

Spatial clusters for demand and supply of early childhood education and care services in Italy[☆]

Angela Andreella^{a,*}, Emanuele Aliverti^b, Federico Caldura^a, Stefano Campostrini^a

^a Department of Economics, Ca' Foscari University of Venice, San Giobbe, Cannaregio 873, Venice, 30121, Italy

^b Department of Statistical Sciences, University of Padova, Via Cesare Battisti, 241, Padova, 35121, Italy

ARTICLE INFO

Keywords:

Clustering
ECEC services
Lasso penalty
Multinomial regression
Supply and demand
Social services
Spatial proximity

ABSTRACT

In recent years, concern has increased about affordable and high-quality early childhood education and care (ECEC) services. Such services can help balance work and family life, increase workforce participation, promote gender equality, and increase fertility rates. ECEC services are then essential, especially for countries facing ultralow fertility rates like Italy. To face these challenges, the Italian government has included an unprecedented investment in the recovery and resilience plan financed with Next Generations EU funds to increase the supply of ECEC services and make them more equitably distributed across the country. However, socio-economic policies need a comprehensive understanding of demand dynamics and the intricate interplay between supply and demand of such services. Therefore, in this article, we propose a novel analysis of the complex combination of supply and demand of Italian ECEC services. We explore patterns of this combination in terms of groups of spatial areas sharing similar structures of service coverage and public expenditure rates in ECEC, as well as other socio-demographic and economic factors such as female employment, education, and grandparent rates. Our empirical findings confirm how Italy is characterized by several distinct territorial models and how some are shared across multiple regions. Current empirical findings suggest a lack of uniform governance across the Italian territory. Policymakers can use these results to plan, discuss, and implement area-specific welfare strategies regarding ECEC services.

1. Introduction

In recent years, there has been increasing attention to the availability of early childhood education and care (ECEC) services. A rich literature on the social and economic impact of ECEC services is present, and economists, educators, and policymakers agree on the relevance of ECEC services [1]. For example, there is shared agreement on how such facilities promote gender employment equality [2–6] and the related fight against an aging population [4,7,8]. In addition, ECEC services have a notable educational and social impact on a child's cognitive development, notably when the service offers high-quality education [9–12]. Moreover, several studies indicate that decreasing service prices and increasing available places impact the families' decision to send the child to daycare, improving the mothers' work participation [2,12–15]. It is important to emphasize that this is not

a unidirectional relationship in which a government policy of funding ECEC services impacts female employment. Instead, the effectiveness of this relationship lies in the circularity of supply and demand effects. Investments bring more availability of ECEC services in the territories, which, in turn, reduces the family commitment of women and stimulates employability, which, in turn, leads to an increase in demand for the service. The result is a virtuous circle in which these two factors regenerate each other [14,15]. Therefore, it is paramount for policymakers to comprehend the intricate interplay between the supply and demand of ECEC services. By clearly understanding the complex relationship between the availability of and need for ECEC services, one can develop social welfare policies. By prioritizing the development of policies that support the provision of ECEC services, policymakers can help ensure that all children have the opportunity to thrive and reach

[☆] Angela Andreella gratefully acknowledges funding from the grant PON 2014-2020/DM 1062 of the Ca' Foscari University of Venice, Italy. The authors acknowledge the agreement signed among the University Ca' Foscari of Venice, the National Institute of Statistics, and the Department of Family Policies at the Presidency of the Ministerial Council that supported and partially funded their work.

* Corresponding author.

E-mail addresses: angela.andreella@unive.it (A. Andreella), emanuele.aliverti@unipd.it (E. Aliverti), federico.caldura@unive.it (F. Caldura), stefano.campostrini@unive.it (S. Campostrini).

URLs: <https://angeella.github.io> (A. Andreella), <https://emanuelealiverti.github.io/> (E. Aliverti), <https://unive.it/data/persona/23921190> (F. Caldura), <https://www.unive.it/data/persona/5593107> (S. Campostrini).

<https://doi.org/10.1016/j.seps.2024.102034>

Received 20 July 2023; Received in revised form 21 July 2024; Accepted 30 July 2024

Available online 2 August 2024

0038-0121/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

their full potential and incentivize uniform governance across Italy as well as in many other European countries.

To our knowledge, no existing studies have undertaken a comprehensive analysis of multiple variables concurrently describing the supply and demand dynamics in ECEC systems, neither in Italy nor other countries. This paper aims to fill this gap, focusing on the Italian territory and proposing and discussing a methodological approach that could also be used elsewhere since several other European countries, e.g., are facing similar problems and could rely on similar data. Mainly, most research has focused on the economic impact of providing ECEC services [16,17]. Other investigations, such as Carlin et al. [18] and Del Boca et al. [19], have explored factors influencing families' decisions to enroll children in ECEC services. Additional studies have analyzed the impacts of welfare policies on the supply and demand of ECEC services [20,21]. Moreover, extensive research has been conducted on examining the availability and overall social expenditure on ECEC services [13,22–25]. In Italy, Dipartimento per le politiche della famiglia et al. [26] has presented exploratory analyses shedding light on the regional variability in the supply and demand of ECEC services, described and expanded in the following subsection.

1.1. Italian context and objectives

In a recent study, Dipartimento per le politiche della famiglia et al. [26] reported substantial variability in terms of supply, focusing on the level of coverage; such a quantity is central to our work and is defined as the ratio of available spots in ECEC structures on children under 3 years. Notably, coverage across the Italian territory is below the European average and the European Union target, which is fixed at 33%, where thirteen European Union countries have reached this target [27], with minimum coverage equal to 0.6% (Slovakia) and a maximum of 71.1% (Denmark). Focusing on the year 2019, North-East and Central Italy confirm the level of coverage above the European target (with 34.5% and 35.3% respectively), and North-West Italy is getting closer to the target (31.4%), while the South and the Islands are still far from the target (with 14.5% and 15.75% respectively). Such variability in the supply is also evident in terms of the type of services (i.e., public and private ones); for example, South Italy has a percentage of public ECEC services equal to 41.1%, while in the North East, it equals 55.1% (above the national mean equals 49.1%).

Most importantly, Dipartimento per le politiche della famiglia et al. [26] noted a substantial variability at a much finer spatial level than those canonical macro areas, such as between regions and also sub-regionally. The largest level of coverage is found in the larger municipalities like Bologna, Firenze, Roma, Milano, and Venezia; however, deep differences are observed between metropolitan and suburban areas. For example, coverage in Torino is close to 35% if the metropolitan area is considered, while coverage is roughly 25% outside the suburban area. Regarding the per-capita public expenditure, ideal situations are in regions like Trentino (a “statuto speciale” region), Valle d’Aosta, Emilia Romagna (i.e., northern regions), Lazio, and Toscana (i.e., central regions). Instead, sub-ideal conditions are observed in Calabria, Campania, Puglia, and Basilicata (i.e., southern regions). Such variability within the macro-area is also consistent if the percentage of public services is analyzed. For example, the large percentage of public services in North-East Italy is mainly due to Emilia Romagna (71.2%), Bolzano (54.1%), and Trento (76%). The remaining northeast regions (i.e., Veneto, Friuli-Venezia Giulia) have percentages of public services significantly below the national average.

Dipartimento per le politiche della famiglia et al. [26] also analyzed some aspects of the demand side of ECEC services in Italy through the sample survey “Aspects of daily life” available at <https://www.istat.it/it/archivio/91926>. In this survey, 25000 households distributed in about 800 Italian municipalities are interviewed about school, work, family and relationship life, housing and area of living, leisure time, political and social participation, health, and lifestyles. The analysis has

found that most of the children that use ECEC services have working parents ($\approx 60\%$), fathers with a high level of education ($\approx 41,6\%$), and no economic problems ($\approx 30\%$); geographical distance and the cost of ECEC services were indicated among the primary motivations for the non-participation to such services.

While the geographical variability in ECEC services supply in Italy has been well documented, recurrent supply and demand patterns still need to be explored. Our study aims to focus on this gap. Specifically, we investigate how Italy is characterized by patterns of supply and demand that do not correspond to administrative territories but are driven instead by subregional models shared across different regions. Rather than centering on welfare policies' influence on social and economic aspects, our paper seeks to identify patterns that might help policymakers formulate strategies to enhance ECEC services. The analysis proposed is especially crucial in nations such as Italy, where the lack of national and/or regional governance has resulted in significant variability in the availability of these services.

The analysis is developed at the administrative unit (*Ambito Territoriale Sociale* (ATS), i.e., intermunicipality social unit) level within the region and the province. From the perspective of political planning, the choice to focus on this micro-scale level to explore the intra-regional variability of ECEC services can be motivated by various factors. In Italy, the educational offer for children from zero to six years old consists of two segments characterized by very different developments, even if they are considered today within a single integrated offer of ECEC services. Indeed, ECEC services have historically been included in social services instead of educational services (such as primary schools). For this reason, they have long been the subject of decentralized decisions with a substantial absence of central governance. In addition to a poorly controlled and uneven proliferation of detailed regulatory acts and interventions at local government levels, this development has brought evident territorial differences in providing childcare services [26]. Regions identify the ATSs to support or replace the municipalities in managing social interventions and services (Italian law d.lgs328/2000) to reduce these micro-territorial differences in social interventions. They are formally defined as inter-municipal aggregations that handle social programming and often intersect with the scheduling of childhood education. Therefore, ATS can be considered a plausible territorial dimension to obtain a description more in line with the reality of accessibility to ECEC services, taking into account their distribution in areas larger than municipal ones and, simultaneously, bringing out any specific differences between certain areas, such as regions and provinces. Indeed, widespread and homogeneous distribution of ECEC services at the municipal level is rare. In most of them, the supply of services is concentrated in some municipalities more than in others, and it is well known that several families bring their children to a different municipality due to the lack of available places in the municipality of residence [e.g., 28].

The main objective of this work is to find, investigate, and describe clusters of ATSs that are similar in terms of supply and demand structure. We allow such clusters to span different regions, characterizing the shared services patterns across the territory and defining sub-regional welfare models. We focus on 2019 data on ECEC services from the ISTAT (the Italian National Institute of Statistics) survey “Survey of early childhood education and care services” that describes the supply side and its structure [29]. These data are integrated with cultural and social indicators that could characterize the structure of the demand, such as the fertility rate, the presence of family support (e.g., grandparents), and the parents' education, from the permanent census of 2019 [30]. More specifically, the analysis is divided into two steps. First, a clustering method with geographical penalization is used to identify groups of ATSs that share similar patterns in terms of the supply and demand combination, using a regionalization method called SKATER (Spatial ‘K’luster Analysis by Tree Edge Removal, Assunção et al. [31]). A distance-based clustering algorithm allows for finding clusters of ATSs that also belong to different regions. At the same time,

the additional spatial constraints facilitate the estimation of smooth clusters across neighbor ATSS, improving the interpretability of the estimated structure. As a second step, we fit a penalized multinomial regression model [32,33] to link the clustering membership with the socioeconomic indicators that characterize the socioeconomic level of each ATS; this approach allows one to select the factors that influence the probability of belonging to a specific cluster and measure their effect, thereby providing a more complete description of the phenomena under investigation.

In summary, the contributions of this paper to the literature on ECEC services are threefold: (1) linkage of high-dimensional data describing the supply and demand combination of ECEC services (i.e., the ISTAT survey “Survey of early childhood education and care services” [29] and the permanent census, ISTAT [30]) at a proper spatial level (i.e., ATS level); (2) analyze with suitable statistical methods the multidimensional structure of this combination (i.e., spatial clustering approach [31] and penalized multinomial regression [32,33]), offering a valuable methodology for similar studies, also in other countries; (3) identify sub-regional models that characterize the supply and demand dynamics of ECEC services in Italy, highlighting the importance of discovering and understanding these patterns for policymakers to formulate effective and targeted policies.

The paper is organized as follows. Section 2 introduces the variables analyzed coming from the two datasets described before, i.e., the ISTAT survey “Survey of early childhood education and care services” [29] and the 2019 permanent census [30] and the theory behind the statistical methods utilized. In particular, Section 2.1 shows some descriptive statistics, while Section 2.2 introduces the SKATER regionalization method [31]. Finally, Section 3 outlines the results of the spatial clustering analysis and penalized multinomial regression model.

2. Data description and statistical analysis

In this work, two data sources are used to capture the complex structure of the demand and supply of Italian ECEC services. We concentrated on analyzing data from the year 2019, as our primary objective is to provide an initial overview of ECEC services in Italy and demonstrate how this analysis can serve as a valuable tool for policymakers. However, conducting a follow-up analysis would be of great interest and is planned for future research.

As the introduction outlined, we considered the ATSS as spatial statistical units. The Italian territory is divided into 20 regions, each divided into provinces, which are, in turn, divided into municipalities, then gathered into ATSS, particularly for the planning and coordination of social services. For example, the Veneto region is divided into 7 provinces and 593 municipalities. Table 1 shows each Italian region's number of ATSS and municipalities; current data comprise $n = 606$ ATSS across 20 regions.

The supply side of ECEC services is described using data from the ISTAT survey “Survey of early childhood education and care services” [29], while the demand side by data from the 2019 permanent census [30]. Table 2 briefly describes the variables used in the analysis. The first two (i.e., coverage and per capita public expenditure rate) represent the supply side, and the remaining variables describe the demand side. The latter variables were chosen to define possible solutions that can be roughly considered an alternative to ECEC services (e.g., babysitters, grandparents, extended families) and to outline the socio-economic characteristics of the population of each ATS.

2.1. Descriptive statistics

We propose some exploratory analysis to illustrate the data variability; refer to Appendix A for a more complete analysis.

Fig. 1 shows the relationship between the coverage and the female employment rate (left panel), grandparents rate (center panel), and the Social and Material Vulnerability Index (IVSM) (right panel) divided

Table 1

Number of ATSS and municipalities for each Italian region.

Region	N. ATSS	N. municipalities
Piemonte	49	1221
Valle d'Aosta/Vallée d'Aoste	5	74
Lombardia	88	1567
Trentino-Alto Adige/Südtirol	17	355
Veneto	21	593
Friuli-Venezia Giulia	18	223
Liguria	18	236
Emilia-Romagna	38	361
Toscana	26	301
Umbria	12	92
Marche	23	245
Lazio	37	378
Abruzzo	24	305
Molise	7	136
Campania	57	552
Puglia	45	259
Basilicata	9	131
Calabria	32	411
Sicilia	55	390
Sardegna	25	532

by Italian macro areas, i.e., North-East (Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Emilia-Romagna), North-West (Valle d'Aosta, Liguria, Lombardia, Piemonte), Center (Toscana, Umbria, Marche, Lazio), South (Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria) and Islands (Sicilia, Sardegna). The black lines represent the corresponding fitted local polynomial regression.

The left panel of Fig. 1 highlights a positive relationship between coverage and female employment rate, as the existing literature suggests [2,5,6]. In particular, northern areas have higher coverage and participation of females at work, starkly contrasting the situation in southern Italy [26]. In the middle panel of Fig. 1, we observe a positive relationship between coverage and grandparent rate with variable strength. However, the grandparent rate should be considered with caution. It was built to represent a viable alternative to daycare, but due to a lack of data, the index was built by considering the total population of retirees. Therefore, it also includes the older population, who can hardly support children. Moreover, this index is influenced by the region's demographic characteristics, i.e., high values in the areas characterized by an older population, such as Liguria and Sardegna. Finally, it is interesting to note the relationship between coverage and the IVSM index, reported in the right panel of Fig. 1, where it is evident that Italian areas with low social and economic levels are deprived of ECEC services. As found by numerous literature (e.g., [26,35–37]), ECEC services are often affordable only for families with specific economic and social characteristics, thus causing a critical social gap.

The four panels of Fig. 2 show the coverage (top-left), female house (top-right), commuter (bottom-left), and fertility (bottom-right) rates at ATS spatial level; thick black segments highlight the regional borders. Looking at the top panels of Fig. 2, there are clear signs of the north-south gradient regarding coverage and female house rate. North Italy is characterized by a proportion of females at home that is smaller than South Italy and also a larger coverage rate, as previously seen in Fig. 1; clearly, the rate of females at home is directly related to the female employment rate analyzed in Fig. 1. However, Fig. 2 provides further insights into the strong variability observed in these two rates at the regional and sub-regional levels, as discussed in the Introduction. For example, some ATSS in the North (e.g., in Emilia Romagna) report larger coverage than the nearby ATSS, and similar patterns are observed in Sardegna and Abruzzo. In contrast, the rate of females at home seems more spatially homogeneous.

The bottom-left panel of Fig. 2 indicates that the commuting rate is greater around large metropolitan areas. This index is meaningful because it suggests families' possible displacement and, therefore, a different need for educational services regarding the lease. Commuter

Table 2

List of the variables analyzed. The first two variables describe the supply side of ECEC services, while the remaining outline the demand side.

Variable	Description
Coverage	Number of daycare places/Number of resident children between 0–2 years old
Per capita public expenditure rate	(calculated on the resident population between 0–2 years old)
Female employment rate	Number of resident females working/number of resident females between 20–64 years old
Female house rate	Number of resident females not working at home/number of resident children between 0–2 years old
Commuter rate	Number of commuters workers outside the municipality /number of workers
Male educational qualification rate	Number of males with a degree higher than high school degree/number of males over 20 years of age
Female educational qualification rate	Number of females with a degree higher than high school degree/number of females over 20 years of age
Foreign rate	Number of foreign residents/number of residents
Grandparent rate	Number of resident retired persons/number of resident children between 0–2 years old
Babysitter rate	Number of resident females not working (studying) between 15–25 years old/number of resident children between 0–2 years old
Number of members in the household	
Fertility rate	Number of resident children with age 0/ number of resident females between 15–49 years old.
Social and Material Vulnerability Index (IVSM)	It comprises seven different indicators capturing household size, education level, employment status, welfare distress, and severe crowding condition in the housing situation. For a more detailed description, please refer to [34]

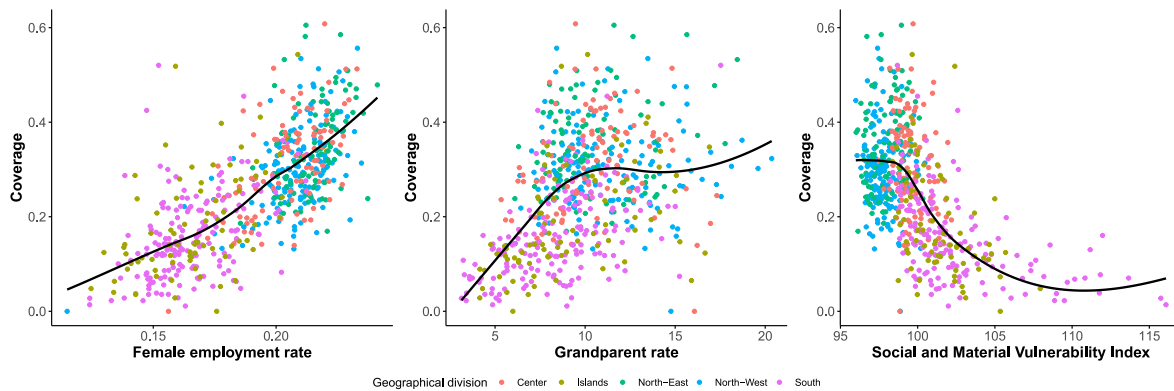


Fig. 1. Scatter plots between coverage and the female employment rate (left), grandparents rate (center), and the Social and Material Vulnerability Index (IVSM) (right) divided by macro areas (Center, Islands, North-East, North-West, and South). The black lines represent the fitted local polynomial regression.

parents will probably choose a daycare near work, for example, in another ATS rather than in the place of residence or near the place of residence [28]. It is well-known in the literature that the probability of a female taking a job decreases as the work travel time increases. Females are more likely to take flexible employment close to home to balance work and private life more easily [38,39]. This again leads to gender inequality, which policymakers need to consider by, for example, creating daycare places in those areas marked by commuting workers. Finally, the right bottom panel of Fig. 2 represents another critical aspect of the ECEC services demand, the fertility rate. Focusing on the national scale, the gap between the North and South is smooth. However, we can point out that the fertility rate of Südtirol, known for its innovative family welfare (e.g., social farming), is among the highest in Italy [40]. While this fact suggests that an efficient social policy can help to increase the birth rate of a population [41], the relationship between ECEC availability and fertility is likely to be complex and influenced by many more factors such as personal preferences, economic considerations, and cultural norms [26].

Overall, there is much evidence of regional and sub-regional variability in the supply and demand of ECEC services. Clearly, these descriptive statistics do not provide sufficient information on the behind complex relationships, and proper statistical models are necessary to achieve this aim.

2.2. Spatial clustering

We propose a cluster analysis to estimate groups of ATSs that are similar in terms of the observed characteristics outlined in Section 2.1. Traditional clustering approaches (e.g., *k*-means [42] and hierarchical clustering [43]) minimize the within-group dissimilarity and maximize the dissimilarity between groups. Often, such similarities are expressed in terms of distances. However, these methods do not consider spatial contiguity, which is a crucial aspect to be accounted for when modeling territorial services; indeed, we expect that spatially contiguous ATSs might share similar levels of services, and this information should be included in the modeling approach. Due to this, we consider clustering algorithms for spatial data that can exploit the available spatial information in constructing the clusters; refer to Xu and Tian [44] for a complete review of clustering methods. A possible method for defining homogeneous spatial clusters following certain variables is regionalization [45]. This technique of spatial statistics is used in various disciplines, and it has been used to capture environmental [46], socio-economic [47], epidemiological [48] and climatic [49] spatial clusters. Regionalization aims to maximize the within-cluster similarity and minimize the between-cluster similarity, taking into account the spatial information of the data. These methods are formulated as constrained

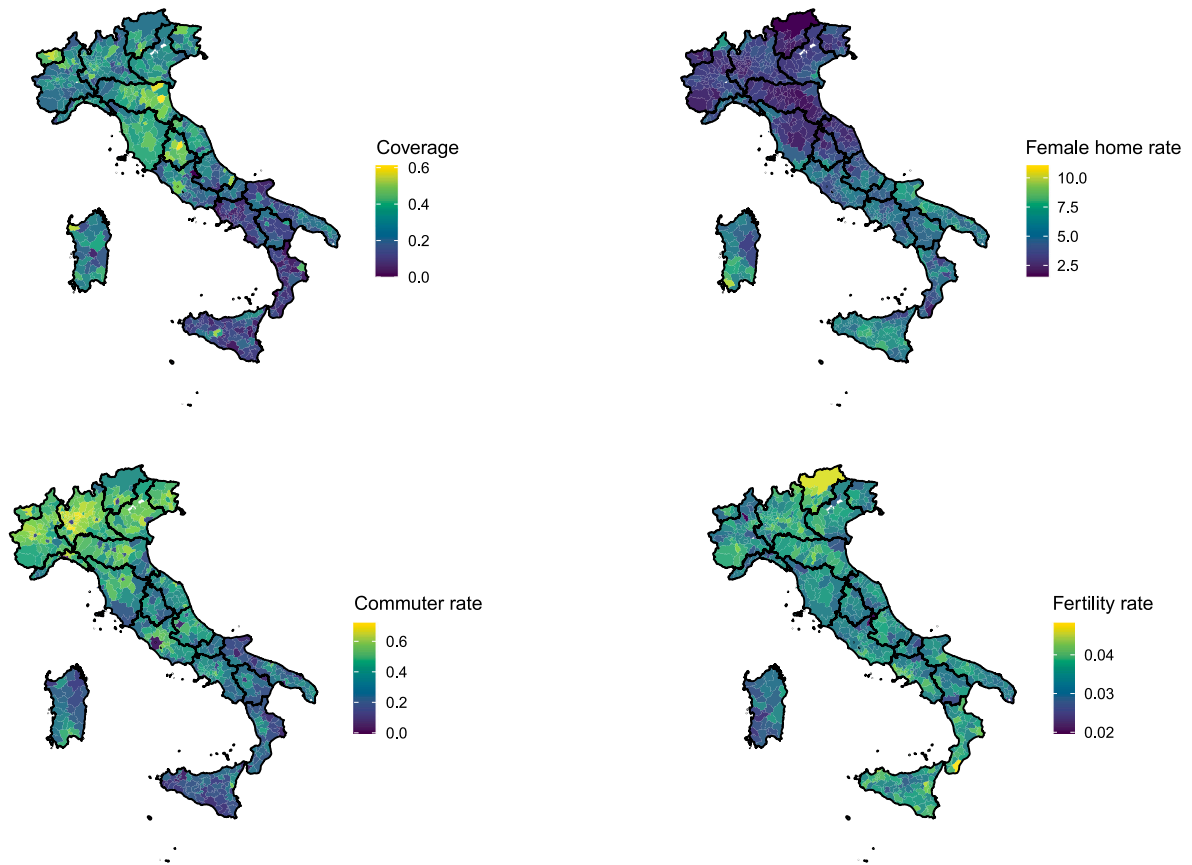


Fig. 2. Geographical map representation of the coverage rate (left top), female home rate (right top), commuter rate (left bottom), and fertility rate (right bottom) at the ATS spatial level. The thick black segments represent the regional borders.

optimization problems, where the constraint is based on the spatial contiguity of the spatial units.

Let us consider the $p = 13$ variables of Section 2 measured for each spatial unit (ATS) as $x_i = (x_{i1}, \dots, x_{i13})$ with $i = 1, \dots, 606$. The regionalization approaches aim to find k groups of spatial units, where units in the same group are characterized by similar values of x and units in different groups by different values. Each spatial area is described as a polygon with vertices and boundaries. The spatial constraint is based on the spatial neighborhood structure defined by geographical adjacency: the spatial areas share at least one boundary or vertex (i.e., queen approach [50]). The spatial constraint can be relaxed by summarizing the spatial units by their centroids. We then define that two spatial areas i and j are spatially contiguous if and only if the distance d_{ij} (e.g., Euclidean distance) is less than a fixed maximum distance $r \in \mathbb{R}^+$. The elements of the adjacency matrix $W = [w_{ij}]$ with dimensions 606×606 are then defined as

$$w_{ij} = \begin{cases} 1 & \text{if } d_{ij} \leq r \\ 0 & \text{if } d_{ij} > r \end{cases} \quad (1)$$

and are used to construct the graph. Here, the distance is computed considering the latitude and longitude of the centroids for each spatial unit i . Because we are considering relatively small areas with fairly regular shapes, the centroid represents a reasonable summary measure of the spatial area described by the polygon.

Leveraging on the SKATER algorithm [31], the spatial units summarized by their centroids can be modeled as a node in an undirected graph [51] represented via the adjacency matrix of Eq. (1). Therefore, the spatial constraint of the SKATER algorithm is included when the initial graph is constructed, i.e., when the adjacency matrix W defined in Eq. (1) is determined. The spatial clusters are then defined as

connected subgraphs that minimize the within-cluster heterogeneity, computed by associating a cost or weight c_{ij} to each edge that connects the nodes (i.e., spatial units) i and j . These weights are based on the squared Euclidean distance between locations i and j with respect to their attribute vectors x_i and x_j . The within-clusters similarity is then described as the Euclidean squared distance between the location attributes in cluster k and the cluster means of these attributes. The SKATER algorithm uses the minimum spanning tree approach to reduce the graph complexity. After constructing the corresponding minimum spanning tree, the SKATER algorithm prunes the tree for the desired number of clusters to minimize the within-cluster variability; refer to Assunção et al. [31] for further details.

After clustering the spatial units, a multinomial regression model with ℓ_1 penalization on the coefficients [32,33] is fitted on the clustering indicators against the supply and socio-demographic factors. Such an approach allows explicitly linking the probability of being in a specified cluster with the covariates described in Table 2. The regularization of the coefficients notably improves interpretation and deals with multicollinearity since only a small number of variables are selected for characterizing the membership to each cluster.

3. Results

Fig. 3 describes the clusters created by the SKATER algorithm imposing $k = 15$ clusters. The value of k was selected via the silhouette index. Fig. 4 shows the silhouette values across several k values (i.e., $k \in \{5, \dots, 20\}$) and r spatial constraints (i.e., $r \in \{15.5 \times 10^4, 55.5 \times 10^4, 95.5 \times 10^4, 175.5 \times 10^4, 10 \times 10^9\}$). Fig. 4 suggests that $k = 15$ provides a reasonable choice even across different values of r ; also, a number of clusters $k > 15$ complicates their interpretation and, therefore, should

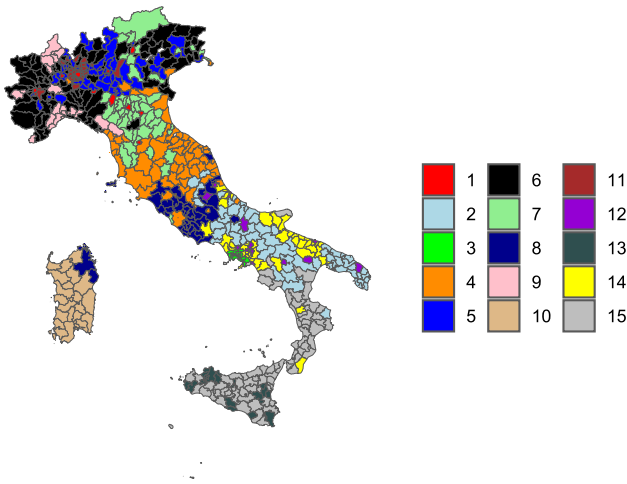


Fig. 3. Geographical representation of the 15 clusters created by the SKATER algorithm with flexible spatial constrain.

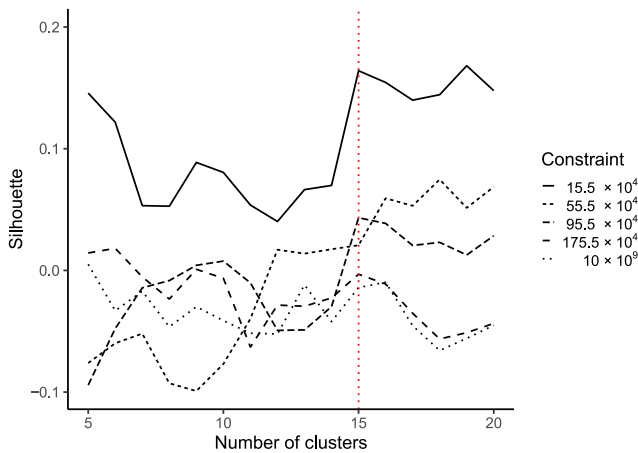


Fig. 4. Silhouette values considering several numbers of clusters k and several values of spatial constraint r . The black solid line corresponds to the spatial constraint presented in this paper, i.e., $r = 15.5 \times 10^4$.

be justified with care. The black solid line represents the scenario of our analysis related to the minimum value r to have a connected graph. As r increases, the spatial constraint is then relaxed. We can note how our case (i.e., $r = 15.5 \times 10^4$) outperforms the other situations in all cases in terms of silhouette. This result confirms that geographic constraint in a clustering analysis on data such as those analyzed improves the quality of clusters in terms of cohesion.

While comprehending the data from 15 clusters is challenging, a detailed analysis can benefit local policymakers. Here, we provide a simplified summary. Therefore, to better understand the significance of the information brought by the proposed approach, we have conducted further analyses, identifying three groups composed of multiple clusters and four clusters that instead have different characteristics and deserve dedicated analysis. Figs. 5, 6, 7, 8 describe these groups of clusters. Each plot reports the ATSS involved in the cluster, and bar plots describe the averages of the scaled variables within the clusters. Table 3 shows the unscaled averages of the variables defined in Table 2 within each cluster and at the national level. After identifying the groups, a penalized multinomial regression model for the cluster membership on the socio-economic factors is estimated to understand the most important determinants of each cluster, selecting the amount of penalization via 10-folds cross-validation. Cluster 8 (i.e., the one colored in dark blue in Fig. 3) is considered the reference category for multinomial

regression since it represents the cluster closer to the national average. It covers mainly the Lazio region without the Roma municipality. This section includes details on the interpretations of the most relevant coefficients, while the complete estimates are reported in Table 4 in Appendix B.

The research aims to understand national and interregional variations in the interplay between the supply and demand of ECEC services, highlighting recurrent patterns. We then interpret the results from the clustering analysis and the penalized multinomial regression, emphasizing the supply and demand variables that define each cluster and presenting potential recommendations for policymakers.

Fig. 5 describes the “South Italy” group composed of clusters 2, 13, 14, 15 characterized by low coverage, high social and material deprivation, a large number of components in the family, and a low female employment rate. However, cluster 2 (covering mainly the southern Appenini region and the Salento areas) has a better ECEC service situation with respect to the other three in terms of coverage, female employment rate, and social and material deprivation as evident from the left and right plots of Fig. 1. This cluster is also characterized by some possible alternatives to the ECEC service, i.e., babysitters and grandparents. We can also note that cluster 15 includes the Sicilia region without the main municipalities, while cluster 13 comprises the Sicilian urbanized areas. In Sicilia, the situation of the main municipalities is worse than in suburban areas.

Focusing on the penalized multinomial regression’s results (Table 4, Appendix B), the female employment rate negatively affects the probability of being in clusters of the “South Italy” group (i.e., Fig. 5). This reflects the scenario described in the left plot of Fig. 1 and in the top right spatial plot of Fig. 2. In addition, the main municipalities of Sicilia (i.e., cluster 13) and the area covering mainly the Salento and Appenino (i.e., cluster 2) are characterized by a low value of commuters and foreign (e.g., log relative risk with respect to cluster 8 decreases by approximately 2 for a one-unit increase in foreign rate). The south of Italy, without the area of Appenini and Salento (i.e., cluster 2), lacks coverage. The log relative risk with respect to cluster 8 decreases by approximately 2 for a one-unit increase in coverage.

Even with an expanding supply, the demand for ECEC services in South Italy might continue to be low if access is impeded by factors like high tuition fees, geographical distance, and perceived low quality of service. Therefore, policies should be designed to ensure widespread accessibility of ECEC services in these areas while simultaneously increasing the overall supply. This approach is crucial in overcoming cultural barriers to service utilization and relieving mothers from the responsibility of early childcare, which can significantly impact employment rates.

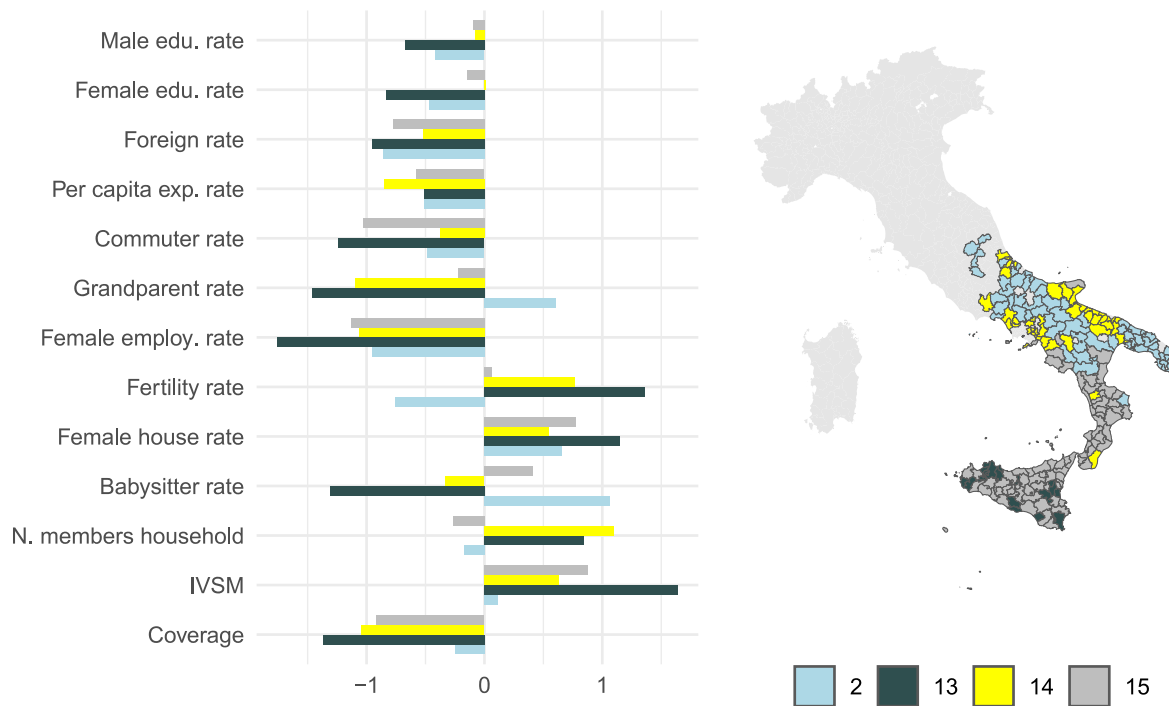
The second group of clusters (called “Center-North Italy”) comprises ATSS from the North and the Center, i.e., clusters 4, 7, and 11 represented in Fig. 6. This group of clusters is characterized by high coverage and female employment rate, as well as low social and material vulnerability index. In addition, clusters 4 and 11 (covering the suburban areas of Milano, the Roma municipality, and some areas of the Marche, Toscana, and Lombardia) are also distinguished by a high level of female and male education. Cluster 7 (covering the Bolzano and Emilia-Romagna areas) is characterized by high foreign and per-capita expenditure rates. Cluster 11 includes ATSS outside the main Italian municipalities like Torino and Milano, where the commuter rate is large and characterized by many workers who live in the suburban areas but work in the city as clear from the top left plot in Fig. 2.

The probability of being in the “Center-North Italy” group, in particular, the area covering part of the Emilia Romagna and the province of Bolzano (i.e., cluster 7), is positively affected by coverage, per capita expenditure rate, and negatively by the female house rate as expected looking at the exploratory plots in Figs. 1 and 2. For example, the log-risk ratio increases by roughly 2 (at least) for a one-unit increase in the per capita expenditure rate and decreases by roughly 2 (at least) for a one-unit increase in the female house rate.

Table 3

Averages of the unscaled variable described in Table 2 stratified by cluster represented in Fig. 3 and at the national level (first column).

Variable	Italy	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Coverage	0.255	0.419	0.219	0.077	0.366	0.266	0.293	0.395	0.236	0.284	0.280	0.358	0.243	0.097	0.134	0.142
Per capita exp. rate	678.134	2311.934	346.207	116.934	1119.004	506.171	750.175	1732.483	336.883	685.582	482.417	880.669	777.702	286.622	214.607	331.926
Female employ. rate	0.191	0.228	0.170	0.147	0.211	0.206	0.207	0.221	0.192	0.198	0.189	0.219	0.194	0.147	0.164	0.163
Female house rate	4.345	2.599	5.390	5.703	3.631	3.337	3.602	2.573	4.734	4.611	6.030	2.850	4.863	5.874	5.489	5.501
Commuter rate	0.392	0.209	0.301	0.375	0.338	0.551	0.514	0.465	0.421	0.492	0.293	0.617	0.197	0.201	0.340	0.230
Male edu. rate	0.093	0.184	0.087	0.082	0.111	0.073	0.084	0.096	0.091	0.077	0.069	0.124	0.169	0.075	0.091	0.091
Female edu. rate	0.107	0.188	0.102	0.095	0.128	0.085	0.095	0.113	0.107	0.086	0.096	0.130	0.186	0.084	0.106	0.106
Foreign rate	0.239	0.574	0.127	0.121	0.317	0.333	0.253	0.370	0.325	0.216	0.079	0.332	0.174	0.146	0.153	0.139
Grandparent rate	9.844	9.363	10.986	4.418	11.416	8.670	11.879	9.913	8.811	16.216	12.475	9.435	9.454	5.475	6.933	8.665
Babysitter rate	1.488	1.323	1.711	1.362	1.534	1.265	1.487	1.348	1.437	1.707	1.741	1.410	1.705	1.260	1.476	1.536
N. members household	2.304	2.081	2.298	2.848	2.285	2.367	2.160	2.294	2.240	1.997	2.226	2.301	2.351	2.485	2.516	2.297
Fertility rate	0.035	0.037	0.032	0.039	0.033	0.038	0.034	0.037	0.034	0.029	0.029	0.036	0.032	0.040	0.036	0.035
IVSM	99.847	98.883	100.159	109.198	99.170	97.840	97.638	98.530	100.214	97.872	99.288	97.271	100.045	103.937	101.418	102.027

**Fig. 5.** “South Italy” group composed of clusters 2, 13, 14 and 15 represented by different colors. Left plot: bar plots representing each cluster’s mean values of the scaled variables. Right plot: geographical representation of the clusters 2, 13, 14, and 15.

These areas of the country are among the most virtuous in Italy, where the historical development of public ECEC services has resulted in a relatively high and evenly distributed supply in the territories and where female labor market participation is among the highest in the country. The “Center-North Italy” group serves as a beacon for potential best practices that could be adopted nationwide. Understanding the factors contributing to their success could inform targeted strategies for replicating this estimated pattern in other regions.

Fig. 7 represents the “North Italy” group, consisting of clusters 5, 6, and 9, which cover most of the northern regions except the Milano and Torino and other northern municipalities. Low female and male educational qualification rates, high commuter rate, a low social and material vulnerability index, and medium-high coverage characterize these areas of North Italy. These ATSs have a relatively good situation; however, cluster 9 (i.e., part of the Liguria and northern Piemonte regions) probably can leverage more alternatives to ECEC services than the others, having high grandparents and babysitter rates. Furthermore, these clusters can be classified as the comparatively less prosperous northern ATSs, with satisfactory social and economic conditions but still inferior to the “Center-North Italy” group (i.e., Fig. 6). Similarly

to the “Center-North Italy” group, this group encompasses ATSs located outside the primary municipalities, distinguished by a significant commuter rate.

The probability of being in clusters covering the suburban areas around Milano (i.e., clusters 5 and 11) and Torino (i.e., cluster 6), the Varese, Domodossola, Como areas (i.e., cluster 9) increases along with the commuter rate and decreases along with the female house rate. In fact, commuter workers are prevalent in most northern-west suburban areas (except for the main municipalities such as Milano and Torino, represented by cluster 1). Again, we can check the distribution of commuter workers by looking at Fig. 2 (bottom left plot).

In these areas, ECEC services must be focused on families composed of commuter working parents. One potential policy approach could involve directing childcare resources toward the municipalities where parents work, encouraging cooperation among local governments to fund educational services (or even encouraging the creation of daycare centers as part of corporate welfare), and enhancing the usability of these services. Providing flexible hours for dropping off and picking up children and extending afternoon hours can significantly boost participation in ECEC services among working families.

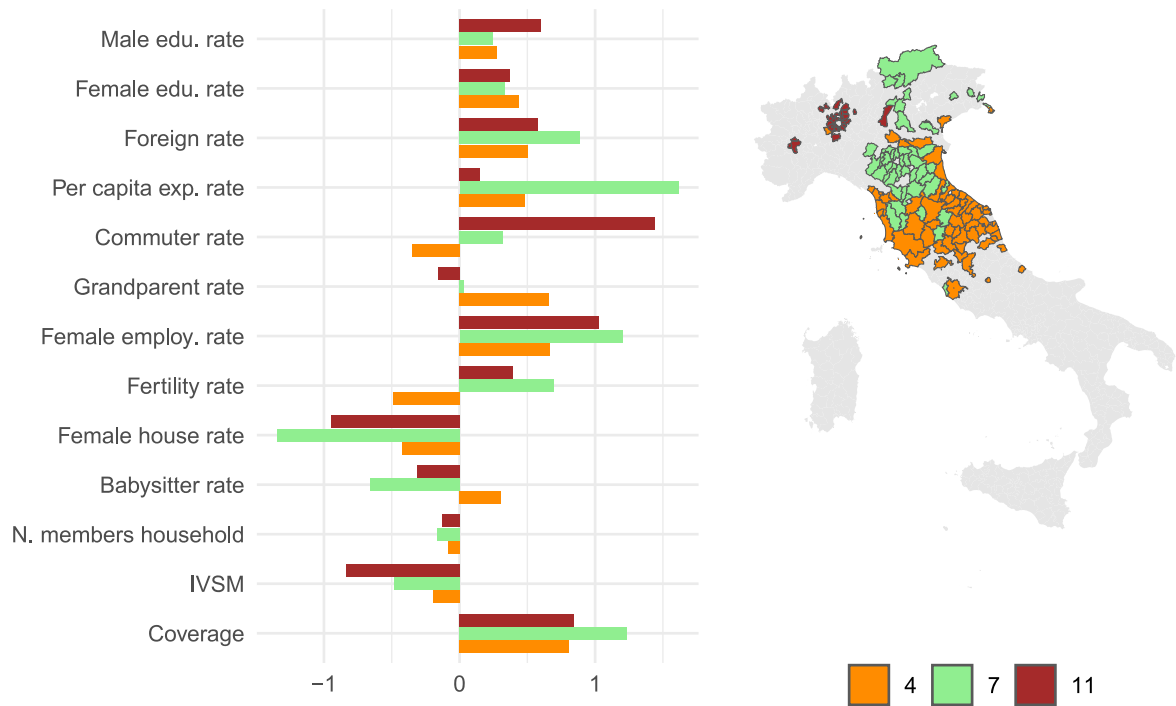


Fig. 6. “Center-North Italy” group composed of clusters 4, 7 and 11, represented by different colors. Left plot: bar plots representing each cluster’s mean values of the scaled variables. Right plot: geographical representation of the clusters 4, 7 and 11.

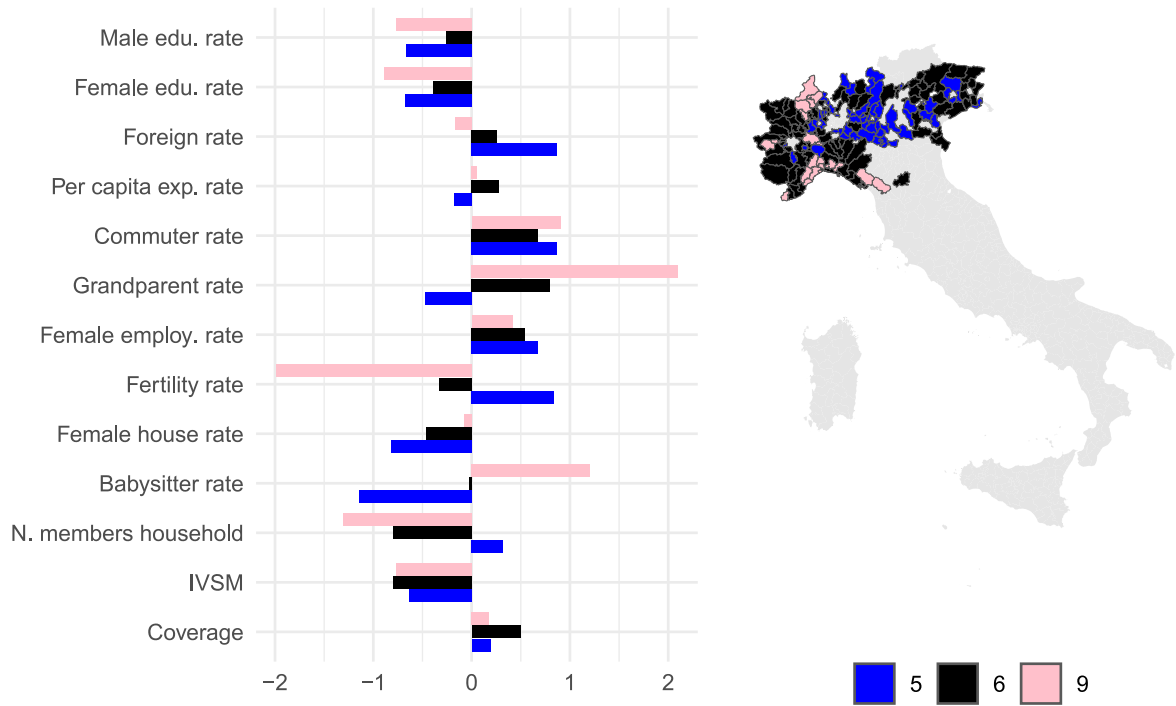


Fig. 7. “North Italy” group composed of clusters 5, 6 and 9, represented by different colors. Left plot: bar plots representing each cluster’s mean values of the scaled variables. Right plot: geographical representation of the clusters 5, 6 and 9.

Finally, clusters 1, 3, 10 and 12 are described in Fig. 8 but they should be analyzed separately. Cluster 10 includes only the Sardegna region, characterized by an old population (i.e., low fertility rate and high grandparents rate). Instead, cluster 1 comprises the main municipalities of North Italy, e.g., Milano and Torino. A high level of education characterizes this cluster for both males and females due to the presence of major universities. Since this cluster is composed of

ATs placed in big cities, the foreign rate is also high, and the commuter rate is low, i.e., these cities are attractive to foreigners, and workers do not move to get to work. Finally, these ATs have a large coverage of ECEC services, as well as good participation of females in the labor force. Cluster 3 includes the area around the city of Napoli. It is the most deprived one and shows large values of social and vulnerability index, low coverage, and low rates of alternative ECEC services such

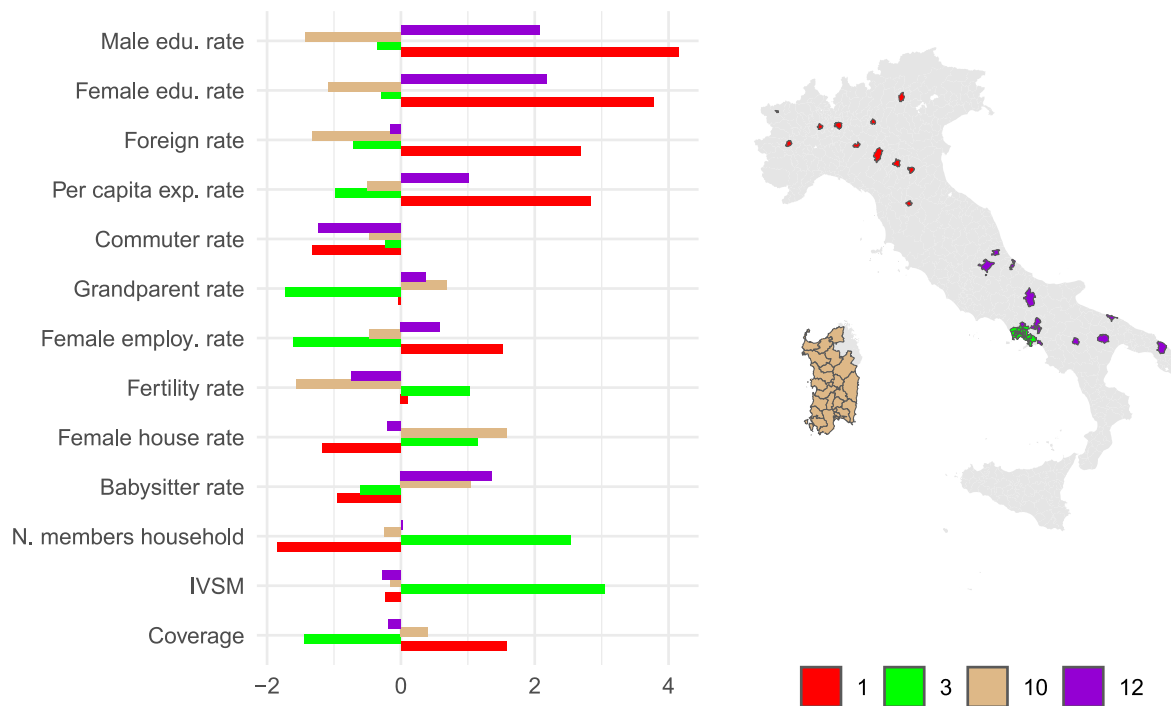


Fig. 8. Clusters 1, 3, 10 and 12 represented by different colors. Left plot: bar plots representing each cluster's mean values of the scaled variables. Right plot: geographical representation of the clusters 1, 3, 10 and 12.

as babysitters and grandparents. Finally, cluster 12 is composed of ATSS placed on central and southern municipalities characterized by the presence of several major universities. In fact, we can observe a high education rate for both females and males and a possibility of employment locally, observing a commuter rate lower than the national average.

The foreign and fertility rates affect the probability of being in the region of Sardegna (i.e., cluster 10). The log relative risk ratio decreases by roughly 2 for one-unit increases in the fertility rate. In fact, Sardegna suffers from an accentuated process of population aging. A similar situation is observable in Liguria and northern Piemonte (i.e., cluster 9) and the Appennino and Salento areas (i.e., cluster 2). In these areas, more so than in others (although Italy as a whole is facing extremely low fertility rates), the coverage of ECEC services is witnessing a significant increase. However, this growth is not due to a rise in supply but rather stems from an overall decline in the number of children. To prevent the risk of under-utilization of ECEC services resulting from reduced user numbers and the prevalence of informal care provided by grandparents and relatives, it becomes crucial to establish conditions that encourage higher participation rates among families. This can be achieved by reducing service utilization barriers and improving accessibility. Additionally, it is necessary to implement robust birth policies alongside efforts to expand the overall supply of ECEC services.

Instead, cluster 1 (i.e., big northern municipalities) is denoted by a negative female house rate (i.e., the log relative risk decreases by roughly 3 for a one-unit increase in the female home rate). Cluster 3 (i.e., Napoli area) is characterized by a positive number of members in the household (i.e., the log relative risk increases by roughly 5 for a one-unit increase in the number of members in the household). Cluster 12 (i.e., universities cities) is denoted by a significantly positive effect of the male educational qualification rate (i.e., the log relative risk increases by roughly 3 for a one-unit increase of the male educational qualification rate) and by a negative effect of the commuter rate (i.e., the log relative risk decreases by roughly 2 for a one-unit increase of the number of the commuter rate). These specific features are important since policymakers can consider these characteristics

to address design welfare policies, e.g., cluster 12 could implement policies to promote the birth rate among young couples attending universities. This can be accomplished by facilitating residency choices and providing ECEC services not only for local students and graduates but also for those coming from outside the region.

These analyses clearly reveal the variability in the combination of supply and demand for ECEC services in Italy. Drawing on the analyses presented here, it becomes possible to design public policies specifically suited to each territory's distinct characteristics. These policies should go beyond considering the mere presence of ECEC services and consider the socio-demographic and economic characteristics of the local population. These factors are crucial in shaping public policies that successfully promote and increase family participation rates in these services.

4. Conclusions

Considering the entire Italian territory, the average coverage rate of ECEC services stands at 27.2 places for every 100 toddlers, below the European minimum target fixed at 33 authorized places for every 100 children and taken up by the Italian National Recovery and Resilience Plan (NRRP) as the goal to be achieved. If we disaggregate the national situation, great differences emerge between territorial divisions, with large geographical areas such as North-East and Central Italy presenting on average coverage above the European target, the North-West close to the target, while the South and the Islands are still far from the target. More in-depth analysis has shown that differences in supply levels are present not only between macro areas and regions but also within them, e.g., peripheral areas versus urban centers and mountainous areas versus non-mountainous ones. This variability appears to stem from historical factors related to demand dynamics (e.g., declining birth rates, female labor participation) and from local and regional government policies in developing social and educational services. The lack of effective national and, in many cases, regional governance has hindered uniform and widespread development. Only some regions have a spatially uniform distribution of services; in most (including many areas in the North), the supply of services is concentrated in

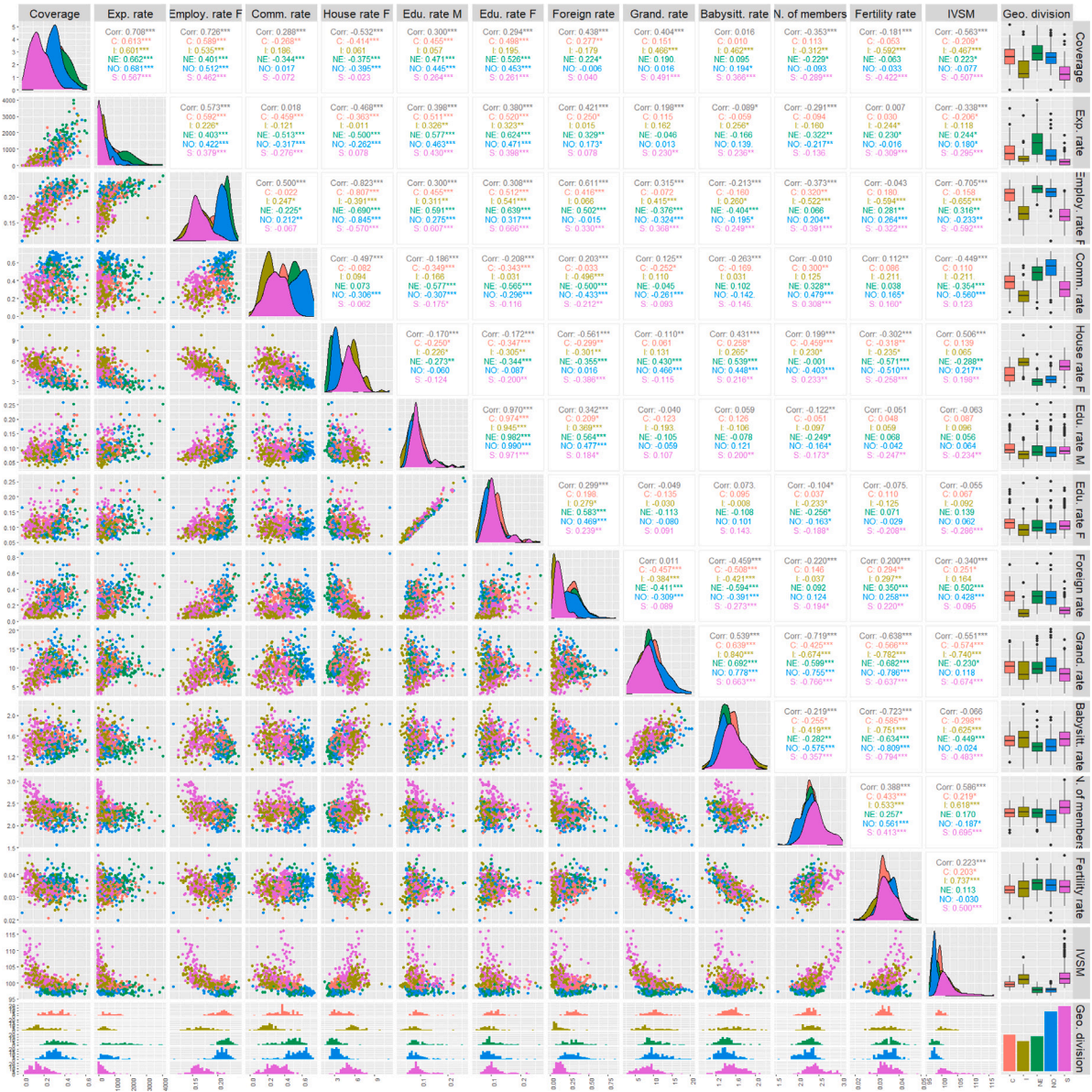


Fig. 9. Exploratory plots (i.e., scatter plots, histograms, and boxplots) for each variable described in Section 2 stratified by macro areas (Center, Islands, North-East, North-West, and South).

Table 4
Estimated coefficients from penalized multinomial model.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 15
Intercept	-3.138	3.477	-6.173	3.546	0.877	3.160	-0.512	4.375	-1.713	-2.219	-2.173	-0.376	-4.087	2.443	2.511
Coverage	0.000	0.350	-0.427	0.891	0.915	0.000	0.847	0.526	-0.078	0.000	0.157	-0.562	-0.424	-1.514	-1.338
Per capita expenditure rate	1.080	-1.247	0.000	0.356	-0.083	0.147	1.905	-1.836	0.000	-0.492	0.000	0.000	0.000	-1.027	0.000
Female employment rate	0.000	-2.653	0.000	1.297	-0.250	0.155	0.985	0.000	0.578	2.522	0.000	0.000	-3.023	-2.101	-1.515
Female house rate	-2.637	-0.749	0.000	-0.117	1.057	-0.032	-1.750	1.029	0.542	1.462	-1.401	0.000	1.283	0.216	0.000
Commuter rate	-0.013	-0.427	0.000	-0.640	0.719	1.101	1.080	0.580	1.119	-1.148	4.031	-1.775	-2.066	0.308	-1.232
Male educational qualification rate	1.318	0.239	0.000	0.000	0.000	0.000	-0.103	-0.398	0.000	-0.885	1.924	2.438	0.000	0.072	-0.512
Female educational qualification rate	0.000	0.000	0.000	0.504	-1.992	-0.531	0.000	-0.152	-1.141	0.895	0.000	0.000	-0.141	0.000	1.071
Foreign rate	1.998	-1.632	0.000	0.177	0.683	0.000	1.127	0.381	0.000	-6.086	0.980	-2.243	-1.456	-0.456	-2.058
Grandparent rate	0.000	0.110	-0.952	1.430	0.000	0.433	2.050	-0.442	1.472	0.777	0.000	0.000	-0.813	-0.258	0.000
Babysitter rate	0.000	0.684	0.000	0.000	-2.773	0.000	-1.160	-0.509	0.537	0.000	0.000	1.343	-1.434	0.000	0.459
Number of members in the household	-0.070	0.000	5.227	1.528	0.527	-1.194	1.451	-0.731	-2.018	-0.954	-0.667	0.959	0.493	2.067	-0.884
Fertility rate	1.050	-1.631	0.244	-0.857	0.848	-0.251	0.405	0.000	-1.717	-2.049	0.000	0.000	1.488	0.141	-0.073
IVSM	0.000	0.000	4.139	0.000	-3.436	-3.311	-0.436	1.185	-1.932	0.000	-2.428	0.000	1.125	0.084	2.752

some territories more than in others, and the latter, in some cases, are entirely lacking. Moreover, Italy is facing ultra-low fertility rates. In this context, the Italian government has approved an extraordinary funding plan, utilizing Next Generation EU funds to bolster the provision of ECEC services. The aim is to address the imbalances in supply levels across different regions of the country.

In this challenging environment, this paper has explored the spatial distribution of ECEC services in Italy, jointly with territorial socio-demographic information that explains the demand side. Therefore, by applying the proposed spatial clustering approach, we showed and explored the strong variability of demand and supply across the Italian territory, confirming a lack of uniform ECEC governance. We then

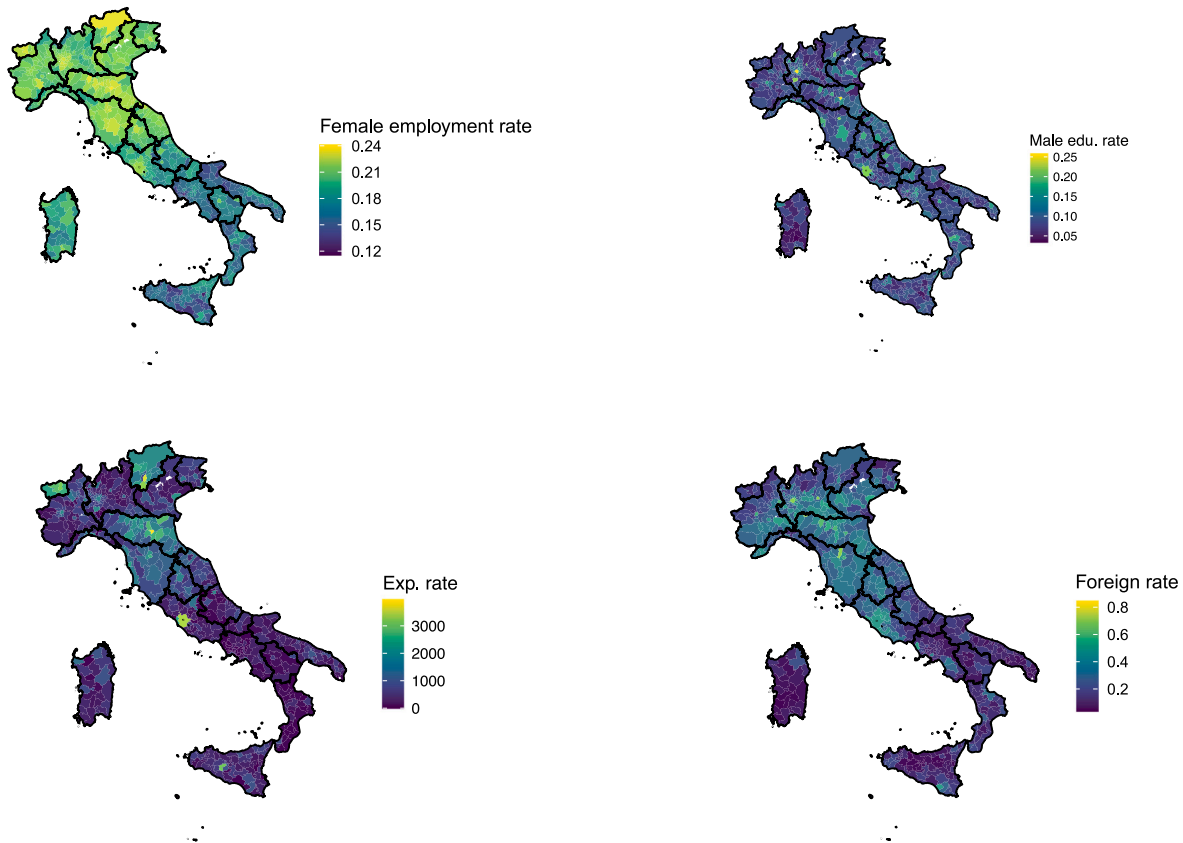


Fig. 10. Geographical map representation of the female employment rate (left top), male educational rate (right top), per capita expenditure rate (left bottom), and foreign rate (right bottom) at the ATS spatial level. The thick black segments represent the regional borders.

identified patterns (i.e., groups of ATSSs) sharing similar supply and demand characteristics. Therefore, employing appropriate statistical methods and integrating data from various sources enabled a better understanding of the relationship between the demand and supply of these services. This approach helps identify territorial commonalities and differences, providing valuable insights into the impact of governance policies or the lack thereof. While only some hypotheses have risen from our analyses on the first aspects, and further research is needed on the observed territorial differences, some interesting aspects emerged.

We noted how the situation in South Italy is critical in terms of demand and supply of ECEC services: the low services provision goes together with significant material and social deprivation with few job opportunities for females. In particular, only in some southern Italian areas (e.g., southern Appennini region and Salento) are present possible alternatives to ECEC services as a family (e.g., grandparents) or external support (e.g., babysitters). In contrast, northern and central Italy has more female employment opportunities even if the supply side is lacking in some areas (e.g., suburban northern areas). In particular, several areas in central Italy are inhabited by large families, thus suggesting the presence of family support, while commuting workers characterize several areas in the North. Finally, areas with particular characteristics were observed, such as university cities where the rate of study is high and commuting is relatively uncommon, the Naples area with considerable social and material deprivation, and suburban areas characterized by high commuting.

Policymakers must observe and account for these patterns to develop efficient welfare policies. For instance, our research has highlighted the effect of the presence of universities in designated areas across the South and North. In response, the government could devise welfare initiatives targeting young couples in these areas, such as

facilitating access to affordable housing and establishing collaborative ECEC services in partnership with universities. Furthermore, many commuting workers are concentrated in specific northern areas. Specific policies should be implemented to help parents organize a working life outside their municipality of residence. Policymakers could design policies that improve accessibility to ECEC services for commuters by, for example, proposing flexible hours for dropping off and picking up children. In addition, a comprehensive examination of commuter patterns between municipalities would provide valuable insights for designing such policies. Socio-economical aspects of the population characterize the demand, and the NRRP should, therefore, aim to understand the multidimensional structure of the supply and demand of ECEC services instead of focusing only on the supply side. Here, we proposed an instrument to look at that. Nevertheless, some limits stand.

Describing ECEC demand is challenging in Italy since the Italian permanent census is the only available data at the moment. This paper attempts to capture the presence of an alternative to childcare by constructing indices such as the grandparent and babysitter rates. However, the grandparent rate includes a portion of the population that is too old to be helpful to the family, and the babysitter rate considers only female students between the ages of 15 and 25. A future direction could be, therefore, improving the description of the presence of alternative childcare if additional data becomes accessible. Other measures could be inserted into the regionalization model to improve the description of the socio-economic structure of the population, such as the Italian citizenship income and fertility rate trends that were not available when the analysis of this manuscript was made. Finally, extending the analysis to include multiple years of observations could provide valuable insights into the temporal dynamics of these clusters, helping to identify virtuous regions that have successfully improved their ECEC policies over time. To properly isolate the effect of effective

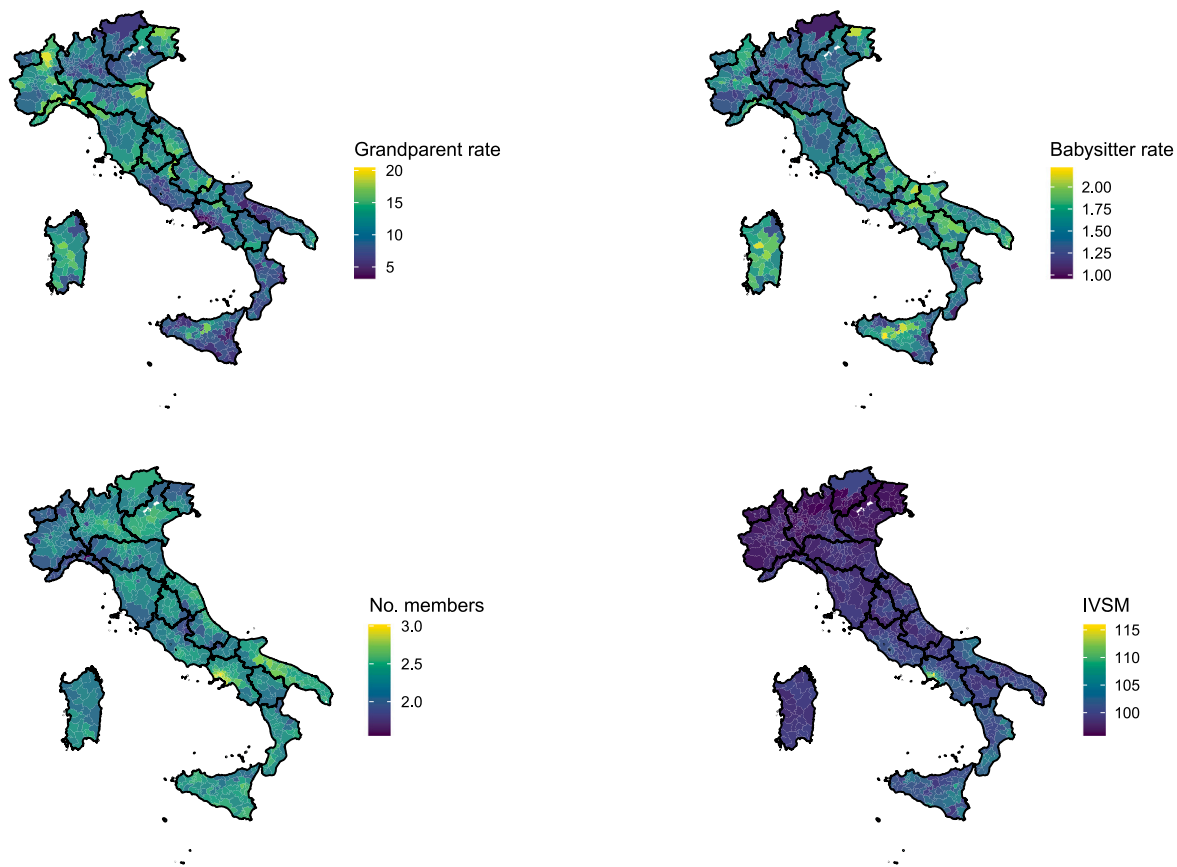


Fig. 11. Geographical map representation of the grandparents rate (left top), babysitter rate (right top), number of members in the household (left bottom), and IVSM (right bottom) at the ATS spatial level. The thick black segments represent the regional borders.

policies, causal inference approaches such as Difference-in-Differences (DiD) and the Average Treatment Effect on the Treated (ATT) can be utilized. However, given that we currently only have observational data and cannot conduct a controlled study, these methods are not feasible at this moment. For future developments, employing these methods could be valuable in evaluating the effectiveness of specific policies.

CRedit authorship contribution statement

Angela Andreella: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization,. **Emanuele Aliverti:** Writing – review & editing, Investigation, Conceptualization,. **Federico Caldura:** Writing – review & editing, Conceptualization,. **Stefano Campostrini:** Writing – review & editing, Supervision, Conceptualization,.

Data availability

Data will be made available on request.

Appendix A. Exploratory plots

Fig. 9 shows some exploratory plots (i.e., pairwise scatter plots, frequency histograms, and boxplots) for examining the variables defined in Section 2 and used in the clustering/modeling analysis. Each color represents one Italian macro area (light red for the Center, dark yellow for the Islands, green for the North-East, light blue for the North-West, and purple for the South).

Figs. 10 and 11 show the geographical distribution at the ATS spatial level of some variables described in Section 2.

Appendix B. Results

Table 4 shows the coefficients estimated from the penalized multinomial model discussed in Section 3.

References

- [1] Plantenga J, Remery C, Camilleri-Cassar F. The provision of childcare services: A comparative review of 30 European countries. Luxembourg: European Commission's Expert Group on Gender and Employment Issues (EGGE); 2009.
- [2] Landivar LC, Ruppanner L, Scarborough WJ. Are states created equal? Moving to a state with more expensive childcare reduces mothers' odds of employment. *Demography* 2021;58(2):451–70.
- [3] Addabbo T, Caiumi A, Maccagnan A. The allocation of time within Italian couples: Exploring its unequal gender distribution and the effect of childcare services. *Ann Econ Stat* 2012;105/106:209–27. <http://dx.doi.org/10.2307/23646462>.
- [4] Del Boca D. The effect of child care and part time opportunities on participation and fertility decisions in Italy. *J Popul Econ* 2002;15(3):549–73.
- [5] Chiuri M. Quality and demand of child care and female labour supply in Italy. *Labour* 2000;14(1):97–118.
- [6] Meyers MK, Gornick JC, Ross KE. Public childcare, parental leave, and employment. In: *Gender and welfare state regimes*. Oxford University Press; 1999. <http://dx.doi.org/10.1093/0198294166.003.0005>.
- [7] Luci-Greulich A, Thévenon O. The impact of family policies on fertility trends in developed countries. *Eur J Popul* 2013;29(4):387–416.
- [8] Haan P, Wrohlich K. Can child care policy encourage employment and fertility?: Evidence from a structural model. *Lab Econ* 2011;18(4):498–512.
- [9] Felfe C, Lalive R. Does early child care affect children's development? *J Public Econ* 2018;159:33–53.
- [10] van Huizen T, Plantenga J. Do children benefit from universal early childhood education and care? A meta-analysis of evidence from natural experiments. *Econ Educ Rev* 2018;66:206–22.
- [11] Felfe C, Nollenberger N, Rodríguez-Planas N. Can't buy mommy's love? Universal childcare and children's long-term cognitive development. *J Popul Econ* 2015;28(2):393–422.

- [12] Brilli Y, Del Boca D, Pronzato CD. Does child care availability play a role in maternal employment and children's development? Evidence from Italy. *Rev Econom Household* 2016;14(1):27–51.
- [13] Henderson A, White LA. Shrinking welfare states? Comparing maternity leave benefits and child care programs in European union and North American welfare states, 1985–2000. *J Eur Public Policy* 2004;11(3):497–519.
- [14] Van Lancker W, Ghysels J. Explaining patterns of inequality in childcare service use across 31 developed economies: A welfare state perspective. *Int J Comp Sociol* 2016;57(5):310–37.
- [15] Van Lancker W. Does the use of reconciliation policies enable single mothers to work? A comparative examination of European countries. In: *The triple bind of single-parent families*. Policy Press; 2018, p. 239–62.
- [16] Kalb G. Children, labour supply and child care: Challenges for empirical analysis. *Aust Econ Rev* 2009;42(3):276–99.
- [17] Kornstad T, Thoresen TO. A discrete choice model for labor supply and childcare. *J Popul Econ* 2007;20:781–803.
- [18] Carlin C, Davis EE, Krafft C, Tout K. Parental preferences and patterns of child care use among low-income families: A Bayesian analysis. *Child Youth Serv Rev* 2019;99:172–85.
- [19] Del Boca D, Locatelli M, Vuri D. Child-care choices by working mothers: The case of Italy. *Rev Econom Household* 2005;3:453–77.
- [20] Naumann IK. Childcare politics in the “new” welfare state. In: *The politics of the new welfare state*. Oxford University Press Oxford; 2012, p. 158–81.
- [21] Farfan-Portet M-I, Lorant V, Petrella F. Access to childcare services: The role of demand and supply-side policies. *Popul Res Policy Rev* 2011;30:165–83.
- [22] Van Lancker W. Reducing inequality in childcare service use across European countries: What (if any) is the role of social spending? *Soc Policy Admin* 2018;52(1):271–92.
- [23] Abrassart A, Bonoli G. Availability, cost or culture? Obstacles to childcare services for low-income families. *J Soc Policy* 2015;44(4):787–806.
- [24] Wrohlich K. Labor supply and child care choices in a rationed child care market. 2011.
- [25] Bonoli G, Reber F. The political economy of childcare in OECD countries: Explaining cross-national variation in spending and coverage rates. *Eur J Polit Res* 2010;49(1):97–118.
- [26] Dipartimento per le politiche della famiglia, ISTAT, Università Ca' Foscari, MIPA. Nidi e servizi educativi per la prima infanzia, stato dell'arte, criticità e sviluppi del sistema educativo integrato 0-6. ISTAT 2020. <https://www.istat.it/it/archivio/244116>.
- [27] Motiejunaite-Schulmeister A, Balcon M-P, de Coster I. Key data on early childhood education and care in Europe, 2019 edition. eurydice report.. Education, Audiovisual and Culture Executive Agency, European Commission 2019.
- [28] Aliverti E, Campostrini S, Caldura F, Zanotto L. Spatial modeling of childcare services in lombardia. In: *Book of short papers - SIS 2021*. 2021, p. 1369–74.
- [29] ISTAT. Indagine su nidi e servizi integrativi per la prima infanzia. 2019, <https://www.istat.it/it/archivio/267291>.
- [30] ISTAT. Censimento permanente della popolazione e delle abitazioni. 2019, <https://www.istat.it/it/censimenti-permanenti/popolazione-e-abitazioni>.
- [31] Assunção RM, Neves MC, Câmara G, da Costa Freitas C. Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *Int J Geogr Inf Sci* 2006;20(7):797–811.
- [32] Hastie T, Tibshirani R, Friedman JH. *The elements of statistical learning: data mining, inference, and prediction*, vol. 2, Springer; 2009.
- [33] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;33(1):1.
- [34] ISTAT. Le misure della vulnerabilità: un'applicazione a diversi ambiti territoriali. ISTAT 2020. <https://www.istat.it/it/files/2020/12/Le-misure-della-vulnerabilita.pdf>.
- [35] Cornelissen T, Dustmann C, Raute A, Schönberg U. Who benefits from universal child care? Estimating marginal returns to early child care attendance. *J Polit Econ* 2018;126(6):2356–409.
- [36] Baker M, Gruber J, Milligan K. Universal child care, maternal labor supply, and family well-being. *J Polit Econ* 2008;116(4):709–45.
- [37] Anderson PM, Levine PB. Child care and mothers' employment decisions. Working paper series 7058, National Bureau of Economic Research; 1999, <http://dx.doi.org/10.3386/w7058>, URL: <http://www.nber.org/papers/w7058>.
- [38] Farré L, Jofre-Monseny J, Torrecillas J. Commuting time and the gender gap in labor market participation. *J Econ Geogr* 2022;lbac037. <http://dx.doi.org/10.1093/jeg/lbac037>, arXiv:<https://academic.oup.com/joeg/advance-article-pdf/doi/10.1093/jeg/lbac037/48362456/lbac037.pdf>.
- [39] Borghorst M, Mulalic I, Van Ommeren J. Commuting, children and the gender wage gap. Technical Report, Tinbergen Institute Discussion Paper; 2021, <https://ssrn.com/abstract=4122091>.
- [40] Gramm V, Dalla Torre C, Membretti A. Farms in progress-providing childcare services as a means of empowering women farmers in South Tyrol, Italy. *Sustainability* 2020;12(2):467.
- [41] Bauernschuster S, Hener T, Rainer H. Does the expansion of public child care increase birth rates? Evidence from a low-fertility country. VfS annual conference 2013 (Duesseldorf): Competition policy and regulation in a global economic order 79909, Verein für Socialpolitik / German Economic Association; 2013, <https://ideas.repec.org/p/zbw/vfsc13/79909.html>.
- [42] MacQueen J. Classification and analysis of multivariate observations. In: *5th Berkeley symp. math. statist. probability*. 1967, p. 281–97.
- [43] Johnson SC. Hierarchical clustering schemes. *Psychometrika* 1967;32(3):241–54.
- [44] Xu D, Tian Y. A comprehensive survey of clustering algorithms. *Ann Data Sci* 2015;2(2):165–93.
- [45] Wise S, Haining R, Ma J. Providing spatial statistical data analysis functionality for the GIS user: the SAGE project. *Int J Geogr Inf Sci* 2001;15(3):239–54.
- [46] Bernetti I, Ciampi C, Sacchelli S. Minimizing carbon footprint of biomass energy supply chain in the Province of Florence. *Metody Ilościowe Badaniach Ekonom* 2010;11(1):24–36.
- [47] Openshaw S, Rao L. Algorithms for reengineering 1991 census geography. *Environ Plan A* 1995;27(3):425–46.
- [48] Haining R, Wise S, Blake M. Constructing regions for small area analysis: material deprivation and colorectal cancer. *J Public Health* 1994;16(4):429–38.
- [49] Fovell RG, Fovell M-YC. Climate zones of the conterminous United States defined using cluster analysis. *J Clim* 1993;6(11):2103–35.
- [50] Bondy JA, Murty USR. *Graph theory with applications*, vol. 290, Macmillan London; 1976.
- [51] Tutte WT. *Graph theory*, vol. 21, Cambridge University Press; 2001.



Angela Andreella got her Ph.D. in statistics from the University of Padova in 2021, supervised by Livio Finos, Jelle Goeman, and James Haxby. She is currently an assistant professor (RTD-A) in social statistics at the Department of Economics of the Ca' Foscari University of Venice, working on several projects related to Italian social welfare and health. Her main interest is applied statistics, i.e., connecting statistics with different fields, such as social sciences, health sciences, psychology, neuroscience, and medicine, improving both areas. Her statistical interests are in methods that permit understanding and managing high-dimensional data, i.e., dimension reduction techniques, statistical shape analysis, permutation tests, multiple testing, and selective inference.



Emanuele Aliverti holds a Ph.D. in statistics from the University of Padova (2020), supervised by Bruno Scarpa and David Dunson. Now he is an assistant professor of statistics at the Department of Statistics at the University of Padova. The focus of his research is on developing and applying novel statistical methods motivated by challenging problems in health and social sciences. He is interested in flexible Bayesian methodologies and scalable algorithms to perform inference on models for complex data, such as high-dimensional structures, networks, and tensors, balancing accuracy with interpretability.



Federico Caldura graduated in political science and governance of public organizations. Since 2020, he has collaborated with the Governance and Social Innovation (GSI) Center at the Department of Economics of Ca' Foscari University of Venice. Previously, he conducted research activities at the Regional Council of Veneto on immigration, disadvantaged municipalities, and personnel policies in the healthcare field. At the GSI Center, he has focused on labor market analysis, governance of public services, social innovation, and social programming. Additionally, he has participated in research projects on Corporate Social Responsibility.



Stefano Campostrini is a full professor of social statistics at the Ca' Foscari University of Venice, formerly a professor at Pavia, associate professor at Padova, and research fellow at Edinburgh (Scotland). His main research interest is applying statistics to support decision-making processes, particularly concerning health and social policies and interventions. This interest led to several research studies on evaluation, needs analysis, information systems, health systems, health inequalities, governance, and social innovation. International experience has regarded particularly the study and surveillance of behavioral risk factors for health and the evaluation of health promotion. He has collaborated with several international institutions and research centers (e.g., CDC in Atlanta, WHO in Geneva and Copenhagen, ILO in Torino, POHS in Adelaide, and Health Board of Singapore).