int/en/home>.

- VERBEKE W., VERMEIR I., BRUNSDØ K. 2007. Consumer evaluation of fish quality as basis for fish market segmentation, *Food Quality and Preference* 18, pp. 651-661.
- VERBEKE W., VANHONACKER F., FREWER L.J., SIOEN I., HENAUW S.D., 2008. Camp J.V. Communicating risk and benefits from fish consumption: impact on Belgian Consumers' perception and intention to eat fish, *Risk Analysis*, Vol.28, N.4.

## SUMMARY

### **The Consumption of Fish in Old Age: a Statistical Analysis on the Dataset “*Multiscopo*”**

A certain incentive to buy in a reasoned way is revealed against a generalized reduction in the agri-food expenditure of Italian households: less “junk” food, more healthy food. Fish, in the specific case for its high nutraceutical content, is among healthy foods: not by chance, the shares of those consuming fish tend to be, though timidly, growing. Thus, a study on the behavior of consumers of fish products is particularly useful for tracing a profile as close as possible to the frequent “perfect consumer” of fish. In the context of a multivariate methodological approach applied on “*Multiscopo*” microdata, it is evaluated how some explanatory variables, specifically identified to describe seven thematic dimensions (demographic, economic, geographical, health, food, social and housing), are more or less significantly related to fish consumption. The emerging information framework, characterized by people in old age, couples and no longer with dependent children, resident in the South and in the islands more than in the North and Central Italy, who are healthy and with a level of education and a modest standard living and moderately satisfied on economic life expectancy of their economic resources, can make it possible to build a market segmentation map on which to correctly place the collected points on the arguments and outcomes emerging from the analysis carried out. This will allow you in economic-commercial terms to think about possible marketing strategies that industry operators may adopt to make the fish product more attractive to the target consumer.

---

Francesco BOZZO, Dipartimento di Scienze Agro Ambientali e Territoriali -  
Università degli Studi di Bari “A. Moro”, francesco.bozzo@uniba.it.

Vincenzo FUCILLI, Dipartimento di Scienze Agro Ambientali e Territoriali -  
Università degli Studi di Bari “A. Moro”, vincenzo.fucilli@uniba.it.

Stefania GIRONE, Sinagri S.r.l. - Spin Off dell’Università degli Studi di Bari “A.  
Moro”, stefaniagirone@hotmail.com.

Alessandro PETRONTINO, Sinagri S.r.l. - Spin Off dell’Università degli Studi di  
Bari “A. Moro”, ale.petrontino@gmail.com

Ruggiero SARDARO, Dipartimento di Scienze Agro Ambientali e Territoriali -  
Università degli Studi di Bari “A. Moro”, ruggierosardaro@gmail.com.

**SOCIETÀ E RIVISTA ADERENTI AL SISTEMA ISDS**  
**ISSN ASSEGNATO: 0035-6832**

---

*Direttore Responsabile:* Dott. 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. Libreria Editrice Università di Padova”  
Via G. Belzoni, 118/3 – Padova (Tel. 049/650261)  
[www.cleup.it](http://www.cleup.it)

# ATTIVITÀ DELLA SOCIETÀ

## A) 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).
- 50esima 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).
- LIX Mobilità territoriale, sociale ed economica: modelli e metodi di analisi (Università degli Studi Internazionali di Catania, 25-26 maggio 2017).