Socio-Economic Planning Sciences 95 (2024) 102014

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: www.elsevier.com/locate/seps

Exploring the territorial unevenness of one-person households and contextual factors of vulnerability: Evidence from the Italian context

Federico Benassi [a,*], Alessia Naccarato [b], Pierluigi Vellucci [b]

[a] Dept. of Political Sciences, University of Naples Federico II, Via Leopoldo Rodinò 22, 80133 Naples, Italy
[b] Dept. of Economics, Roma Tre University, Via Silvio D'Amico 77, 00145 Rome, Italy

ARTICLE INFO

Keywords:
Living alone people
Decision algorithm
Geographical pattern
Local analysis

ABSTRACT

New family structures have emerged in Italy in recent decades, with a trend towards smaller nuclear families due to demographic, social and economic changes. An aging population, marital disruptions, declining fertility, and later marriages have contributed to this trend. It is important to understand the changing needs of families, especially the vulnerable, from both an economic and social perspective. Vulnerability is often related to economic factors, but people living alone are often at risk. The goal of this study is to classify Italian municipalities based on the prevailing characteristics of their one-person households, identifying areas of greater or lesser fragility. This classification constitutes a tool to plan people-based policies. Starting from the 2020 Italian Permanent Population and Housing Census data, a decision algorithm was used to identify municipalities according to the different types of their one-person households and to study their geographical distribution throughout the country. Our results show there is an unexpected heterogeneity that goes far beyond the classical North–South divide, emphasizing the urgency of approaching the study of economic and social processes at the local level.

1. Introduction

In the last decades, important changes in family structures and dynamics have occurred in Italy, and new forms of families have spread across the country [1–3]. Many demographic and economic factors have contributed to this transformation. The aging of the population, declining fertility, longer lengths of stays in the family of origin by young adults, marital disruptions, and greater age at the time of one's first marriage, are the most important determinants of the changes in family structures and dynamics that have occurred so far [4–8]. These changes are usually associated with the Second Demographic Transition [9], a macro-demographic transformation that characterize nowadays most of the Western societies, with different magnitudes and dynamics [10,11]. In this framework, one of the major changes is the so called nuclearization process, manifested by the progressive decrease in the average number of family members and the spread of a family structure in which multigenerational families are increasingly rare [12]. This has led to an increase in the number of one-person households. In the view of Cámara et al. [13], this is one of the most significant demographic and socioeconomic phenomena that Western societies have undergone since the mid-twentieth century. Such transformations are related to macroeconomic changes such as the process of globalization [14] and the growth of the urban population [15]. But, as is reasonable, the economy is also affected by these changes.

The growth in one-person households has implications for consumption [16], housing [17,18] and environmental resources [19]. For instance, when the process of growth of one-person households is linked with the aging process, the issue of living standards and well-being emerges [20]. The growth of one-person households is a process of particular interest in societies where the family has traditionally been held to be an important factor in determining living arrangements [13]. This is reinforced by the history of Italy, a Catholic country in which the concept of family was (and in some way still is) highly influenced by traditional values and religiosity [21]. The relevance of Italy is also due to the aging process of its population, which is one of the causes of the growth in one-person households, and which in Italy has reached a high level [22].

From an economic and social planning perspective, it is necessary to understand how the needs of families have changed, especially those of the most vulnerable. The vulnerability of a family is generally related to economic and social factors [23,24]. However, people living alone can be reasonably considered to be more likely to be in fragile situations [25]. In this general framework, the territorial dimension of the one-person households emerges as a main topic and a challenge

* Corresponding author.
E-mail addresses: federico.benassi@unina.it (F. Benassi), alessia.naccarato@uniroma3.it (A. Naccarato), pierluigi.vellucci@uniroma3.it (P. Vellucci).

https://doi.org/10.1016/j.seps.2024.102014
Received 13 January 2024; Received in revised form 20 May 2024; Accepted 5 July 2024
Available online 11 July 2024

for applied quantitative research. As proved by many recent contributions regarding the Italian context [3,26–28], the population dynamics and changes leading to the transformations fuelled by the Second Demographic Transition are essentially spatial processes. Therefore, to implement efficient and useful actions of regional planning and policy interventions, it is necessary to consider how these processes manifest themselves across spaces [29]. Indeed, territorial contexts are not neutral elements with respect to the demographic and socio-economic process that occurred in such areas [30]. For this reason, to study the geographical distribution of the phenomenon is a fundamental step for planning and promoting place-based policies [31]. These are policies which attempt to reduce the gaps between territories in disadvantaged conditions and those in better conditions, thereby increasing the economic competitiveness of the former and the overall level of well-being. From a more purely demographic point of view, interest lies in a particular subcategory of these policies, namely those defined by Ladd [32] as place-based people strategies, which remain territorially targeted but are oriented towards directly supporting specific groups of the population who reside in those territories, in our case, the one-person households. Naturally, to produce adequate statistical information it is necessary to carry out our analysis at a very fine territorial scale (e.g., a local scale), that is to say, with reference to Italy, at the municipal level.

In other words, it is necessary to think locally [33]. Before proceeding further, it is necessary to clearly define some key conceptual issues as well as the objectives of this paper. Regarding the first point, it is necessary to provide more details on what is meant, in this paper, by the term "vulnerability". As is well known, vulnerability is a multidimensional concept whose definition involves considerable complexity [34]. Vulnerability, understood as the increased exposure of a given population to exogenous risk factors and, therefore, as a potential fragility of the individuals composing that population, encompasses a range of dimensions including economic, social, and psychological ones. Thus, phenomena such as economic deprivation, loneliness and isolation, social exclusion, and marginalization, may all be linked, more or less interchangeably, to the broader concept of vulnerability [35]. These phenomena are not mutually exclusive but often interact with each other, fuelling self-perpetuating vicious cycles [36]. Indeed, limited economic resources can lead to social integration problems and, consequently, processes of isolation – even spatially – of specific segments of the population, which will gradually become more fragile and more exposed to exogenous shocks, thereby increasing their level of vulnerability. Territorial contexts where critical conditions prevail are likely to negatively impact the level of fragility of an individual residing in those areas, following a macro–micro logic. From an empirical perspective, the type of dimension of vulnerability to observe is closely linked to the availability of the data and, in the case of territorial studies, to the geographic scale at which we intend to investigate the phenomenon. These two conditions act in opposite directions. Usually, as we move down the territorial scale, the availability of data, and particularly individual variables, decreases. Many survey investigations, while addressing issues closely related to vulnerability, do not provide estimates below the regional level, a rather coarse territorial level. A classic example is the Eu-Standard Income and Living Conditions (Eu-Silc) survey. Our contribution arises from the need to observe the phenomenon of single-person households at a very granular geographic scale, that of the municipality. This condition, as just mentioned, has influenced the choice of data from among the somewhat limited set of options. Vulnerability is therefore approached on a meso-scale, characterizing the places (municipalities) where single-person households tend to concentrate more. The empirical analysis focuses on a combined study of the distribution of single-person households and the characteristics of the spatial contexts in which they live. The general idea, therefore, is not so much to measure the level of vulnerability of single-person households per se, but to study their territorial distribution on a local scale, taking into account, at the same time, some characteristics of the places where these families reside, which may indicate situations of potential criticality. In doing so, we will naturally consider some socio-demographic characteristics of the household as well. For a single-person household to reside in an area where the percentage of people receiving basic income is high and where demographic dynamism is low may be indicative of a (potentially) high level of vulnerability. Therefore, this paper is based on two assumptions: (i) The vulnerability of a one-person household depends on both individual characteristics and ecological variables, and (ii) The vulnerability of one-person households is unevenly distributed over Italy. The objectives we intend to pursue are: (i) to detect the major categories of one-person households and (ii) to illustrate their distribution over Italy.

In order to classify Italian municipalities according to the types of one-person households within them, data from the 2020 Italian Permanent Population and Housing Census (PPHC) have been used. We employ a decision algorithm to identify the types of one-person families, using various characteristics related to both the family and the municipality. The decision algorithm leads us to four prevalent types of one-person household, enabling us to analyse their geographical distribution and create specific thematic maps.

The remainder of this paper is organized as follows: Section 2 provides an in-depth examination of the phenomenon of the one-person household in Italy. In Section 3, we introduce the decision algorithm employed for classifying Italian municipalities based on the key characteristics of one-person households, accompanied by the definition of the Prevalent One-Person Households type.

Moving forward, Section 4 offers a comprehensive overview of the data used in applying the decision tree to Italian municipalities. Section 5 presents the outcomes, focusing on the geographical distribution of the Prevalent One-Person Households typology. Lastly, Section 6 engages in a discussion of the results and concludes the paper.

2. One-person households in Italy. Evolution across time and space

The family is the fundamental and primary unit of any economic and social system. On the family, and its structure, depends the structure of consumption and the dynamics of investment, and also the adoption of demographic and social behaviours (e.g. having less or more children, migrating, etc.) having economic effects. These processes, in turn, influence the family's structure. Demographers and social scientists have explored the evolution of the Italian family and the changes that have defined it [21,37–39].

An extremely significant element has been the marked growth and spread of one-person households. In Italy this process had assumed impressive dimensions, making deep changes in the social structures of society. Fig. 1 shows the temporal dynamics of this process from the beginning of the 20th century to the present day, both as an absolute value (panel a) and as a relative weight with respect to the total number of Italian households (panel b). In 1901, the number of one-person households was 614,816, constituting 8.8% of all households. In the same year, the mean size of the households was 4.5, with the larger households (i.e. the ones with at least 6 persons) constituting 30.0% of all households. Data coming from the permanent census, in 2021, certified that the number of one-person households now exceeds 9.6 million, constituting 36.8% of all households and the mean size is 2.2 persons per household. Recently, Lo Conte et al. [39] predicted that the scenario of the next 20 years (2041) will be characterized by similar trends: an increase in the number of households, a decrease in their size, and a significant growth in the number of one-person households.

This is at a national scale, but the differences in the levels and the dynamics of such changes vary greatly across the country [3]. In 2021, there is a North–South gradient: the highest percentage of one-person households, considering the total number of households in each macro-geographical area, is recorded in the Northwest (39.0%), while in the South, the level is the lowest (32.0%). Referring to the

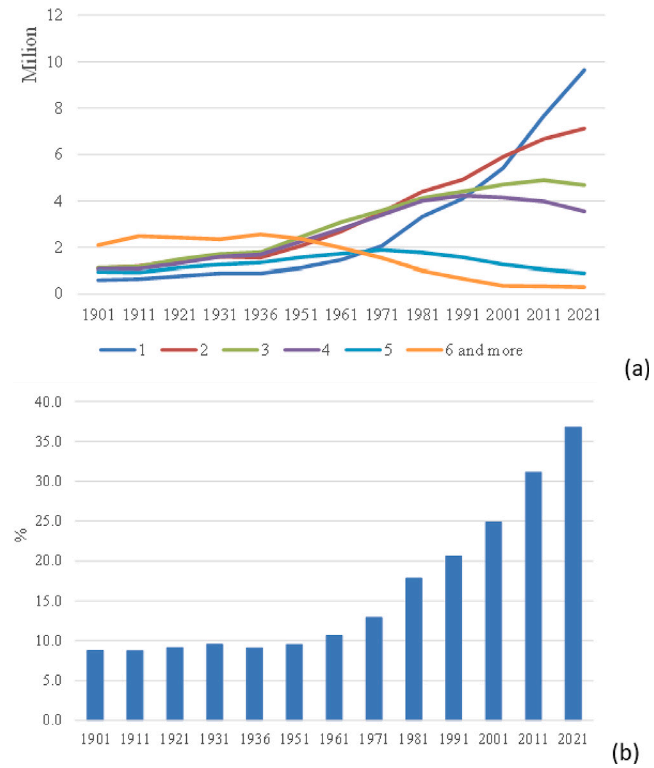

Fig. 1. The evolution of household composition in Italy, 1901–2021: households by number of components (a); percentage proportion of one-person households (b).
Source: Own elaboration based on Istat "Serie Storiche" data (1901–2011) and Istat Permanent Population and Housing Census (PPHC) 2021 data.

total number of one-person households in Italy, the lowest percentage is recorded in the Islands (only 10.4% of all one-person households) while the highest is recorded in the Northwest (29.6%). Territorial differences increase when we consider a finer geographical scale. A very high spatial variability is recorded at the municipality level (Fig. 2). In the left panel (a) the percentage of one-person households at the municipality level is mapped using the quantile maps. In the right panel (b) a categorical map is shown, indicating in green the municipalities where the percentage of one-person households is not higher than the national level in 2021 (36.8%), red indicates those where the percentage of one-person households is higher.

Evidently, the territorial distribution of the one-person households does not present a clear geographical pattern. High percentages (fifth quintile) of one-person households are recorded both in the inland and remote areas of Italy, e.g. the Apennine and foothill areas and the rural or in any case non-urban areas of both northern and southern Italy and the Islands, and also in the large urban and metropolitan municipalities such as Rome, Genoa, Milan, Turin, and Venice. Furthermore, the North–South distinction is quite tenuous and, in any case, not generalizable. On the other hand, the lowest percentages of one-person households (first quintile) are both in the North, typically in the municipalities of the Po Valley, but more in general in the surroundings areas of the large metropolitan municipalities affected by processes of extensive suburbanization, and in the South and the Islands. This is the case of Apulia, but also of large areas of coastal Campania and parts of Sardinia.

The clear differentiation between municipalities with a percentage of one-person households below the national average and those surpassing it (as depicted in the map on the right) underscores the lack of a distinct North–South pattern. It also emphasizes the significance of the local context. Moreover, the intra-regional variation is notably pronounced, with only a few regions, particularly Apulia and Lombardy, exhibiting some attenuation in this variability.

What we have observed clearly indicates the importance of the local dimension in the study of these phenomena, that is to say the urgency of thinking locally [33]. In the following paragraphs we will therefore try to refine the classification analysis taking into account the local dimension and the different characteristics that may affect the profiles of one-person households which, as is evident, are internally very heterogeneous in terms of demographic and socio-economic characteristics.

3. Methods

We use some characteristics of the municipalities and the one-person households residing in them to identify, by means of a decision algorithm, the geographical distribution of Italian municipalities classified according to the type of one-person household. To do this, we indicate the municipality by m, with Y the target variable defined as the ratio between the number of one-person households in 2020 and the total population in each municipality m, and x_{mj} denoting the jth variable observed in municipality m ($m = 1, \dots, M$; $j = 1, \dots, J$) referring to the characteristics of both households and municipalities.

An initial classification is obtained by considering the quartiles of the distribution of Y. The municipalities are then divided into four groups based on the percentage of one-person households residing in them. Using quartiles can simplify the initial classification process. The division into quartiles can reduce the complexity of the problem, making it easier to handle homogeneous groups of municipalities before applying a more detailed decision algorithm. Applying the decision algorithm to each of the four groups identified by the quartiles of the target variable, the municipalities are classified according to the different combinations of the categories of the variable X.

3.1. The decision algorithm

A classification tree is chosen for our classification, due to its simplicity and interpretability. The process of building the tree follows the greedy procedure outlined in Algorithm 1 of Mastroeni et al. [40], which seeks to find the optimal data partition. This approach is commonly used by CART [41] and its popular implementations [42, 43].

In decision tree algorithms, a cost function is used to measure the effectiveness of splitting a node into child nodes during the construction of the tree. The goal of a decision tree is to recursively split the data into subsets based on certain criteria, and the cost function helps determine the best way to make those splits. The cost function is closely related to the splitting criteria, which are conditions used to decide how to split a node into two or more child nodes. The splitting function, i.e. a function for splitting, used to evaluate the "goodness" of the alternative splits for a feature, determines the best feature and its corresponding threshold as follows:

$$\left(j^*, t^*\right) = \arg \min_{j \in \{1,\dots,J\}} \min_{t \in \mathcal{T}_j} \left[\text{cost}\left(\left\{\mathbf{x}_m, y_m : x_{mj} \le t\right\}\right) + \right.$$
$$\left. + \text{cost}\left(\left\{\mathbf{x}_m, y_m : x_{mj} > t\right\}\right)\right]$$

where the cost function for a given dataset will be defined below. Here, $\mathcal{T}_j$ represents the set of possible thresholds for feature j, which can be derived by sorting the distinct values of x_{mj}. For example, if feature 1 takes the values $\{4.5, -12, 72, -12\}$, then we define $\mathcal{T}_1 = \{-12, 4.5, 72\}$. In the case of real-valued inputs, we compare a feature x_{mj} to a numeric value t. For categorical inputs, we consider splits of the form $x_{mj} = c_k$ and $x_{mj} \neq c_k$ for each possible class label c_k. The training set $\mathcal{D}$ is then divided into two subsets: the left subtree $\mathcal{D}_L$ and the right subtree $\mathcal{D}_R$, using a single feature j and a threshold t.

The function that determines whether a node is suitable for splitting employs a stopping heuristic based on the cost reduction Δ:

$$\Delta \triangleq \text{cost}(\mathcal{D}) - \left(\frac{|\mathcal{D}_L|}{|\mathcal{D}|}\,\text{cost}\left(\mathcal{D}_L\right) + \frac{|\mathcal{D}_R|}{|\mathcal{D}|}\,\text{cost}\left(\mathcal{D}_R\right)\right)$$

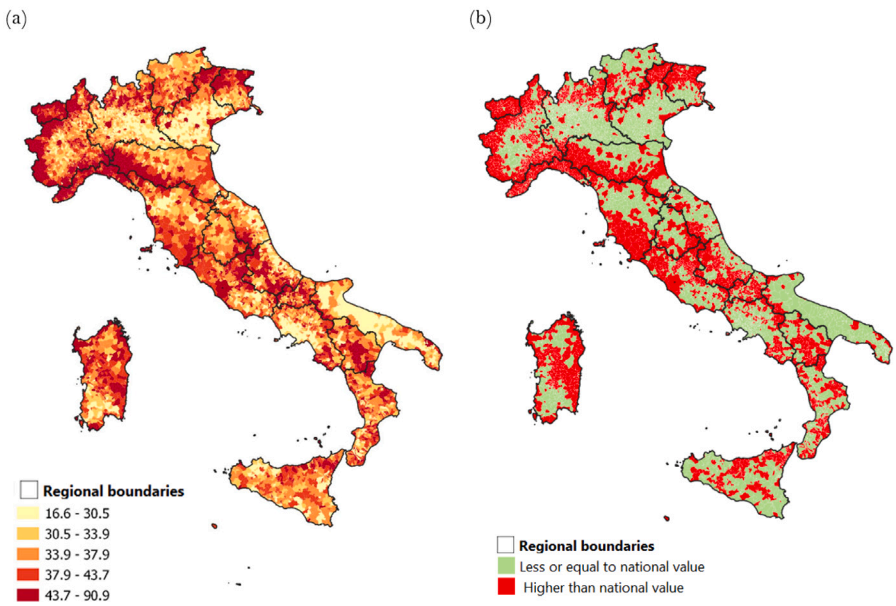

Fig. 2. Geographical distribution (%) of one-person households in Italy, 2021 (a) and quantile details (b).
Source: Own elaboration based on Istat Permanent Population and Housing Census (PPHC) 2021 data.

To define the cost of a classification, we use impurity functions. Let f denote an impurity function, and define the impurity of a node A as follows:

$$I(A) = \sum_{k=1}^{C} f\left(p_{kA}\right) \tag{1}$$

Here, p_{kA} denotes the proportion of individuals in node A belonging to class k (in our case, $C > 2$).

We consider two potential candidates for f: the information entropy $f(p) = -p\log(p)$ and the Gini index $f(p) = p(1-p)$. In this paper, our algorithm splits the tree based on the Gini index. To determine the importance of a variable, we calculate the reduction in the Gini index associated with each variable at each split and then sum these reductions.

The approach adopted in this context is to initially grow a large tree, denoted by T_0, and then prune it to identify an optimal subtree. We define a subtree, T, as any tree that can be derived by pruning T_0, and we use $|T|$ to denote the number of terminal nodes in T. Each node, A_h, corresponds to a region with N_h observations, and the node A referenced in Eq. (1) is just one of them. Therefore, we introduce a cost complexity criterion as follows, following [44]:

$$L_\alpha(T) = \sum_{h=1}^{|T|} N_h I\left(A_h\right) + \alpha|T| \tag{2}$$

In this equation, $I\left(A_h\right) = I\left(A_h, T\right)$ as defined by Eq. (1), and α is a complexity parameter. The objective is to identify the optimal subtree that minimizes $L_\alpha(T)$. Larger (smaller) values of α result in smaller (larger) subtrees. For instance, when $\alpha = 0$, the solution corresponds to the full tree, T_0.

We used the method of cross-validation, with 10-fold cross-validation repeated 10 times. This is a technique used to evaluate the performance of a machine learning model. In this approach, the dataset is randomly divided into k non-overlapping subsets (folds). The model is trained on $k-1$ folds and tested on the remaining fold. This process is repeated k times, using each fold exactly once as the test set. In the case of repeated cross-validation, this process is repeated multiple times over different random splits of the data. This helps reduce the variance in the estimation of the performance of the model resulting from the randomness of the data split.

Ten-fold cross-validation is a specific instance of cross-validation, one in which the dataset is divided into 10 folds. Out of these, 9 folds are used for training and 1 fold is used for validation. This process is repeated 10 times so that each fold is used exactly once as the test set. Finally, the performance of the model is computed as the average of the performances obtained in each of the 10 iterations.

The complexity parameter is chosen from among those that, over the different folds (10 in our case), minimize the accuracy.

3.2. Multiclass metrics assessment for map classification

Since we consider the quartiles of the distribution of Y, we have a multiclass problem unlike what we had in a previous paper [45], where we examined the binary case.

For multiclass models, the definition of the accuracy can be extended from binary classification to encompass multiclass problems. In the case of multiclass accuracy assessment, the overall accuracy is calculated by considering the number of true positives across all observations. The evaluation of accuracy in multiclass classification involves using a confusion matrix, where the predicted class is depicted in the rows and the observed class is depicted in the columns [46,47].

To determine the overall accuracy for a multiclass problem, the main diagonal of the confusion matrix is summed, giving the total number of correct observations. The resulting sum is then divided by the total number of pixels in the error matrix. This approach provides a comprehensive measure of accuracy for multiclass classification. For instance, in Eq. (3),

$$\Omega = \begin{pmatrix} \text{Class} & 1 & 2 & \dots & C & \text{Total} \\ 1 & \omega_{1,1} & \omega_{1,2} & \dots & \omega_{1,C} & \omega_{1,\cdot} \\ 2 & \omega_{2,1} & \omega_{2,2} & \dots & \omega_{2,C} & \omega_{2,\cdot} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ C & \omega_{C,1} & \omega_{C,2} & \dots & \omega_{C,C} & \omega_{C,\cdot} \\ \text{Total} & \omega_{\cdot,1} & \omega_{\cdot,2} & \dots & \omega_{\cdot,C} & n \end{pmatrix} \tag{3}$$

the confusion matrix Ω illustrates the relation between the predicted and the observed classes. E.g., $\omega_{k,k'}$ denotes the number of observations classified as type k' when they are actually of type k; C is the number of rows and columns of the matrix and corresponds to the number of classes. By $\Omega[k,k']$, we denote the entry of the confusion matrix at row k and column k'.

The overall accuracy is computed as the sum of the number of correct observations for each class (ω_{kk}) divided by the total number of observations (n):

$$\text{Multi-Class Accuracy} = \frac{\sum_{k=1}^{C} \Omega[k,k]}{\sum_{k,k'} \Omega[k,k']} = \frac{1}{n}\sum_{k=1}^{C} \Omega[k,k] \tag{4}$$

In the context of binary classification tasks, Precision is defined as the ratio of True Positives to the total number of positive predictions of the model (the column sum of the predicted positives). Specifically, True Positive refers to elements labelled as positive by the model that are indeed positive, while False Positive refers to elements labelled as positive by the model that are actually negative.

$$\text{Precision} = \frac{TP}{TP+FP}$$

'Recall' is the ratio of True Positive elements to the total number of units actually belonging to the positive class (the row sum of actual positives). In this case, False Negative represents elements labelled as negative by the model that are actually positive.

$$\text{Recall} = \frac{TP}{TP+FN}$$

'Recall' assesses the model's predictive accuracy for the positive class, measuring its ability to identify all positive units in the dataset.

The extension to multi-class problems involves aggregating values over rows or columns of the confusion matrix. Assuming the matrix is oriented such that a specific row corresponds to a particular "truth" value, the Precision and Recall for class k are calculated as follows:

$$\text{Precision}_k = \frac{\Omega[k,k]}{\sum_{k'} \Omega[k',k]}$$

$$\text{Recall}_k = \frac{\Omega[k,k]}{\sum_{k'} \Omega[k,k']}$$

In other words, Precision represents the proportion of instances where class k was correctly identified out of all instances where the algorithm identified k, while Recall is the proportion of instances where class k was correctly identified out of all instances where the true state of the world is k.

3.3. Prevalent one-person household type

Applying the decision algorithm to the municipalities of Italy, we obtain multiple classes of municipalities for each quartile, namely, the algorithm leads to a total of 50 classes of municipalities (c_k, where $k = 1,\dots,50$). Since the aim of this paper is to classify Italian municipalities according to the prevailing characteristics of their one-person households, only the 'barycentre' class was considered for each quartile. We define the barycentre class as the one to which the largest number of municipalities belong, i.e. the class that saturates the quartile the most in terms of statistical units (municipalities). The J variables' categories combination corresponding to the barycentre, defines the Prevalent One-Person Household type (POPH).

This yields 4 groups of municipalities – one for each quartile – with which to study the geographical distribution. Municipalities in each quartile that do not fall in the 'barycentre' (residuals) are excluded from the analysis.

Fig. 3 shows the distribution of the percentage frequencies of municipalities, by quartiles and classes. It is evident that in each quartile, the municipalities excluded from the analysis – those that do not belong to the barycentre class – are distributed among the remaining classes quite evenly with a very low frequency. For this reason, in order to summarize the characteristics of the most frequent one-person households in Italy, we have excluded from the analysis all municipalities that do not fall within the POPH of each quartile. However, since the percentage of excluded municipalities is not negligible (Table 3), it is necessary to verify whether they define a specific pattern, different from that of the municipalities falling within the POPH. The results of the analysis conducted on the municipalities excluded from the POPH are reported in Section 5.

4. Data description

In order to classify Italian municipalities according to the types of one-person households, data from the 2020 Italian Permanent Population and Housing Census (PPHC) on households residing in Italian municipalities are used, jointly with information coming from other administrative archives of Istat.

Specifically, the characteristics referring to the one-person households are age (AGE), sex (SEX), citizenship (CIT), ownership of one or more dwellings (DWE), ownership of one or more cars (CAR), educational level (EDU), and whether they receive citizenship income or not (CIN). For each of the characteristics of the one-person households, we have a set of variables corresponding to each of the categories of a specific characteristic. For example, for AGE, whose categories are 18–34, 35–66, and 67 and over, we have three variables: the number of one-person households between 18 and 34 years old; the number of one-person households between 35 and 66 year old; and the number of one-person households over 67 years of age. Specifically, the characteristics of the one-person households are defined as follows:

- For SEX, we have two variables, counting, respectively, the number of male (M) and female (F) one-person households in each municipality.
- For CIT, there are two variables, counting, respectively, the number of Italian (IT) and foreign (NoIT) one-person households in each municipality.
- For DWE, there are three variables, counting, respectively, the number of one-person households owning 0, 1 or 2 or more houses, in each municipality.
- For CAR, there are three variables, counting, respectively, the number of one-person households owning 0, 1 or 2 or more cars, in each municipality.
- For EDU there are eight variables, counting the number of one-person households that are Illiterate (ILL), Literate with no educational qualification (LWD), having a primary school level (PRI), a secondary school level (SEC), an high school level (HIG), a first level degree (FLD), a secondary level degree (SLD), and a Ph.D. (PHD), in each municipality.
- For CIN, there are two variables, counting, respectively, the number of one-person households receiving (CIYes) or not (CINo) citizenship income, in each municipality.

Istat also provides the interactions between some of these variables: for example, we have the variable counting the number of male (or female) one-person households for each of the three age classes 18–34, 35–64, and 65 and over. The available interactions refer to the categories of AGE combined with those of all the other characteristics of the one-person households just described. Moreover, we consider 6 contextual characteristics of municipalities: geographic macro-area (GA), and region (REG), referring respectively to the administrative division of the Italy into 5 macro-areas (Northwest, Northeast, Centre, South, Islands) and 20 regions; degree of urbanization (DU) [48], classifying municipalities into Cities (CITY), Small towns or suburbs (SCITY) and Rural areas (RURAL) according to the classification based on the criterion of geographical contiguity and minimum density and population thresholds of the regular grid with 1 km^2 cells; capacity of elderly care homes (BED), considering two variables, BED1 – number of beds available in elderly housing facilities – and BED2 number of guests in elderly housing facilities. Furthermore, we consider the local wealth variable (WEA). The variable WEA is a proxy of local wealth obtained from the total amount of the personal income tax paid and the number of taxpayers in each municipality, both derived from 2019 tax returns.

As for the geographical data (shape files), these are related to the Italian municipalities and are provided by the Italian National Institute of Statistics on its official website. The thematic maps were obtained using QGIS Desktop version 3.20.2.

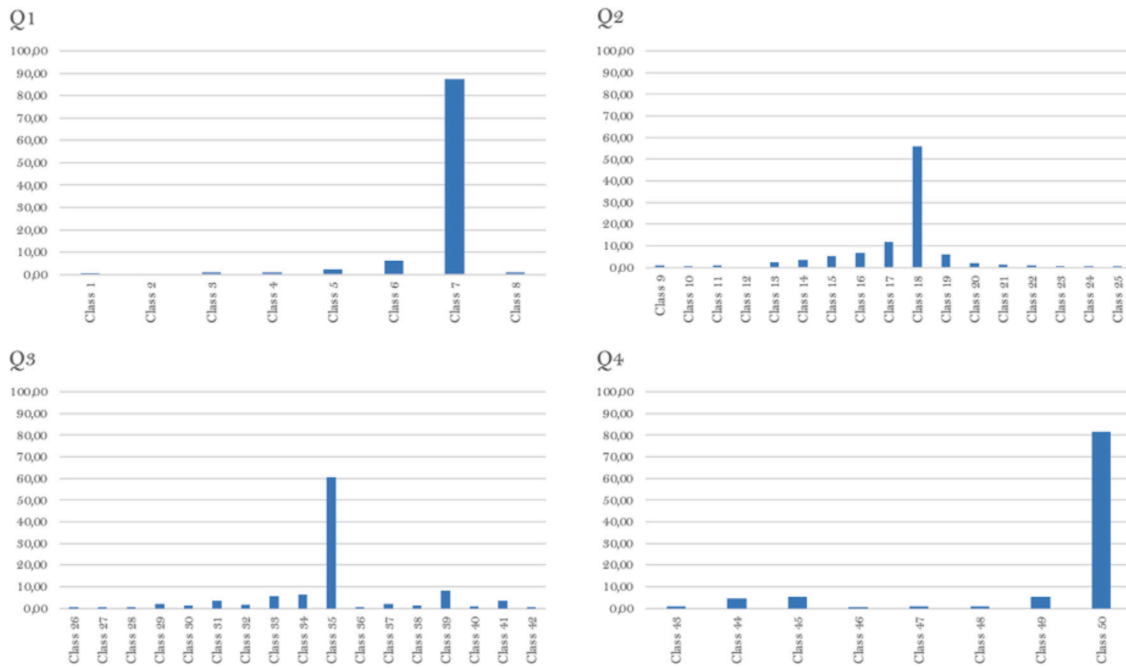

Fig. 3. Distribution of municipalities (%), by quartiles of target variable and decision tree classes.

Table 1
Municipalities distribution (%) by geographical area, region and quartile.

Geographical area	Region	$Q1$	$Q2$	$Q3$	$Q4$	Total
North	Piedmont	11.26	22.35	29.47	36.92	100.00
	Valle d'Aosta	0.00	1.35	22.97	75.68	100.00
	Lombardy	37.58	32.20	16.33	13.88	100.00
	Veneto	61.46	18.29	9.95	10.30	100.00
	Friuli-Venezia Giulia	15.35	29.77	30.23	24.65	100.00
	Emilia-Romagna	13.41	36.89	26.52	23.17	100.00
	Liguria	0.85	4.70	21.37	73.08	100.00
	Trentino-Alto Adige	9.22	37.94	39.01	13.83	100.00
Total North		32.51	28.48	19.63	19.37	100.00
Centre	Marche	9.22	37.94	39.01	13.83	100.00
	Lazio	12.17	22.75	30.69	34.39	100.00
	Tuscany	22.34	23.08	31.87	22.71	100.00
	Umbria	9.78	30.43	46.74	13.04	100.00
Total Centre		13.85	27.71	34.73	23.71	100.00
South and Islands	Abruzzo	21.64	21.31	22.95	34.10	100.00
	Basilicata	8.40	22.14	29.01	40.46	100.00
	Calabria	18.32	22.28	34.41	25.00	100.00
	Campania	39.82	21.09	22.36	16.73	100.00
	Molise	5.88	19.12	29.41	45.59	100.00
	Apulia	51.36	24.51	16.34	7.78	100.00
	Sardinia	13.53	22.81	27.32	36.34	100.00
	Sicily	20.51	25.90	33.08	20.51	100.00
Total South and Islands		25.14	22.59	26.82	25.45	100.00
Italy		26.85	26.12	24.68	22.35	100.00

5. Results

As we have already described in Section 3, the municipalities were initially classified according to the quartiles of the target variable; thus, for example, the first quartile contains all municipalities in which the percentage of one-person households out of the total number of residents in the municipality does not exceed 25%. This initial classification shows that the regional distribution of one-person households in Italy is not homogeneous (Table 1).

Differences are found both between regions belonging to the same macro-area and between regions belonging to different macro-areas. Among the northern regions, Piedmont, Valle d'Aosta and Liguria register the highest percentage of municipalities belonging to Q4, with percentages above 70% in the latter two regions. These regions are marked by a limited number of urban centres and a monocentric structure. It is noteworthy that Lazio, in the Centre, and, in the breakdown of the South and the Islands, Abruzzo, Basilicata, Molise, and Sardinia, share this common characteristic.

Nevertheless, the heterogeneity is substantial, and each regional context possesses its unique identifying profile. Geographically extensive and populous regions such as Lombardia and Veneto, in the North, Campania and Apulia, in the South and the Islands, record the majority of municipalities classified as Q1. Emilia Romagna, located in the macro-area of the North, is the only regional context where a majority of municipalities classified in Q2 are recorded, whereas the remaining regions record the majority of municipalities classified in cluster Q3. Note that this occurs in three out of four regions in the Centre while in only two out of eight regions in both the North and the South and the Islands.

For a deeper assessment of the adequacy of the choice of using quartiles instead of a different classification of municipalities, we compared the results in terms of the accuracy of the decision tree, with those obtained in the case of a different distribution of municipalities. In particular, Appendix A shows the results for the transition from the 4 groups defined by the quartiles to the 7 groups inferred from the empirical distribution of the units. The performance of a 7-group classification is significantly lower than that obtained by grouping municipalities into quartiles.

A decision tree as described in Section 3.1 is applied with the full set of the J variables described in Section 4. The resulting multi-class accuracy is 0.96, with the recall and precision presented in Table 2. In our case, the complexity parameter α, defined by (2), is near to zero: $\alpha = 0.0002812148$, indicating a decision tree that is close to being the full tree.

The application of the decision algorithm to each of quartiles results in a classification into subgroups of homogeneous municipalities with respect to the category combinations of the J variables.

The decision algorithm leads to the identification of 50 classes of municipalities; some of these classes have a small number of municipalities. To synthesize the phenomenon according to its most frequent characteristics and thus provide a key for easily reading its distribution

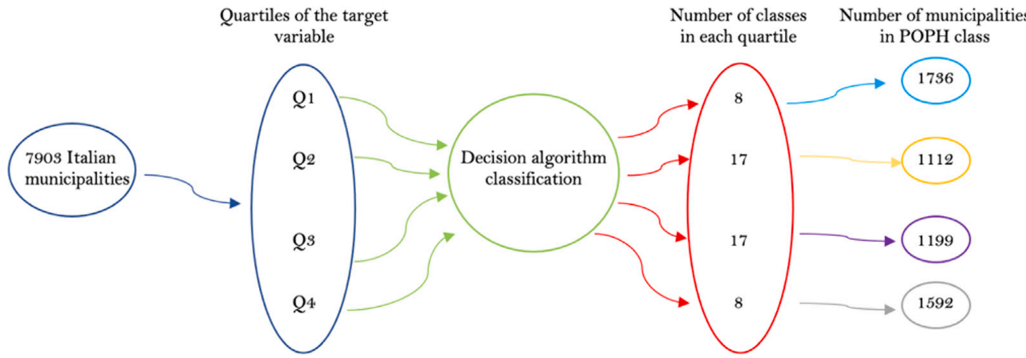

Fig. 4. Outline of the classification procedure for Italian municipalities.

Table 2
Multi-class precision and recall.

Class	Precision	Recall
Q1	0.970	0.980
Q2	0.954	0.949
Q3	0.950	0.959
Q4	0.985	0.970

across Italy, from the classification provided by the decision algorithm we consider for each quartile only the class that contains the largest number of municipalities (i.e. the barycentre).

The barycentre in each quartile is defined as the Prevalent One-Person Household type (POPH). Fig. 4 shows an outline of the procedure for classifying Italian municipalities into POPHs.

The barycentre classes, i.e. the four POPH types, have different weights in terms of the number of municipalities (Table 3). The residual municipalities, i.e. those that do not fall into any of the four POPH types, number 2264 and correspond to 28.6% of the total number of Italian municipalities. However, since the number of municipalities excluded from the territorial representation is quite high, especially in quartiles Q2 and Q3 (Table 3), we conducted a further analysis to verify whether the excluded municipalities exhibit significantly different characteristics from those included in the POPH. Firstly, we compared the means and standard deviations of each of the J variables used, calculated for Italy, for the four quartiles, and for the 4 POPH. The results (Table B.6, in the Appendix) show the consistency of the results between the different groups of municipalities Secondly, to investigate further the differences between the municipalities included in the POPH and those excluded, for each quartile of the target variable we conducted a test for the difference between means (average number of one-person households exhibiting a specific variable's category) of all the variables used to define the profiles, considering two reference groups for each quartile: municipalities belonging to the POPH and those excluded. At a significance level of $\alpha = 0.01$, the results of the analysis (Table B.7, in the Appendix) show that for the first and fourth quartiles, the difference between means is significantly equal to zero for all variables, while in the second quartile, the means are significantly different from zero only with reference to four variables: Education-Illiterates, Bed1 (beds in facilities for the elderly) and Bed2 (guests in facilities for the elderly), and Wealth. In the first case, it concerns a very rare characteristic of individuals in Italy (the percentage of illiterates in 2020 is 0.5%) [49]. The second and third cases concern the variables counting the number of beds and guests in facilities for the elderly, which differ significantly from the average of municipalities included in POPH2 only for 6 large cities (Genoa, Milan, Padua, Bologna, Rome, Palermo). Large cities in Italy can be considered as special cases with respect to the variable BED, as they exhibit values of the capacity of elderly care facilities that are very different from those of smaller and more numerous municipalities. The fourth case refers to a proxy of the wealth of municipalities, which in the second quartile presents average values of 19,431.04 euros in municipalities included in POPH2 and 18,400.04 euros in municipalities excluded from POPH2. Since the differences highlighted by the test appear to be of slight magnitude and related to particular cases, we prefer a synthetic representation of the phenomenon using the group of municipalities in POPH2 already identified. The test provides completely different results in the case of the third quartile. For municipalities in the 3rd quartile, there exists a difference between the means of the two groups of municipalities, significantly different from zero for almost all considered variables. Therefore, with reference to only the municipalities in the 3rd quartile that do not fall into POPH3, we repeated the classification using the decision algorithm to determine whether these municipalities, not represented in terms of household characteristics by the POPH3 profile, present specific peculiarities defining spatial patterns different from those already identified with the first classification. The results of this analysis are presented in the Appendix and clearly show that the decision algorithm does not classify municipalities into homogeneous groups with acceptable reliability. In other words, this highlights that one-person households residing in municipalities excluded from POPH3, although having different characteristics from households in municipalities included in POPH3, are very heterogeneous internally and do not allow for a geographic representation that can highlight particular patterns of their spatial distribution.

In Fig. 5, the top 10 variables used in the classification model, ranked by their importance, are presented: Citizenship Income, Citizenship, Sex, Age, and Car Ownership. As already seen in Section 3, the algorithm splits the tree based on the Gini index. To determine the importance of a variable, we calculate the reduction in the Gini index associated with each variable at each split and then sum these reductions, which are presented in Fig. 5 for the top 10.

It is worth emphasizing that the inclusion or exclusion from citizenship income plays a pivotal role in defining all categories within the POPH classification. In each split of the decision tree, the importance of a variable is calculated by measuring how much the split based on that variable has contributed to reducing the overall Gini index compared to the situation where no such split would have occurred. The importance assigned to a variable in a specific split is the difference between the Gini index before and after the split. The greater the reduction in the Gini index, the higher the importance attributed to that variable. The importance of a variable is then tabulated by summing the reductions in the Gini index attributed to that variable across all splits in the tree. It is important to note that even variables that were not actually used to split the data in a particular division will contribute to their importance, as the model considers all variables as potential candidates for splitting.

Table 4 explicitly presents the combinations of features of the J variables defining the POPH class.

From Table 4, it is possible to deduce the prevalent characteristics of the municipalities belonging to the POPH in each of the quartiles. The

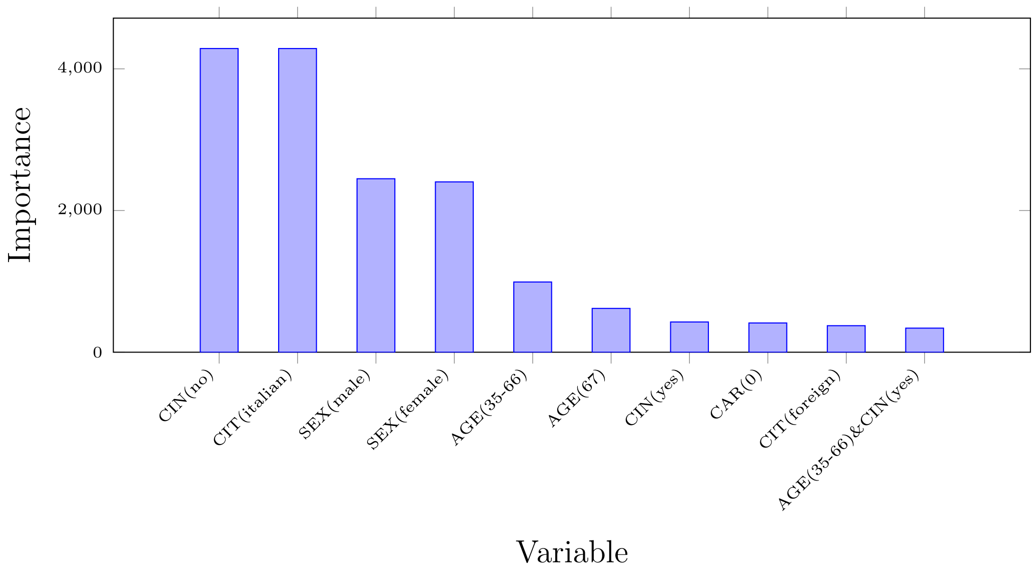

Fig. 5. Variable importance (top 10).

Table 3
Number of municipalities in the POPH class, by quartiles.

Quartile	Number of municipalities	Number of municipalities in POPH class	Percentage of municipalities in the POPH class out of total municipalities in each quartile
Q1	1976	$1736(c_7)$	87.8
Q2	1976	$1112(c_{18})$	56.2
Q3	1975	$1199(c_{35})$	60.7
Q4	1976	$1592(c_{50})$	80.5
Total	7903	5639	

Table 4
Combination of J variables' categories defining the POPH classes, by quartiles.

POPH	Variables			
	Citizen Income	Citizen Income	Citizenship	Age
POPH1: foreign citizens and in working age with high potential vulnerability	Percentage of one-person households not receiving citizenship income less than 29%		Percentage of Italian one-person households less than 29%	Percentage of non-Italian one-person households aged between 18 and 34 less than 1.4%
POPH2: medium–high potential fragility, with mixed citizenship who own a car	Percentage of one-person households not receiving citizenship income between 29.7% and 32.9%	Percentage of one-person households receiving citizenship income between 0.5% and 3.3%	Percentage of Italian one-person households less than 32.8%	Percentage of one-person households aged between 18 and 34 without a car less than 3%
POPH3: medium–low potential fragility, with a mixed composition of citizenship	Percentage of one-person households not receiving citizenship income between 29.6% and 38.4%		Percentage of Italian one-person households between 33.8% and 39.9%	Percentage of one-person households aged between 35 and 66 between 12.3% and 21.5%
POPH4: low potential fragility	Percentage of one-person households not receiving citizenship income greater than or equal to 40.7%			

POPH of the first quartile (POPH1) consists of municipalities (c_7) where the percentage of one-person households receiving citizenship income is high (at least 71% of the families in POPH1 receive citizenship income). It is therefore characterized as a group of municipalities with high economic vulnerability and a high presence of one-person households of non-Italian citizenship, especially of working age (mostly over 34 years old). POPH1 can be defined as "foreign citizens and of working age with high potential vulnerability". The spatial distribution of the municipalities (Fig. 6) reveals diverse patterns with certain consistencies: Emilia Romagna stands out with a notably low number of municipalities falling into this category, while considerable concentrations are observed in Apulia and Campania, particularly in proximity to the regional capitals. Additionally, there are notable concentrations in some central regions such as Marche, especially along the coast, and in specific areas of southern Lazio. The Islands (Sicily and Sardinia) also exhibit a significant presence, particularly close to their respective regional capitals. Regions characterized by high population density include those with intense urbanization in Veneto and Lombardy, outlining a densely populated and intricate spatial continuum.

Among the 14 metropolitan city capitals in Italy, only Naples and Palermo belong to POPH1, underscoring how the potential vulnerability dimension primarily impacts central urban areas. In other metropolitan cities, especially in the North, the suburbs are more

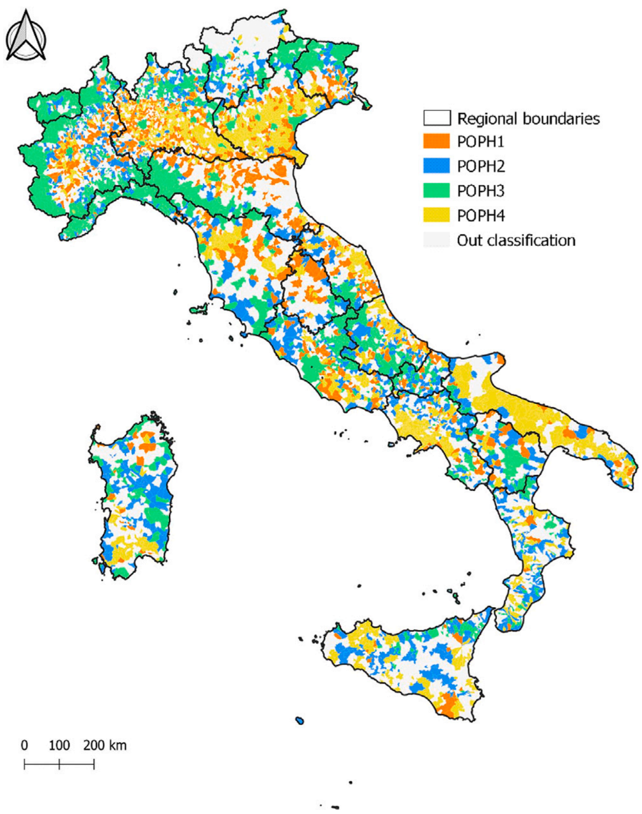

Fig. 6. Italian municipalities by POPH of each quartile.

affected by these dynamics, signalling extensive suburbanization processes that have, to some extent, "weakened" the outskirts of these areas, mirroring trends observed in other Southern European contexts [28,50,51].

The POPH of the second quartile (POPH2) comprises a collection of municipalities (c_{18}), primarily situated in non-central areas of Italian provinces (Fig. 6), where the percentage of one-person households receiving citizenship income is slightly lower than in POPH1. Additionally, the presence of one-person households with foreign citizenship is slightly less compared to POPH1.

Municipalities falling into POPH2 also exhibit a distinct characteristic, which is ownership of cars. POPH2 can be defined as "medium-high potential fragility, with mixed citizenship (Italians and foreigners) who own a car". The occurrence of municipalities falling under POPH2 is more evident in Central-Northern Italy than it is in the South, where the number of such municipalities is notably lower. Specifically, in Central-Northern Italy, a substantial number of POPH2 municipalities are situated in the proximity of major metropolitan city, such as Milan, Turin, Venice, Florence, Bologna, and Rome, albeit with varying degrees of intensity and geographic spread. The predominant characteristic of the municipalities within POPH2 is their non-rural nature, and the overall geographic distribution mirrors that of municipalities representing the third quartile (POPH3).

Municipalities falling under POPH3 (c_{35}) are typically small in size and are often situated in inland areas of Italy, predominantly characterized as non-urban. This typology can be described as "medium-low potential fragility, with a mixed composition of citizenship". These municipalities are distributed to varying degrees across all regions, with a notable concentration in the South, especially in insular areas such as the Ogliastra region in Sardinia, which experiences a comparatively higher impact from this type of municipality.

The municipalities within the POPH4 category (c_{50}) are primarily distinguished by their geographical features. POPH4 encompasses all the principal municipalities of Metropolitan Cities in Central-Northern Italy: Milan, Turin, Genoa, Venice, Bologna, Florence, and Rome, as well as Cagliari, which belongs to the southern and island macro-area. Municipalities classified under POPH4 are characterized by "low potential fragility", mainly consisting of single Italian individuals with a relatively higher age. These are households of elderly individuals residing alone in major cities, often owning the property they live in and receiving pensions (with a lower percentage of citizenship income recipients compared to other POPH categories). The vulnerability of one-person households in POPH4 municipalities is, therefore, lower. This might seem counterintuitive given their elderly demographic; however, one-person households in urban areas, particularly in Central-Northern Italy, often include individuals without profiles of fragility. It is noteworthy that the other metropolitan city capitals, specifically those in the South, excluding Cagliari, do not fall within this particular classification.

Supporting this conclusion, various empirical studies have indicated that the well-being of individuals living alone, primarily in metropolitan areas, tends to be comparatively higher among the elderly demographic than among the younger population [25]. Additionally, it is crucial to highlight that metropolitan regions, particularly in Italy, and specifically in Central-Northern contexts, enjoy a greater provision of infrastructure and demographic dynamism [52]. In the POPH4 classification, there are also municipalities that are notably aged and typically non-urban, concentrated mainly along the Apennine ridge and in regions characterized by advanced age, such as Liguria, which registers a significant number of municipalities falling under POPH4 (Fig. 6). Similarly, these municipalities are often found along border areas, particularly in alpine and pre-alpine zones.

6. Discussion and conclusion

Space is a continuous variable, and any classification and synthesis of it is necessarily partial and improvable [53]. However, the loss of information due to synthesis is an unavoidable aspect when dealing with spatially varying processes, on the one hand, and with fragmented and heterogeneous territories on the other. Their classification, furthermore, becomes a necessary condition for the adoption of effective policies that truly support the needs of inhabitants. To achieve this goal, it is necessary for policies to be territorially targeted; the territory, in fact, is not a neutral element for these policies but interacts with them, often determining their success or failure [54].

Italy is a privileged laboratory for studies and research in this field. It is characterized by profound and lasting territorial imbalances, both of an economic and demographic-social nature. Growing territories, typically urban areas in the Centre-North, are opposed to others in systematic decline, becoming increasingly marginal and more exposed to exogenous shocks, both natural and economic [55]. These differential trends hide many other diversities and are often linked to broader processes, such as the spread of single-person households. The latter phenomenon, considered by many as the most significant change in industrialized societies in recent decades, is projected to continue relentlessly and unavoidably in the near future, according to recent forecasts [39].

In the present paper, which constitutes an exploratory analysis, Italian municipalities have been categorized according to certain attributes of their one-person households, establishing connections between these features and the geographical placement of the municipalities. The decision algorithm employed for the classification has facilitated the examination of the probability distribution across numerous combinations of family and municipal variables, leading to an intelligible synthesis of the phenomenon.

A decision algorithm, something still rarely applied in geographic studies, provides some useful insights in terms of representing a social phenomenon on a local scale. In particular, the extensive use of much statistical information at a very detailed and granular geographic level has allowed us to propose a taxonomy of Italian municipalities that shows how local heterogeneity is an indispensable element and confirms the urgency and importance of thinking locally [33]. All of this also underscores the importance of producing increasingly detailed and timely municipal-based statistics to refine classification studies and, hopefully, enable the adoption of ever more efficient and effective policies.

The present study is not without some limitations that are important to briefly mention. The analysis refers to only one year and should therefore be considered partial. Some aspects that emerged could be linked to temporary conditions which, especially when the analysis is conducted at a very fine geographic scale like that of municipalities, can sometimes have a significant impact on the results. On the other hand, having a greater number of temporal observations could allow the identification of more structural aspects as well as possible transitions in the state of municipalities. Another limitation is that the classification method adopted here is essentially a non-spatial approach, in the sense that the spatial attributes of the elementary units (the municipalities) do not directly (explicitly) enter the analysis, which indeed does not consider any neighbourhood structure (no spatial weights matrix) or any spatial constraint. While this allows obtaining a classification less constrained by contiguity conditions, which often bias the results, especially when contiguous spatial partitions (i.e. regionalization) are desired, it is a limitation both methodologically and substantively, as the clusters obtained are often formed by municipalities that are very far apart from each other, posing a challenge to the implementation of specific policies.

Research funding

The paper was conceived and realized as part of the PRIN 2022-PNRR research project “Foreign population and territory: integration processes, demographic imbalances, challenges and opportunities for the social and economic sustainability of the different local contexts (For.Pop.Ter)” [P2022 WNLM7], Funded by the European Union—Next Generation EU, component M4C2, Investment 1.1. The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them”.

CRediT authorship contribution statement

Federico Benassi: Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Alessia Naccarato:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Pierluigi Vellucci:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

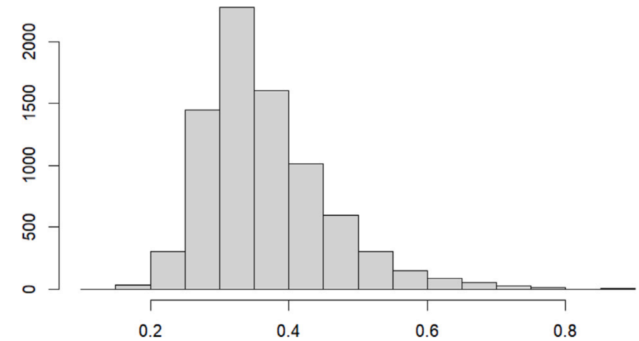

Fig. A.7. Histogram of the Target variable.

Table A.5
Multi-class precision and recall.

Class	Precision	Recall
Q1	0.942	0.971
Q2	0.927	0.959
Q3	0.909	0.925
Q4	0.907	0.845
Q5	0.887	0.922
Q6	0.962	0.850
Q7	0.924	0.984

Appendix A. Transition from 4 to 7 classes

We considered the empirical distribution of a target variable given by the histogram represented in Fig. A.7. We assigned the data in the first two columns of the histogram to class Q1, the data in the five central columns to classes Q2–Q6, and the remaining tail of the data to class Q7. This classification scheme increases the number of classes in the classification problem from 4 to 7.

In our case, the complexity parameter α, as defined by (2), is approximately ten times greater than that for the case with 4 classes, specifically $\alpha = 0.002173484$. In Fig. A.8, the depiction of the top 10 variables employed in the classification model with 7 classes is presented.

The resulting multi-class accuracy is 0.915404, with the recall and precision presented in Table A.5. The expansion of our classification problem from 4 to 7 classes has led to notable changes in the performance metrics. In particular, the overall accuracy has decreased from 0.9646 in the 4-class scenario to 0.9154 in the 7-class scenario. Examining the precision and recall for individual classes provides deeper insights into the nuances of the model's performance. In the 4-class setting, the precision ranged from 0.950 to 0.985, with recall ranging from 0.949 to 0.980. Notably, the model achieved high precision and recall for each of the four classes, indicating a robust ability to correctly identify instances belonging to those categories. However, with the transition to a 7-class problem, the precision and recall exhibit more variability. While precision remains relatively high for several classes, there is a noticeable decline for some, with Q4 decreasing from 0.985 to 0.907. This reduction in precision suggests that, in the 7-class scenario, the model is less precise in correctly classifying instances for certain classes. Additionally, the values of the recall also vary, with Q7 standing out with a remarkable increase from 0.970 to 0.984. On the other hand, Q4 shows a substantial decrease in recall from 0.970 to 0.845. These shifts in recall indicate changes in the model's ability to capture all instances belonging to a particular class. Hence, the expansion to a 7-class classification problem introduces both challenges and opportunities. The model's overall accuracy decreases, reflecting the increased complexity of distinguishing between seven classes. Precision and recall highlight the trade-offs and variations in the model's performance across individual classes, emphasizing the importance of a nuanced evaluation when dealing with a higher number of classes in a multiclass classification problem.

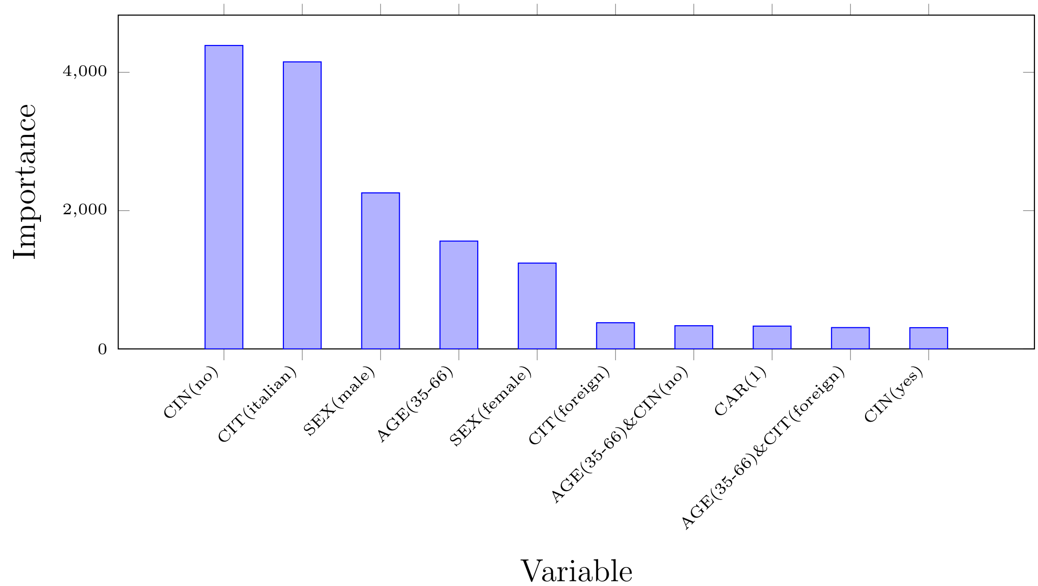

Fig. A.8. Variable importance (top 10) for the classification model with 7 classes.

Table B.6
Mean and standard deviation of the number of one-person households for each of the characteristics of the one-person households, by group of municipalities.

	Variables	Group of municipalities																	
		Italy		Q1		POPH1		Q2		POPH2		Q3		POPH3		Q4		POPH4	
		Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD	Mean	StD
Age	18 – 34	137.89	1361.71	98.03	313.89	101.27	333.56	111.16	239.72	112.93	238.16	130.27	373.23	90.53	242.94	212.11	2667.55	234.48	2962.53
	35 – 66	559.92	4544.90	459.90	1567.41	477.38	1667.09	491.22	958.71	525.31	967.08	550.62	1448.18	392.39	1028.67	737.95	8782.17	804.53	9752.32
	67+	472.56	3022.53	423.72	1378.41	443.50	1466.77	445.31	849.09	461.18	822.17	476.35	1169.85	361.05	884.07	544.85	5705.66	588.03	6332.59
Sex	Male	532.33	4009.19	440.38	1378.33	456.07	1465.29	475.13	895.35	498.80	885.43	526.79	1322.76	383.69	941.74	687.00	7734.63	746.26	8588.23
	Female	638.05	4851.73	541.27	1877.47	566.08	1998.17	572.56	1143.36	600.63	1137.11	630.45	1656.63	460.29	1209.14	807.90	9303.91	880.78	10 329.35
Citizenship	Italian	1059.91	7475.10	913.34	2926.00	951.47	3112.37	969.51	1859.47	1013.95	1836.70	1053.72	2647.81	787.81	1963.02	1303.05	14 298.56	1409.59	15 870.19
	Not Ita.	110.47	1420.05	68.32	354.37	70.68	376.14	78.18	193.29	85.48	189.99	103.52	347.71	56.17	196.17	191.85	2788.36	217.44	3100.43
Dwelling	0	443.63	3620.04	360.05	1467.45	377.51	1562.22	375.35	781.49	379.34	775.45	442.87	1191.86	311.71	822.14	596.23	6943.56	653.14	7710.38
	1	569.05	3666.87	513.98	1383.28	533.40	1468.90	544.36	1009.00	576.52	980.02	566.95	1420.91	424.13	1043.58	650.92	6988.36	699.45	7753.90
	2+	157.70	1623.72	107.62	431.45	111.23	459.05	127.98	270.54	143.56	281.87	147.42	409.97	108.13	305.77	247.75	3179.56	274.44	3533.98
Car	0	515.67	4515.14	432.51	1770.22	455.99	1885.10	437.82	960.93	438.60	950.75	482.60	1269.80	360.17	928.19	709.73	8708.96	784.97	9679.70
	1	631.11	4295.10	526.12	1459.32	542.24	1549.71	587.91	1062.54	637.17	1054.65	650.15	1667.38	465.36	1194.70	760.26	8230.58	816.51	9127.22
	2+	23.59	109.58	23.02	69.23	23.91	73.35	21.95	34.77	23.66	35.69	24.49	58.67	18.45	35.80	24.92	196.45	25.56	217.21
Cit. Inc.	Yes	69.00	488.80	77.70	459.47	84.83	489.54	61.05	192.30	47.23	154.20	66.03	243.90	55.11	177.45	71.21	805.12	73.22	890.09
	No	1101.38	8410.65	903.96	2805.01	937.31	2982.47	986.63	1861.84	1052.19	1874.57	1091.21	2776.15	788.87	1990.56	1423.70	16 242.92	1553.81	18 036.09
Edu.	ILL	11.12	49.87	14.61	54.08	15.75	57.38	11.31	28.87	7.82	22.49	10.12	27.28	9.53	19.13	8.46	73.67	8.21	81.27
	LWD	41.17	182.99	48.23	144.07	51.70	152.87	42.57	93.48	36.26	80.94	40.73	97.82	33.49	64.65	33.14	307.90	32.83	340.20
	PRI	234.94	1049.64	245.88	658.80	255.83	700.12	240.39	409.03	252.11	396.50	234.90	520.48	179.04	382.91	218.62	1880.31	228.83	2080.91
	SEC	279.74	1716.97	253.19	802.86	262.78	853.89	261.36	466.73	275.48	451.72	280.14	684.86	208.95	504.50	324.26	3234.30	346.68	3586.65
	HIG	395.96	3300.96	295.96	957.56	305.76	1018.13	346.97	679.56	373.62	672.43	401.39	1059.07	283.20	766.46	539.50	6408.07	591.85	7117.78
	FLD	41.87	459.48	25.96	85.02	26.75	90.22	32.11	67.96	35.31	70.94	40.30	121.62	27.12	82.04	69.09	903.86	77.48	1004.25
	SLD	158.64	2089.05	94.56	553.20	100.15	589.62	109.23	322.80	114.47	348.51	144.17	492.70	99.16	351.49	286.57	4096.94	323.60	4554.25
	PhD	3.25	55.31	1.53	11.46	1.61	12.21	1.74	6.90	2.04	8.52	2.55	11.02	1.62	7.34	7.16	109.15	8.24	121.35

Table B.7
Test for the difference between means $H_0 : \mu_{Q_i} - \mu_{POPH_i} = 0 \quad i = 1, 2, 3, 4.$

	Variables	Q1-POPH1	Q_2 – POPH2	Q3-POPH3	Q4-POPH4
Age	Age 18–84	t(3580) = −0.3036	t(2315) = −0.1980	t(3157) = 3.6312	t(3236) = −0.2348
		P-value = 0.7614	P-value = 0.8431	P-value = 0.0002	P-value = 0.8148
	Age 35–66	t(3579) = −0.3276	t(2286) = −0.9433	t(3095) = 3.5882	t(3236) = −0.2118
		P-value = 0.7432	P-value = 0.3456	P-value = 0.0003	P-value = 0.8322
	Age 67+	t(3579) = −0.4215	t(2365) = −0.5088	t(3024) = 3.1440	t(3237) = −0.2115
		P-value = 0.6734	P-value = 0.6109	P-value = 0.0017	P-value = 0.8325

(*continued on next page*)

Appendix B. Classification of excluded municipalities

We conducted further analysis on the municipalities excluded from the Prevalent One-Person Household type (POPH) in the third quartile. Specifically, we applied the classification algorithm with both two and four classes to these excluded municipalities, yielding the following results.

The algorithm achieved an accuracy of 65% for the two-class problem, while the multi-class accuracy was 35%, confirming that the exclusion of minority classes may lead to a significant loss of information regarding the diverse characteristics of municipalities within each quartile. If we consider that the accuracy for the entire dataset is 96%, this indicates that the algorithm's performance significantly drops when applied to the subset of municipalities excluded from the Prevalent One-Person Household type in the third quartile. This suggests that these excluded municipalities possess distinct characteristics that cannot be adequately captured by the classification approach.

Table B.7 (*continued*).

	Variables	Q1-POPH1	Q_2 – POPH2	Q3-POPH3	Q4-POPH4
Sex	Male	t(3580) = −0.3346	t(2324) = −0.7102	t(3093) = 3.5491	t(3236) = −0.2141
		P-value = 0.7379	P-value = 0.4777	P-value = 0.0004	P-value = 0.8305
	Female	t(3578) = −0.3881	t(2313) = −0.6571	t(3067) = 3.3313	t(3236) = −0.2189
		P-value = 0.698	P-value = 0.5112	P-value = 0.0009	P-value = 0.8267
Citizenship	Italian	t(3579) = −0.3830	t(2326) = −0.6425	t(3048) = 3.2333	t(3287) = −0.2083
		P-value = 0.7017	P-value = 0.5206	P-value = 0.0012	P-value = 0.8350
	Not Italian	t(3582) = −0.1956	t(2336) = −1.0180	t(3156) = 4.9013	t(3233) = −0.2563
		P-value = 0.8449	P-value = 0.3088	P-value = 1e−06	P-value = 0.7977
Dwelling	0	t(3578) = −0.3494	t(2317) = −0.1367	t(3121) = 3.6616	t(3236) = −0.2290
		P-value = 0.7268	P-value = 0.8913	P-value = 0.0002	P-value = 0.8188
	1	t(3581) = −0.4131	t(2359) = −0.8662	t(3059) = 3.2503	t(3237) = −0.1941
		P-value = 0.6796	P-value = 0.3865	P-value = 0.0012	P-value = 0.8461
	2+	t(3579) = −0.2457	t(2225) = −1.4961	t(3041) = 3.0769	t(3234) = −0.2344
		P-value = 0.8059	P-value = 0.1348	P-value = 0.0021	P-value = 0.8147
Car	0	t(3578) = −0.3895	t(2323) = −0.0217	t(3065) = 3.125	t(3234) = −0.24113
		P-value = 0.6969	P-value = 0.9827	P-value = 0.0018	P-value = 0.8093
	1	t(3581) = −0.3249	t(2317) = −1.2424	t(3086) = 3.6253	t(3238) = −0.1911
		P-value = 0.7453	P-value = 0.2142	P-value = 0.0003	P-value = 0.8484
	2+	t(3584) = −0.3794	t(2253) = −1.2865	t(3171) = 3.6023	t(3248) = −0.0915
		P-value = 0.7044	P-value = 0.1984	P-value = 0.0003	P-value = 0.9271
CIN	Yes	t(3577) = −0.4559	t(2730) = 2.1831	t(3070) = 1.45472	t(3244) = −0.0702
		P-value = 0.6485	P-value = 0.02911	P-value = 0.1458	P-value = 0.9440
	No	t(3580) = −0.3496	t(2289) = −0.9352	t(3086) = 3.5613	t(3236) = −0.2238
		P-value = 0.7267	P-value = 0.3498	P-value = 0.0004	P-value = 0.8229
EDU	ILL	t(3582) = −0.6215	t(2781) = 3.7804	t(3107) = 0.7165	t(3248) = 0.0935
		P-value = 0.5343	P-value = 0.0002	P-value = 0.4737	P-value = 0.9255
	LWD	t(3582) = −0.7084	t(2585) = 1.9656	t(3149) = 2.5072	t(3245) = 0.0283
		P-value = 0.4788	P-value = 0.0494	P-value = 0.0122	P-value = 0.9774
	PRI	t(3580) = −0.4441	t(2368) = −0.7798	t(3057) = 3.4679	t(3242) = −0.1521
		P-value = 0.657	P-value = 0.4356	P-value = 0.0005	P-value = 0.8791
	SEC	t(3579) = −0.3507	t(2366) = −0.8237	t(3056) = 3.3565	t(3238) = −0.1939
		P-value = 0.7258	P-value = 0.4102	P-value = 0.0008	P-value = 0.8463
	HIG	t(3580) = −0.3008	t(2323) = −1.0531	t(3076) = 3.6357	t(3235) = −0.2282
		P-value = 0.7636	P-value = 0.2924	P-value = 0.0003	P-value = 0.8195
	FLD	t(3582) = −0.2708	t(2221) = −1.2210	t(3137) = 3.6401	t(3235) = −0.2591
		P-value = 0.7866	P-value = 0.2222	P-value = 0.0003	P-value = 0.7956
	SLD	t(3576) = −0.2969	t(2159) = −0.4114	t(3091) = 2.9946	t(3234) = −0.2523
		P-value = 0.7666	P-value = 0.6808	P-value = 0.0028	P-value = 0.8008
	PhD	t(3577) = −0.2022	t(1930) = −0.9895	t(3144) = 2.8820	t(3234) = −0.2782
		P-value = 0.8597	P-value = 0.3225	P-value = 0.0040	P-value = 0.7808
BED	1	t(3621) = 0.1809	t(2106) = −3.2514	t(2977) = 3.7241	t(3252) = −0.2443
		P-value = 0.8564	P-value = 0.0012	P-value = 0.0002	P-value = 0.8070
	2	t(3626) = 0.2248	t(2115) = −3.3547	t(2993) = 3.7805	t(3252) = −0.2448
		P-value = 0.8222	P-value = 0.0008	P-value = 0.0001	P-value = 0.8067
WEA		t(3669) = 1.4285	t(2591) = −7.8214	t(2605) = 5.9043	t(3424) = −2.4016
		P-value = 0.1532	P-value = 7.539e−15	P-value = 4.002e−09	P-value = 0.01638

References

[1] De Rose A, Racioppi F, Zanatta AL. Italy: Delayed adaptation of social institutions to changes in family behaviour. Demogr Res 2008;19:665–704.

[2] Salvini S, Vignoli D. Things change: Women's and men's marital disruption dynamics in Italy during a time of social transformations, 1970–2003. Demogr Res 2011;24:145–74.

[3] Caltabiano M, Dreassi E, Rocco E, Vignoli D. A subregional analysis of family change: The spatial diffusion of one-parent families across Italian municipalities, 1991–2011. Popul Space Place 2019;25(4):e2237.

[4] Billari FC, Manfredi P, Valentini A. Macro-demographic effects of the transition to adulthood: Multistate stable population theory and an application to Italy. Math Popul Stud 2000;9(1):33–63.

[5] Billari FC, Rosina A. Italian "latest-late" transition to adulthood: An exploration of its consequences on fertility. Genus 2004;71–87.

[6] Gabrielli G, Vignoli D. The breaking-down of marriage in Italy: Trends and trendsetters. Popul Rev 2013;52(1).

[7] Reher D, Requena M. Living alone in later life: A global perspective. Popul Dev Rev 2018;427–54.

[8] Kertzer DI, White MJ, Bernardi L, Gabrielli G. Italy's path to very low fertility: The adequacy of economic and second demographic transition theories: Le cheminement de l'Italie vers les tres basses fécondités: adéquation des théories économique et de seconde transition démographique. Eur J Popul/Rev Eur Démogr 2009;25:89–115.

[9] Lesthaeghe R, van de Kaa DJ. Twee demografische transities? (two demographic transitions?). In: Bevolking: groei en krimp (population: growth and decline). Deventer, Van Loghum Slaterus; 1986, p. 9–24.

[10] Lesthaeghe R. The second demographic transition: A concise overview of its development. Proc Natl Acad Sci 2014;111(51):18112–5.

[11] Hall R. Household trends within Western Europe 1970–1980. In: West European population change. Routledge; 2023, p. 18–34.

[12] Vignoli D, Tomassini C. Introduzione. La demografia delle famiglie. In: Vignoli D, Tomassini C, editors. AISP – associazione Italiana per gli studi di popolazione. rapporto sulla popolazione. le famiglie in Italia. forme, ostacoli, sfide. Bologna: Il Mulino; 2023, p. 7–25.

[13] Cámara AD, Rodríguez-Guzmán C, Barroso-Benítez I, Morente-Mejías F. Sociodemographic analysis of an accelerated transition: The rise of solo living in Spain. Eur Soc 2021;23(1):161–89.

[14] Jamieson L, Simpson R. Living alone: Globalization, identity and belonging. Springer; 2013.

[15] Ogden PE, Hall R. Households, reurbanisation and the rise of living alone in the principal French cities, 1975–90. Urban Stud 2000;37(2):367–90.
[16] Lee YH, Han JM. The rise of single-person households and changes in consumption patterns. Korea institute for industrial economics and trade research paper, 2013, p. 4–2, It is unusual to have a number for the issue but no number for the volume. (13).
[17] Wulff M. Growth and change in one person households: Implications for the housing market. Urban Policy Res 2001;19(4):467–89.
[18] Gram-Hanssen K, Scherg RH, Christensen RS. One-person households–A growing challenge for sustainability and housing policy. In: European housing research network conference. vol. 9, 2009, p. 1–15.
[19] Williams J. One-person households–A resource time bomb? WIT Trans Ecol Environ 2005;81.
[20] Piekut M. Living standards in one-person households of the elderly population. Sustainability 2020;12(3):992.
[21] Vignoli D, Salvini S. Religion and union formation in Italy: Catholic precepts, social pressure, and tradition. Demogr Res 2014;31:1079–106.
[22] Capacci G, Rinesi F. An overview of demographic ageing in Italy. In: Ageing, lifestyles and economic crises: The new people of the mediterranean. 2017.
[23] Barbi E, Casacchia O, Racioppi F. Cause-specific mortality as a sentinel indicator of current socioeconomic conditions in Italy. Demogr Res 2018;39:635–46.
[24] Benassi F, Naccarato A. Households in potential economic distress. A geographically weighted regression model for Italy, 2001–2011. Spatial Stat 2017;21:362–76.
[25] Fritsch NS, Riederer B, Seewann L. Living alone in the city: Differentials in subjective well-being among single households 1995–2018. Appl Res Qual Life 2023;18:2065–87. http://dx.doi.org/10.1007/s11482-023-10177-w.
[26] Vitali A, Billari FC. Changing determinants of low fertility and diffusion: A spatial analysis for Italy. Popul Space Place 2017;23(2):e1998.
[27] Salvati L, Benassi F, Miccoli S, Rabiei-Dastjerdi H, Matthews SA. Spatial variability of total fertility rate and crude birth rate in a low-fertility country: Patterns and trends in regional and local scale heterogeneity across Italy, 2002–2018. Appl Geogr 2020;124:102321.
[28] Benassi F, Busetta A, Gallo G, Stranges M. Neighbourhood effects and determinants of population changes in Italy: A spatial perspective. Vienna Yearb Popul Res 2023;21:1–28.
[29] De Castro MC. Spatial demography: An opportunity to improve policy making at diverse decision levels. Popul Res Policy Rev 2007;26:477–509.
[30] Voss PR. Demography as a spatial social science. Popul Res Policy Rev 2007;26:457–76.
[31] Neumark D, Simpson H. Place-based policies. In: Handbook of regional and urban economics. vol. 5, Elsevier; 2015, p. 1197–287.
[32] Ladd HF. Spatially targeted economic development strategies: Do they work? Cityscape 1994;1(1):193–218.
[33] Fotheringham AS, Sachdeva M. On the importance of thinking locally for statistics and society. Spatial Stat 2022;50:100601.
[34] Alwang J, Siegel PB, Jorgensen SL, et al. Vulnerability: A view from different disciplines. Technical report, The World Bank; 2001.
[35] Brown K, Ecclestone K, Emmel N. The many faces of vulnerability. Soc Policy Soc 2017;16(3):497–510.
[36] Gallie D, Paugam S, Jacobs S. Unemployment, poverty and social isolation: Is there a vicious circle of social exclusion? Eur Soc 2003;5(1):1–32.
[37] Micheli GA. New patterns of family formation in Italy. Which tools for which interpretations? Genus 1996;52:15–52.
[38] De Rose A, Vignoli D. Families "all'italiana": 150 years of history. Riv Ital Econ Demogr Stat 2011;65(2):121–44.
[39] Lo Conte M, Corsetti G, De Rose A, Marsili M, Meli E. The future of the Italian family: Evidence from a household projection model. In: Schoen R, editor. The demography of transforming families. Berlin: Springer; 2023, p. 93–118.
[40] Mastroeni L, Naldi M, Vellucci P. Who pushes the discussion on wind energy? An analysis of self-reposting behaviour on Twitter. Qual Quant 2023;57(2):1763–89.
[41] Murphy KP. Machine learning: A probabilistic perspective. MIT Press; 2012.
[42] Kuhn M. Caret: Classification and regression training. Astrophys Source Code Libr 2015;ascl–1505.
[43] Therneau T, Atkinson B, Ripley B. rpart: Recursive partitioning and regression trees. 2023;4:1–9. R package.
[44] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Berlin: Springer; 2001.
[45] Vellucci P, Benassi F, Naccarato A, Gallo G. Living alone in Italian municipalities. In: Book of short papers IES 2023. Chieti: Il Viandante; 2023, p. 317–22. http://dx.doi.org/10.60984/978-88-94593-36-5-IES2023.
[46] Congalton RG. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens Environ 1991;37(1):35–46. http://dx.doi.org/10.1016/0034-4257(91)90048-B, URL: https://www.sciencedirect.com/science/article/pii/003442579190048B.
[47] Grandini M, Bagli E, Visani G. Metrics for multi-class classification: An overview. 2020, arXiv preprint arXiv:2008.05756.
[48] Eurostat. Applying the degree of urbanisation. 2021, https://ec.europa.eu/eurostat/documents/10186/11395216/DEGURBA+manual_20210120.pdf/9f338a4f-82a8-ee78-e088-4404558e5c1e?t=1611236885784.
[49] Istat. Popolazione residente e dinamica demografica. 2021, https://www.istat.it/it/files/2022/12/CENSIMENTO-E-DINAMICA-DEMOGRAFICA-2021.pdf.
[50] Benassi F, Iglesias-Pascual R, Salvati L. Residential segregation and social diversification: Exploring spatial settlement patterns of foreign population in Southern European cities. Habitat Int 2020;101:102200.
[51] Benassi F, Naccarato A, Iglesias-Pascual R, Salvati L, Strozza S. Measuring residential segregation in multi-ethnic and unequal European cities. Int Migr 2023;61(2):341–61.
[52] Buonomo A, Benassi F, Gallo G, Salvati L, Strozza S. In-between centers and suburbs? Increasing differentials in recent demographic dynamics of Italian metropolitan cities. Genus 2024;80(1):1.
[53] Howell FM, Porter JR, Matthews SA, editors. Recapturing space: New middle-range theory in spatial demography. Berlin: Springer; 2016.
[54] Barca F, McCann P, Rodríguez-Pose A. The case for regional development intervention: Place-based versus Place-Neutral approaches. J Reg Sci 2012;52(1):134–52.
[55] Reynaud C, Miccoli S, Benassi F, Naccarato A, Salvati L. Unravelling a demographic 'mosaic': Spatial patterns and contextual factors of depopulation in Italian municipalities, 1981–2011. Ecol Indic 2020;115:106356.

Federico Benassi (Ph.D.) is a researcher of Demography at the University of Naples Federico II, Department of Political Sciences. His main research interests are related to spatial demography, human population modelling, human mobility (migration and commuting), residential segregation and spatial inequalities, urban growth and regional development, foreign presence in Italy.

He is part of the Editorial Board of Spatial Demography and of International Journal of Population Studies (where acts as Section Editor for "Human Geography and Spatial Analysis"). He is currently the P.I. of the Prin – Pnrr research project "Foreign population and territory: integration processes, demographic imbalances, challenges and opportunities for the social and economic sustainability of the different local contexts (For.Pop.Ter)"

Alessia Naccarato is an Associate Professor of Statistics in the Department of Economics at the University of Roma Tre. She obtained her Ph.D. in Statistical Methods for Economics and Finance in 2000 from the University Roma Tre. Her current research focuses on functional data analysis applied to environmental and economic data, simultaneous equation models, and the integration of data from various sources, as well as spatio-temporal models.

Pierluigi Vellucci is a researcher in Mathematical Methods for Economics and Actuarial and Financial Sciences at the Department of Economics of the University of Roma Tre. He earned his Ph.D. in Mathematical Models for Engineering, Electromagnetism, and Nanosciences in 2017 at the University La Sapienza in Rome. His most recent research interests include multi-agent systems for the study of public opinion formation and methods from the theory of representation (frames, Gabor systems, and wavelets) for the analysis of financial series.

He is on the Editorial Board of PLOS ONE. He has also served as a visiting professor at the Department of Mathematics and Statistics, Florida International University (Miami).