STATISTICAL INDICATORS FOR THE ANALYSIS OF SOCIAL INEQUALITIES IN ITALY[1]

Lucia Mongelli, Paola Perchinunno, Samuela L'Abbate, Antonella Massari

Abstract. The analysis of social inequalities is a topic of growing academic and policy interest, recognized as a key dimension in assessing individual and collective well-being. A fundamental prerequisite for any rigorous statistical assessment of inequality is the adoption of a shared, multidimensional definition that captures the complex nature of social disadvantage. This study addresses the need to identify small-area territorial units or population subgroups affected by hardship or severe exclusion. A fuzzy logic approach (Totally Fuzzy and Relative) and a spatial clustering algorithm (DBSCAN) are adopted to construct and interpret indicators of social disadvantage. The analysis draws on 2024 data from the *"Equitable and Sustainable Well-being" (BES)* framework published by ISTAT, focusing on three domains: socio-demography, employment, and personal safety at the provincial level. This integrated methodology enables the detection of both the intensity and spatial configuration of disadvantage across Italian territories. It provides a robust foundation for evidence-based and place-sensitive policy interventions.

1. Introduction

Social inequalities are inherently complex and multidimensional phenomena, encompassing socio-demographic, economic, educational, and safety-related dimensions.

This study aligns with international initiatives such as the OECD Better Life Index and the UNDP Human Development Index, which promote a multidimensional and non-monetary conception of well-being.

The increasing availability of disaggregated territorial statistics enables a more rigorous and data-driven approach to these issues.

---

[1] The contribution is the result of joint reflections by the authors.
This study was carried out within GRINS - Growing Resilient, Inclusive and Sustainable and received funding from the European Union Next-Generation EU (NATIONAL RECOVERY AND RESILIENCE PLAN NRRP MISSION 4, COMPONENT 2, INVESTMENT 1.3 - D.D. 1558 11/10/2022, PE00000018 Spoke 8 Social Sustainability).

The aim of this study is to propose an integrated methodology capable of capturing both the intensity of social disadvantage and its spatial distribution across Italian territories, through the combined use of fuzzy logic and density-based clustering techniques.

The analysis relies on BES (Equitable and Sustainable Well-being) data, published by ISTAT (2024), and introduces a methodological framework that combines a fuzzy synthetic index and spatial clustering to enhance the understanding of multidimensional disadvantage.

2. Objectives and Methodological Strategy

The analysis of social inequalities requires statistical approaches capable of capturing the complexity and multidimensionality of the observed phenomena, moving beyond the traditional dichotomous perspective that rigidly distinguishes between inclusion and exclusion, well-being and deprivation.

The overarching objective of this study is to develop an integrated model that can measure, represent, and interpret social inequalities across Italian territories through a nuanced and spatially explicit representation of disadvantage.

The study has four main aims:

1. Construction of a synthetic index of social disadvantage
    - The index reflects the multifaceted nature of the phenomenon by aggregating multiple thematic indicators.
    - It employs fuzzy membership functions, allowing for a gradual rather than binary measurement of deprivation.
2. Overcoming limitations of traditional thresholds
    - We employ the Totally Fuzzy and Relative method, which assigns each territorial unit a score within the [0,1] interval, reflecting its relative degree of disadvantage with respect to the national distribution.
3. Application of spatial clustering techniques
    - The DBSCAN algorithm is applied to identify homogeneous areas in terms of social distress intensity.
    - DBSCAN is particularly suitable for irregular spatial distributions, as it identifies clusters and outliers without requiring a predefined number of groups.
4. Support for evidence-based public policy
    - The methodology provides a robust basis for designing targeted interventions and moving beyond uniform strategies, which often prove ineffective in capturing local disparities.

The methodological strategy consists of two main and complementary phases. In the first phase, the social disadvantage index is constructed using the Totally Fuzzy and Relative method based on data from the BES (Equitable and Sustainable Well-being) framework. Key indicators are selected from three domains deemed particularly relevant in explaining social sustainability: demography, employment, and personal safety. The indicators are statistically standardized to ensure comparability before being aggregated into a fuzzy synthetic index. In the second phase, territorial clusters are identified through the DBSCAN algorithm, using as input both the synthetic Totally Fuzzy and Relative scores and, in a subsequent step, the raw values of the individual indicators. This choice facilitates a comparative assessment of clustering strategies: one based on the synthesized information using the Totally Fuzzy and Relative method and the other more closely aligned with the original multidimensional structure of the phenomenon.

This integrated approach, which combines fuzzy measurement and spatial clustering, allows for a deeper and more nuanced understanding of social inequalities. It supports a more accurate representation of reality and provides a solid foundation for more informed, differentiated, and sustainable policy interventions.

3. Data Sources and Indicator Selection

The primary data source used for this analysis is the system of territorial indicators from the BES framework (Benessere Equo e Sostenibile - Equitable and Sustainable Well-being) developed and published annually by ISTAT, the Italian National Institute of Statistics. Specifically, the study refers to the most recent 2024 edition, which compiles and disseminates a structured set of indicators disaggregated at the provincial level (NUTS-3), covering all 107 Italian provinces and metropolitan cities.

The BES dataset provides a multidimensional information base aimed at evaluating well-being not only in economic terms, but also from social, environmental, and relational perspectives. It represents both a conceptual and operational framework inspired by the principles of integrated progress measurement beyond GDP, as recommended by the Stiglitz-Sen-Fitoussi Commission (2009), and aligned with the goals of the United Nations 2030 Agenda. It reflects the commitment of statistical institutions to represent territorial inequalities through comparable, timely, and policy-relevant data.

For the purposes of this analysis, three thematic domains were selected as particularly relevant for examining territorial social inequalities. These domains were chosen based on the existing scientific literature and consolidated empirical experience in the socio-economic field.

A) *Socio-demography:* includes three socio-demographic indicators reflecting population structure and health conditions. These variables highlight the entrenched and cumulative nature of disadvantage over time and across generations:

- life expectancy at birth (inverted to reflect higher disadvantage),
- avoidable mortality in the 0–74 age group,
- share of individuals aged 25–64 who have not completed upper secondary education.

B) *Employment:* this domain includes indicators related to labour market participation and the economic sustainability of households. The selected indicators are the following:

- non-participation rate in the labour force among the active population,
- share of NEET youth (Not in Education, Employment or Training) aged 15–29,
- percentage of pensioners receiving incomes below the minimum threshold (as an indicator of economic vulnerability among the elderly population).

C) *Personal Security:* encompasses dimensions related to both perceived and actual exposure to risk. These elements significantly affect urban quality of life and the sense of individual safety. Selected indicators include the following:

- prison overcrowding rate,
- homicide and violent crime frequency,
- number of residential burglaries,
- reported incidents of pickpocketing.

All indicators were subjected to a process of statistical normalization, essential to ensure comparability across variables and to allow for accurate aggregation in subsequent phases of the analysis. Each indicator was also assessed in terms of temporal robustness, complete territorial coverage, and discriminative capacity within the national distribution.

The use of the NUTS-3 territorial level proves particularly suitable for the type of analysis proposed, as it allows for sufficient granularity to capture local specificities, while maintaining adequate sample robustness for advanced statistical elaboration. This balance is crucial for supporting empirically grounded, territorially targeted public policy design.

4. Adopted Statistical Methods

4.1. The Fuzzy approach

The Totally Fuzzy and Relative method, introduced by Cheli and Lemmi (1995), extends the concept of fuzzy sets originally formulated by Zadeh (1965). It allows the assignment of a continuous membership degree in the interval [0,1] to each territorial unit, representing its level of inclusion in the fuzzy set of social disadvantage. This continuous representation overcomes the limitations of traditional binary classification (e.g., inclusion vs. exclusion), offering a more nuanced, realistic, and accurate interpretation of social vulnerability as a gradual phenomenon.

The weights associated with each indicator are calculated based on their discriminative power within the observed distribution, following the approach proposed by Cerioli and Zani (1990). These weights reflect the relative contribution of each factor to the overall determination of disadvantage.

Assuming that k indicators are observed for each unit (e.g., a household or a territorial entity), the membership function for the i-th unit to the fuzzy set of disadvantage can be formally expressed as follows:

$$f(x_{i.}) = \frac{\sum_{j=1}^{k} g(x_{ij}).w_j}{\sum_{j=1}^{k} w_j} \qquad i = 1,\ldots,n \quad (1)$$

where $g(x_{ij})$ represents the degree of membership of unit i to the fuzzy subset associated with indicator j, while w_j is a weight reflecting the relative importance of indicator j.

The weighting system w_j is derived from a generalized extension of the method originally proposed by Cerioli and Zani (1990). It is based on the discriminatory power of each indicator in distinguishing among different observed conditions of disadvantage:

$$w_j = ln\left[1/\overline{g(x_j)}\right], \qquad \text{where} \quad \overline{g(x_j)} = \frac{\sum_{i=1}^{n} g(x_{ij})}{n} \quad j = 1,2,\ldots,k \quad (2)$$

When the average function $\overline{g(x_j)} = 1$ the corresponding weight w_j is equal to zero, while when $\overline{g(x_j)} = 0$, w_j is not defined, or rather X_j is not an appropriate indicator for that collective.

This formulation ensures that indicators contributing more substantially to the differentiation of units within the observed distribution receive higher weights, thereby enhancing the sensitivity of the composite index to real disparities in social conditions.

4.2 The DBSCAN approach

After the construction of the fuzzy index, the second phase of the analysis involves the application of the DBSCAN algorithm (Density-Based Spatial Clustering of Applications with Noise), originally developed by Ester *et al.* (1996), to identify clusters of provinces with similar levels of social disadvantage and to detect potential outlier units. DBSCAN relies on local density estimation, defining a cluster as a set of points that are all reachable from each other through areas of sufficiently high point density. Unlike conventional clustering methods such as k-means, DBSCAN does not require the number of clusters to be specified a priori and is particularly well-suited to irregular or non-convex spatial distributions, making it ideal for complex territorial contexts.

The algorithm relies on two key parameters: ε (epsilon), which represents the maximum distance within which two points are considered neighbors, and *MinPts*, which indicates the minimum number of neighboring points required to form a dense region. To calibrate these parameters, parameter calibration was conducted via k-distance plot analysis. This technique identifies the "elbow point" of the curve, which corresponds to the empirically optimal value for ε. In contexts characterized by high dimensionality or strong territorial heterogeneity, the selection of ε and *MinPts* significantly affects clustering quality, requiring a careful balance between accurately identifying groups and minimizing the number of noise points.

In this study, DBSCAN was applied in two configurations: first, on the aggregated Totally Fuzzy and Relative index to detect clusters of overall social disadvantage; and second, on the standardized values of individual indicators to reveal more detailed clusters specific to the individual domains of demography, employment, and personal security. The integration of the Totally Fuzzy and Relative methodology with spatial clustering via DBSCAN provides a representation of both the intensity ("how much") and the spatial localization ("where") of social disadvantage. This integrated approach provides a robust tool to inform public policy, guide local planning, and evaluate high-impact territorial interventions.

5. Analysis Results

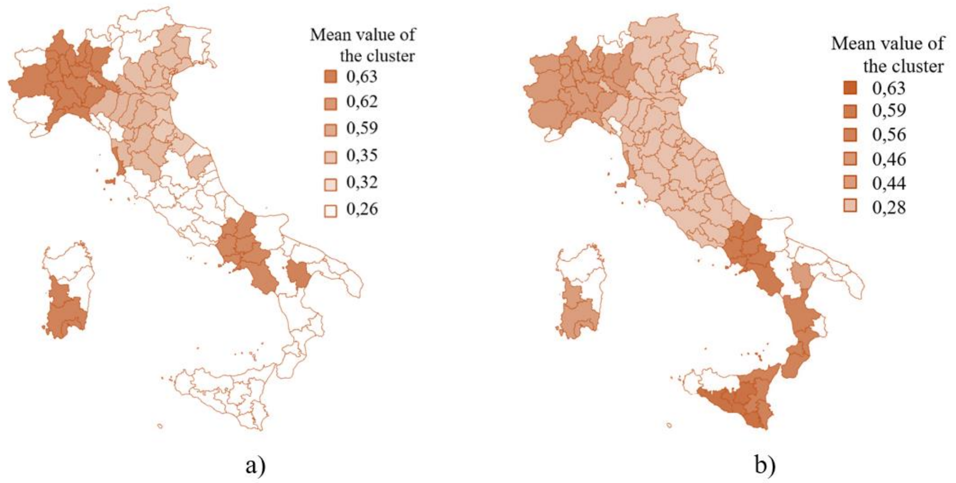
The following section presents the empirical findings emerging from the combined application of the Totally Fuzzy and Relative methodology and DBSCAN clustering. Results are presented by thematic domain: demography, employment, and personal security and by spatial configuration, highlighting both aggregated and domain-specific results. The analysis reveals a complex and fragmented geography of social disadvantage across Italian provinces, challenging linear narratives and calling for nuanced, territorial policy responses.

The provinces were grouped using the DBSCAN algorithm. On the maps, the provinces shown in white were classified by the algorithm as noise points. This does not mean they lack a high fuzzy value, but rather that they are surrounded by fewer than MinPts provinces, and none of those are core points. In other words, noise provinces are located at a distance greater than ε from any core province.

5.1. Socio-Demographic Domain

For the first set, socio-demographic inequality, six clusters were obtained in both configurations mentioned above. The legend shows the average fuzzy values for each resulting cluster; darker colours indicate higher levels of disadvantage among the grouped provinces.

Figure 1 – *Socio-demographic inequality with DBSCAN* **a)** *based on the fuzzy value (ε=0.7 MinPts=4);* **b)** *directly on the three indicators (ε=0.87 MinPts=4).*



Clusters of significant socio-demographic disadvantage are mostly found in the South, particularly in provinces of Calabria, Sicilia, and Campania. However, signs of critical conditions also appear in certain inland provinces, such as Molise and the Abruzzo hinterland.

A notable divergence emerges between the two DBSCAN configurations: the first configuration, based on the aggregated Totally Fuzzy and Relative index, identifies clusters broadly aligned with overall disadvantage levels; the second configuration, based on the three raw indicators, highlights more compact areas.

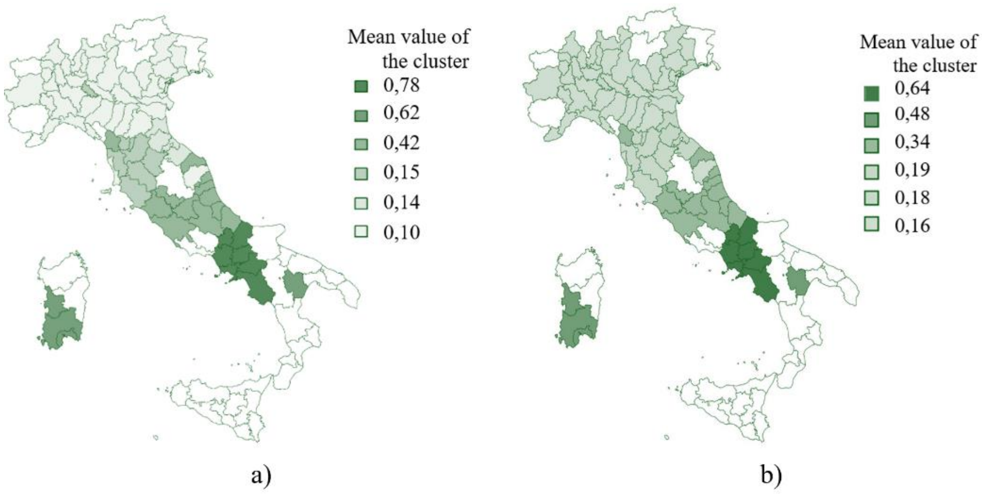
This divergence is largely explained by the presence of southern provinces where educational attainment levels are relatively high, but demographic conditions are poor (e.g., ageing, avoidable mortality).

Therefore, the second configuration better captures these localized mixed conditions of disadvantage. Therefore, the second configuration better captures these mixed localized conditions of disadvantage, also including the inland areas of Molise and Abruzzo.

5.2. Employment Domain

For the employment inequality analysis, six clusters were also identified in both DBSCAN configurations, revealing notable territorial disparities across Italy.

Figure 2 – *Employment inequality with DBSCAN* **a)** *based on the fuzzy value (ε=0.64 MinPts=4);* **b)** *directly on the three indicators (ε=0.745 MinPts=4).*



The findings indicate that the highest concentration of employment-related disadvantage emerges in the regions of Basilicata and Campania, where structural weaknesses in the labour market and persistent socio-economic challenges contribute to elevated levels of vulnerability.

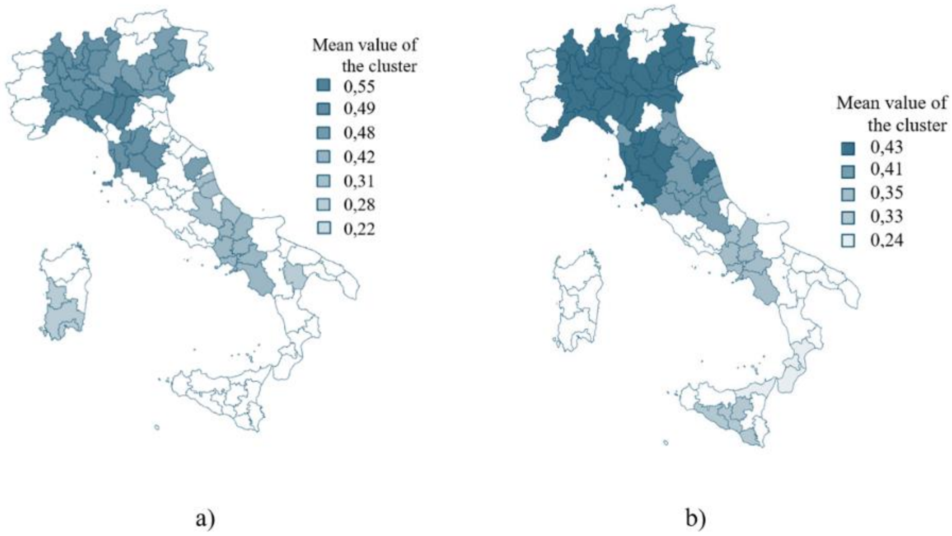
Significant levels of disadvantage are also evident in Molise and in several provinces of Puglia, suggesting the presence of broader economic difficulties within parts of Southern Italy.

Conversely, the northern regions are generally characterized by fewer high-density clusters, a pattern that reflects a stronger overall performance in terms of labour market participation and economic stability, with more resilient local economies mitigating the intensity of employment-related inequalities.

5.3 Personal Security Domain

For the personal safety domain**,** DBSCAN produced seven clusters when applied to the Totally Fuzzy and Relative composite index, while five clusters were obtained when the algorithm was applied using the raw indicator values**.**

Figure 3 – *Personal safety inequality with DBSCAN* **a)** *based on the fuzzy value (ε=0.75 MinPts=4);* **b)** *directly on the four indicators (ε=0.855 MinPts=4).*



The results indicate that disadvantage clusters are primarily concentrated around major metropolitan areas such as Naples**,** Milan**,** and Rome**.** Additionally, several central-northern provinces exhibit critical conditions**,** particularly when analyzing the raw indicators.

Furthermore, some border and coastal provinces in the central-southern regions are also affected, although the distribution in this domain appears to be more fragmented compared to the others.

6. Conclusions and Future Developments

The findings of this study highlight the effectiveness of integrating fuzzy logic and spatial clustering techniques in measuring social inequalities at the territorial level. The combination of the Totally Fuzzy and Relative method with the DBSCAN algorithm allowed for a multidimensional and spatially explicit representation of disadvantage, thereby overcoming the binary logic that traditionally segments populations into “included” and “excluded” groups.

The Totally Fuzzy and Relative method makes it possible to assign a relative score to each territorial unit, reflecting its position within the national distribution of social disadvantage through gradual measurement. This is complemented by DBSCAN, which effectively identifies clusters of similar intensity and territorial outliers, bringing to light situations that might otherwise remain hidden in analyses. This dual perspective, both quantitative and spatial, adds significant value to policy design by enabling place-based public interventions, in line with the European Commission’s guidelines on economic, social, and territorial cohesion.

These findings are consistent with international literature that emphasizes the crucial role of institutional quality and administrative impartiality in shaping territorial disparities in well-being. In particular, Charron, Lapuente, and Rothstein (2013) stress how the quality of government, defined by efficiency, transparency, and absence of corruption, is a structural determinant of socio-economic inequalities across European regions. Integrating governance-related indicators into future iterations of this model may thus enhance its explanatory power and provide further insights for evidence-based policymaking.

Moreover, future methodological developments may include the adoption of more advanced clustering techniques. Recent innovations, such as the HDBSCAN algorithm (Campello, Moulavi & Sander, 2013), offer the ability to detect clusters of varying densities and require fewer parameter specifications than DBSCAN. HDBSCAN’s flexibility makes it particularly suitable for heterogeneous spatial distributions and for identifying complex or irregular cluster structures, which are common in socio-territorial data.

The results show that social inequalities are not exclusively concentrated in Southern Italy but are instead present throughout the entire country, including areas in Central and Northern Italy. This evidence underscores the need for differentiated policy strategies capable of capturing local dynamics and responding through flexible, multi-level approaches.

Several avenues for future development are suggested to further strengthen this line of research. One possible direction is the adoption of HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), an advanced version of DBSCAN that can detect clusters with variable density while requiring only one parameter (*MinPts*); this could simplify adaptability to heterogeneous spatial distributions and improve outlier detection. Another avenue involves expanding the range of BES domains considered: beyond demography, employment, and personal safety, future studies could integrate indicators related to the environment, education, social cohesion, and the quality of public services, thereby broadening the interpretative capacity of the model. Additionally, adopting a longitudinal perspective would allow for the analysis of temporal dynamics in inequality trends and enable the monitoring and evaluation of policy impacts over time. Finally, comparative integration with international frameworks, such as EU-SILC (European Union Statistics on Income and Living Conditions) and the United Nations' Sustainable Development Goals (SDGs), would further enhance international comparability and support the interpretation of results across different territorial and governance scales.

Overall, the proposed approach offers a replicable, flexible, and scalable model that can effectively support territorial policies aimed at reducing social inequalities and advancing sustainable development goals.

References

BARCA F. 2009. *An agenda for a reformed cohesion policy*. European Commission. https://ec.europa.eu/regional_policy/archive/policy/future/pdf/report_barca_v0306.pdf

CAMPELLO R. J. G. B., MOULAVI D., SANDER J. 2013. Density-based clustering based on hierarchical density estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, G. Xu (Eds.), *Advances in knowledge discovery and Data Mining. PAKDD 2013*. *Lecture Notes in Computer Science* (Vol 7819, pp. 160–172). Springer. https://doi.org/10.1007/978-3-642-37456-2_14

CERIOLI A., ZANI S. 1990. A Fuzzy Approach to the Measurement of Poverty. In Dagum, C., Zenga, M. (Eds.), *Income and wealth distribution, inequality and poverty* (pp. 272–283). Springer. https://doi.org/10.1007/978-3-642-84254-1_14

CHELI B., LEMMI A. 1995. A Totally Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty. *Economic Notes,*Vol. 24, No. 1, pp. 115–134.

CHARRON N., LAPUENTE V., ROTHSTEIN B. 2013. *Quality of government and corruption from a European perspective*. Edward Elgar Publishing. https://doi.org/10.4337/9781782545591

ESTER M., KRIEGEL H. P., SANDER J., XU X. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of the second international conference on knowledge discovery and Data Mining (KDD'96)*, (pp.226–231). AAAI Press.

ISTAT 2024. *Il benessere equo e sostenibile in Italia – Indicatori territoriali*. https://www.istat.it/it/archivio/benessere-e-sostenibile

OECD 2020. *How's Life? Measuring Well-being*. OECD Publishing. https://doi.org/10.1787/9870c393-en

STIGLITZ J. E., SEN A., FITOUSSI J. P. 2009. *Report by the Commission on the Measurement of Economic performance and social progress*. Commission on the Measurement of Economic Performance and Social Progress. https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report

UNDP 2023. *Human Development Report: Breaking the Gridlock*. UNDP. https://hdr.undp.org

ZADEH L. A. 1965. Fuzzy Sets. *Information and control*, Vol. 8, No. 3, pp. 338–353. https://doi.org/10.1016/S0019-9958(65)90241-X

Lucia MONGELLI, Istat, Italian National Institute of Statistics, Department for the Development of Methods and Technologies for the Production and Dissemination of Statistical Information, Directorate for Data Collection, mongelli@istat.it

Paola PERCHINUNNO, Department of Economics, Management and Business Law, University of Bari Aldo Moro, Italy, paola.perchinunno@uniba.it

Samuela L'ABBATE, Department of Humanities Research and Innovation, University of Bari Aldo Moro, Italy, samuela.labbate@uniba.it

Antonella MASSARI, Department of Economics, Management and Business Law, University of Bari Aldo Moro, Italy, antonella.massari@uniba.it