Università
Ca'Foscari
Venezia

Master's Degree

in Economics and Finance

Final Thesis

Measuring the early child care supply in Italy, a spatial approach

Supervisor

Ch. Prof. Danilo Cavapozzi

Graduand

Enrico Fornasiero

871877

Academic year

2018/2019

Contents

1 Introduction

This essay measures the formal early child care provision in Italy investigating the spatial issue, the aim is to introduce a spatial variable representing the accessibility to the service exploiting the dispersion of the Italian population and of the early child care facilities. We then analyse how this variable correlates with the familiar decision of enrolling the child to an early child care facility; finally we carry out an empirical exercise to asses how using the early child care system affects the maternal employment, taking into account the endogeneity of the child care utilization in this model, implementing an instrumental variable approach based on our accessibility measure.

The importance of having an efficient early child care system in a developed country represents a shared opinion between economists, politicians and researchers. The main arguments treated in the literature on this topic are related to the human capital accumulation of the children, the cognitive development deriving from the presence of qualified personnel in the dedicated facilities, and to the role that the early child care system plays in the maternal labour force participation, a support for the mothers of newborns that are willing to continue working and demand a service able to take care of the infant during the working time. The existing literature on the topic (both American and European) focuses on the impact that variations in the price and in the quantity supplied, intended as opening hours and spots available, have on familiar decisions and on the maternal labour force participation. The impact of policies aiming to improve the formal early child care system varies according to the country analyzed, the period of time considered and the econometric specifications adopted; but almost all the researchers agree in finding a positive effect of improvements in the early child care service on both the maternal labour force participation and the cognitive development of the children. The human capital accumulation of children may in

fact be influenced by the presence of professional personnel, in Italy for example, exploiting the results of a national test for students of the primary school, it has been possible to analyze the impact of the early child care system confronting the results obtained by students from regions with different levels of child coverage; on average the score for the Language test was higher for students living in the regions with an high level of child coverage (Brilli et al., 2016). For the maternal labour force participation, the researchers agree in finding a positive effect of the early child care system on the mothers' working status; negative variations in the price decreases the cost of the child care, encouraging medium-low income families to try to enroll the child to the system (Anderson and Levine, 1999; Blau and Currie, 2006; Baker et al., 2008). A similar effect emerges when considering an increase in the number of spots available, in particular when initially the service was scarce; reducing the number of children for which the service was precluded, lead to an increase in the probability for the mothers to work (Del Boca et al., 2002; Del Boca and Vuri., 2007; Del Boca et al., 2008; Brilli et al., 2016; Chiuri, 2000). Differently from part of the literature that considers as a whole the effects of the service dedicated to the children younger than the age required for the mandatory school, we focus only to the service dedicated to children between 0 and 2 years old of age to check the effect that the service has on the families during the first years of life of the child.

According with European statistics, Italy performs badly both in the child coverage ratio, the number of spots supplied over the potential demand, and in the female labour force participation. An improvement in the early child care system may then be seen as a first step in the elimination of the barriers that discourage the female participation in the labour market and as a sign of attention towards the families and the new generations of citizens.

In a country in which the regional differences represent a relevant is-

sue, also the early child care system suffers of a strong heterogeneity in the provision of the service across the nation. While there exist regions with an efficient system, others register a low number of spots and a limited number of facilities. Moreover, the service is not present in all the municipalities, families have then to take the child to a structure located far from the house, investing time and money to reach the facility. The child coverage ratio, being an aggregate measure, does not consider that a family can be forced not to use the early child care system not only because the total number of available spots is scarce, but also because the closest facility is located too far from home; families do not know the level of child coverage when considering to enroll the child to the early child care system, but firstly observe which facilities are spatially accessible in terms of monetary cost and opportunity cost. Only few studies have focused on this spatial issue; the territorial diffusion of the facilities may encourage the use of the service by the families that are located near the structures (home, workplace and relatives) or discourage it when the service is far and scarce. The spatial analysis would then be useful for the legislator to identify the areas of the country in which there is a lack in the provision of the service and to plan the correct location of new facilities.

Our essay tries to explore the Italian spatial diffusion of the early child care system, analyzing if the service is efficiently located in the country according to the distribution of the Italian population. In Section 2 we resume the results obtained in the literature on the topic, describing also the Italian data on female labour force participation and provision of the early child care system, comparing Italy with other European countries. In the last part of the section we introduce the spatial approach and we describe how it has been analyzed in the literature. The spatial model is examined in Section 3; initially we describe the reasons that led us to analyze the spatial issue, while the main part of Section 3 is dedicated to the presentation of

the model and to the identification of the spatial accessibility variable, starting from ISTAT data. The spatial model has been built using a Geographic Information System (GIS), we describe the tools and the procedure used in the definition of the model and the spatial accessibility variable; for the list of the early child care facilities we exploited the ISTAT database, while the data on the distribution of the Italian population (population census 2011) and the data for the construction of the GIS model have been acquired by the ISTAT "Territory and Cartography" database (ISTAT, 2019b). Section 4 is entirely dedicated to the econometric specification of the research, which aim is to describe the effects of the formal early child care system on the probability for a family to enroll the child to the service and on mothers' working condition; in particular we focus on the impact of the spatial accessibility variable. The sample used is an ISTAT survey called "Multiscopo - Aspetti della Vita Quotidiana" with information on the familiar composition and on the use of the early child care system. The spatial variable would then estimate the effect of the spatial distance between the location of the family and the location of the facilities, on the probability for a family to enroll the child to the early child care system. In the last part of the section, measuring the impact of the system on the mothers' working status, we exploit the spatial accessibility variable as instrument to estimate the impact that having a child enrolled to a facility has on the probability for a mother to work.

The introduction of the spatial variable contributes to the existing literature characterizing the supply of the service also for its geographic territorial presence, we demonstrate that the spatial issue plays a role in the likelihood to enroll a child to the early child care system and on the maternal employment decision.

2 Early child care system in the literature

2.1 Effects of early child care system on children cognitive functioning and women employment

The importance of having a widespread and efficient early child care system in a developed country is a well-known issue in the economic literature. The main variables related to the care of children in age 0-2 years old are the level of the child coverage (intended as the ratio between the number of slots available over the population of interest) and the price of the service; both of them have an impact on the employment decision of the mother, on the human capital accumulation of the child and on other non-economic variables concerning psychological aspects that we are not going to treat in this research. The early child care system can be divided in formal and informal; in the former the service is provided by external structures (private or public) with professional personnel, in the latter the care of the infant is entrusted to family members or friends. The early child care puzzle has to be analyzed as a two-generations framework from which both the parents and the children can take advantages; the former looking for the child care facilities to maintain their position in the labour market, the latter benefiting from a high quality system in the development of their human capital.

Referring to the cognitive development of the child and its human capital accumulation, educate the child inside the familiar environment (by parents or other members of the family like grandparents) or outside it, through formal service, may lead to different results in term of skills acquired in the first years of life of the child. Firstly, in a formal structure the child has the opportunity to interact and build relations with other children acquiring skills that are more difficult to have when the child care is informal and there is a lack of contact with other children of the same age; secondly the presence of qualified

childminders in the formal system can help the child to develop earlier skills that can result to be postponed in the case of informal child care. The economic literature analyzes in particular this last aspect; what is the impact on the human capital accumulation derived by the attendance of a formal child care system. Despite the difficulties on measuring the final effect, due to the age of the children, Brilli et al. (2016) analyzing the INVALSI results for primary school in Italy (an Italian national test on mathematical and language abilities), found out that there exists a correlation between the public child coverage ratio, the mother participation on labour market and the results of the INVALSI test. The impact is positive for the score of the Language test while for the Mathematical one the effect is null; in particular they found that this result is stronger for the areas of the country in which the provision of the service, by the public sector, is scarce. The relation between educational attainment and early child care attendance is acquiring importance in the national debate, not only for the occupational consequences but also in light of the PISA results of the last years, in which Italy performed below the average level of the European countries (OECD, 2018). Exploring the correlation between these two variables can in fact suggest a way to improve the human capital accumulation of the students of the primary school and improve, even if marginally, the overall results of the PISA statistics. Similarly, Del Boca et. al (2018) found out that for the United Kingdom the informal child care, provided by grandparents, has a positive impact on the vocabulary abilities of the infants but also that those children perform worse on cognitive test than children enrolled in formal child care.

The analysis of the effects of the formal child care on the cognitive abilities of the infants in age 0-2 years old have been widely analysed in the economic literature, and almost all the researchers agree in finding a long-term benefit on the human capital accumulation of the child

(Barnett, 2011; Morrissey, 2017). Even if it is impossible to compare nations with very different educational schemes and different cultures, it appears clearly important to deeper investigate the effects that the formal child care system have on the human capital accumulation of the children.

As stated before the provision of child care facilities impacts mainly on two variables: the human capital accumulation of the child and the labour supply decision of the mother. The final impact on these variables is driven in particular by three factors: variations in the price of the service, variations in the quantity supplied, in terms of slots available and opening hours of the service, and the quality level of the service, both formal and informal.

The literature on female labour market and women's decisions, clearly explains how the birth of a child influences the behavior of the mother (Becker, 1965). The presence of a child increases the opportunity cost of time outside the labour market and contemporary decreases the effective wage through the cost of the child care; it follows that greater are the responsibilities linked to the child-rearing, less are the probabilities that the mother will participate in the labour market. The provision of child care facilities has to be considered as a social strategy aiming to reduce the opportunity costs of children and to provide opportunities to combine child-rearing responsibilities and work. An efficient and available early child care system combined with a sufficient number of part-time opportunities, will then increase the likelihood to have more mothers in the labour market (Del Boca et al., 2008; Bratti et al., 2005). An increase in the number of places available impacts positively on the probability for the child to access to the service and reduces the number of hours that the mother has to dedicate to the infant, while a decrease in the price of the service increases the net wage through a reduction of the cost of child care. Both these policy measures pursue a unique objective, the reduction

of the opportunity cost of children moderating the negative impact they have on mothers' labour supply decisions. The literature suggests that a reduction in the price of the service, through subsides or other policy measures, and/or an increase in the number of places in the public and private child care system, lead to an increase in the attendance of the service and have a positive impact on mothers' labour force participation. The results obtained by the researchers vary on the basis of the country analyzed, of the database used and of the econometric specification implemented, but almost all of them demonstrates the final positive impact of policies aiming to improve the early child care system. For the general analysis of the literature, we are not going to consider only the specific age group between 0 and 2 years old but a more general bandwidth comprehending children between 0-6 years old of age; in fact in the referring literature only few economists have focused on the smallest bandwidth while the majority treated the topic as a whole, considering together all the children that are too young for the mandatory school.
The literature related to this topic can be divided in two main categories: the American point of view, which highlights the importance of the price of the service and analyzes variations in the policy measures to help low income families to access to the service, and the European research that focuses on the availability of the service in terms of child coverage ratio and labour market policies to help mothers, and fathers, to reconcile work and child-rearing. For the purpose of our essay we will treat these two categories separated to stress the attention on the different features they propose.

In the United States the child care service is mainly composed by the private sector and, differently from European countries, it is not strongly regulated or subsidised by the municipalities and the central government (Ruhm, 2011). The lack of public structures has led to the creation of an expensive child care market with few affordable places

for the poorest families. This framework has driven the American authors to investigate primarily what would be the effect of a reduction in the total expense of a family on the mother employment decision, analyzing in particular the effects of the early childhood policies focused on low income families.

A decrease in the cost of the service, meaning a decrease in the total expense that a family has to spend for the child, increases the value of employment in particular for the family with a level of income around the poverty line or with a low level of educational skills (Anderson and Levine, 1999). The high price imposed by the market acts as an entrance barrier for those families that simply can not afford the service, even if these families are those who would benefit probably more by it. The birth of a child in a low income family usually lead to a strong use of the informal child care system, relying on grandparents in the majority of the cases; or to the decision, by the mothers, to exit the labour market to dedicate their time to the child. The lack of low cost formal options is then a crucial determinant of the employment decision for this part of the population. In Hofferth (1999) is also illustrated as several Government policies instituted to help the economically distressed families, through the build of new public structures and subsides to the family, as the "Head Start" program, have not sufficiently increased the number of affordable structures for children from low income families. These policies have led to a strong use of the high-quality centers by the rich families (that did not benefit of the policy but could afford the service) and by the poorest one (that benefited from the program), leaving the children of families with an income level just a little above the poverty line and the middle-class, with no access to the service or with only the economical possibility to enroll the child to a lower quality service (Hofferth, 1999).
The differences in the quality of the child care systems are then expected to generate a gap between the skills acquired by children from

families with different levels of education; reducing this gap may level the disparity that exists in the American educational system where the human capital accumulation depends mostly on the family background. As demonstrated in Anderson and Levine (1999), the children of low skilled mothers have almost twice the likelihood to use the informal child care system than those born from mothers with higher skills. The use of the informal service is strongly related to the social condition of the family, trusting the care of the infant to the grandparents is commonly viewed as a sign of an availability lack of the formal child care service, both considering the coverage ratio level and a daily price too high for the family.

Moving to the price effect on maternal labour force participation, the majority of the studies found a negative association between these two variables, confirming the underlined microenomic theory. These studies, even if are in concordance about the sign of the relation, show heterogeneous results depending on the time span considered, on the country analyzed and on the database used. Blau and Currie (2006), gathered the different researches on this topic and listed the main results obtained by the authors. They collected the final results and found that the estimated price elasticities ranged between 0.06 and -3.60 with variation according to country, skills, income, family background and ethnicity. This approach allows only to identify the employment variation linked to changes in the prices of the paid service, the formal one, and does not consider the existence of an unpaid option, generally the informal system. Taking a closer look, following Blau and Currie, we would also need to consider the interaction between the paid and unpaid options; in fact variations in the price of the formal system can drive the mothers from the use of the paid (formal) service towards the unpaid one (informal), while not changing their employment situation.

Summarising, an increase in the cost of the formal child care sys-

tem affects the familiar choice to use the formal solution reducing the likelihood to enroll the child to a facility. Conversely, if available, the family can react to an increase in the price of the formal sector entrusting the cure of the infant to a family member or a friend, increasing the likelihood to use the informal sector. We can state then that, a positive variation of the price affects the labour force participation of the mothers only when an informal option for the cure of the child is not available; when a mother can not afford the formal sector is forced to exit the labour market only when she has not a relative that can take care of the child. Furthermore it may happen that the formal and the informal sectors are both used simultaneously by the family; this is the case when the opening hours of the facilities do not cover the entire length of the working time of the mother and a relative is entrusted to stay with the child after the closing time of the facility, waiting for the arrival of the mother. In this situation an increase in the price of the early child care facility may lead to the decision to totally entrust the care of the child to the relative to avoid the effect that the variation of the price would have on the family. It is straightforward to consider the opposite case, a reduction in the price of the service leads to an increase in the labour force participation by the mothers only when the informal solution is not available; mothers can in fact decide to switch from an informal care to a formal one if the price is low, not causing variation in the total supply of labour. The consequence of a decrease in the price of the service has an additional effect on the increase in the potential demand; increasing the demand of the service when the number of spots is fixed, may damage the mothers that are willing to work but are forced to stay outside the labour market because they do not have the possibility to use an informal solution. There may exist in fact eligibility criteria used to ration the excess of demand that reward the children of working mothers, mothers that may have decided to exploit the low price of

a facility to switch from an informal to a formal system, leaving the unemployed mothers outside the formal service and not increasing, in this case, the overall level of labour force participation.
A confirmation to this framework is given by Doiron and Kalb (2005); their work on the Australian context demonstrates how the informal care may be a substitute for the formal one and that the price elasticities on labour supply is negative and depends on the age of the child and on the characteristics of the mothers (Doiron and Kalb, 2005). They also measured the price elasticities with respect to married men but they found negligible effects, this result shows us how, still today, the care of the child, which is a time intense occupation, is carried out mainly by the mothers. It also demonstrates the so called "breadwinner" pattern for which, in absence of an efficient child care system, it is the father to be the member of the family designated to earn money for the whole family.

One of the most important quasi-experimental analysis computed in the last decades is the Baker et al.(2008) one. In this paper the authors investigated the effects of a large expansion of the formal child care system happened in Quebec, Canada, measuring the impact it had on the child care utilization, on the maternal labour force participation and on the familiar well-being. To carry out the analysis they exploited a change in the provincial regulation that extended the full-time kindergarten to all the 5 years old children and provided child care with an out of pocket cost of 5 dollars per day to all the 4 years old children. Universal child care was then extended in the subsequent years with other benefits for lower age children. For their analysis they compared, using a difference-in-difference model, the variations in the considered variables with the other provinces of Canada to check the real impact of the policy. The final result is an analysis of a universal subsidized child care program with findings strongly in line with the theoretical framework. The effect on the mothers' labour force partici-

pation was positive and significant but only half as large as the impact on the child care utilization. The probable explanations are due to the possibility that many women were using the service without working or that they switched from an informal care to a formal subsidized one. Notable it is the result they obtained when they analyzed the impact of the policy on the child and family well-being, the expansion of the service has led the children to spend more time outside the familiar environment leading to an increase in the level of stress of the parents and in the aggressiveness and anxiety levels of the children. The authors explained that this result, that concerns the psychological aspects related to the use of the formal early child care service, may be a temporary short-run effect; the family may need time to adjust their expectations and behavior in order to fully benefit from the universal child care system. This interpretation seems to be the logic, otherwise we would find a pattern in which families would exploit the introduction of the policy even if it would cause a worsening in the outcome for both child and parents. Another interpretation explains that this effect may be offset by other unmeasured benefits at the familiar level related to the introduction of the policy, as an increase in the family income (Baker et al., 2008).

Moving from the American continent to Europe, we focus our attention not only to the implications derived from price variations, but mainly on the impact that rises in the child coverage ratio may have on the families. The great variety of welfare models and cultures that characterizes the European continent, makes interesting the comparison of the results obtained in the different countries. In fact we would expect strong heterogeneity in particular between Nordic countries like the Scandinavian ones (with strong subsidized welfare system) and the Mediterranean ones (where services are usually insufficient and the unemployment ratio is higher with respect to the European average).

Norway was one of the first nation worldwide to adopt a universal public child care system, and in 1975 began an expansion of subsidized child care slots for children 3-6 years old of age. Havnes and Mogstad (2011) examined the effects of this expansion using a difference-in-difference methodology exploiting the space and time variation in the implementation of this policy to understand the impact it had on the maternal labour supply. They found that to a 1 percentage point increase in child coverage it followed only a marginal 0.06 percentage points increase in maternal employment; the social cost of the policy resulted to be not efficient for such an increase in the maternal labour supply. Moreover they illustrated that the development of a universal child care system may require a large amount of resources just to have a little impact on the employment level. The two authors explained this result stating that in the Norway specific situation, this policy had only the effect to crowd out other forms of child care having no relevant effect on the maternal labour supply (Havnes and Mogstad, 2011).

In Italy the literature related to the child care system, its availability and the consequences it has on the maternal labour force participation, has acquired more and more importance in the last decades. The lack of the service in many Italian regions and the well-known weak points of the internal labour market structure, have contributed to develop a country specific literature to examine how the combination of these variables impacts on the family choices. This topic has acquired relevance also after the intervention of the European Union; the positive effect that a widespread early child care system can have on the society, has led the European Council to set a threshold level for the early child coverage ratio of 33% to be reached in 2010 by all the countries of the European Union (European Council Presidency Conclusion, 2002). Italy ranks as one of the European country with the lowest level of coverage ratio related to the system dedicated to

children between 0 and 2 years old of age.

Table 1: Main European countries and relative child coverage ratio, 2018.

Country	Coverage ratio
Denmark	63.2%
Netherlands	56.8%
Norway	55.6%
Belgium	54%
Spain	50.5%
Portugal	50.2%
France	50%
Sweden	49.3%
Greece	40.9%
Great Britain	38.7%
Ireland	37.7%
Finland	37.1%
Germany	29.8%
Italy	**25.7%**
Austria	20%
Poland	10.8%

EUROSTAT.

In Table 1 we can observe how Italy compares with the other European countries in 2018, not only with the Nordic nations, but in particular with the other Southern European countries as Spain, Portugal and Greece. Italy ranks low, situated between Germany and Austria and far from Spain and Portugal that had been able to reach the minimum level required by the European Council. For the table we maintained the definition of formal child care given by EUROSTAT that excludes certain typology of centers, as those in which it may coexist the presence of a sort of connection between the child and the family, that are indeed taken into consideration by ISTAT and in our research.

The implementation of an efficient formal child care system has been indicated as one of the first steps to remove disincentives to female labour force participation and, in this particular context, also as one of the many variables that can impact the fertility decision of the women. In all the industrialized countries, a fall in the fertility rate has been

accompanied by a contextual increase in the female labour force participation; the microeconomic models explain how an increase in the wage rates of women, leads to an increase in the female employment rate and a decrease in the fertility rate. What has been commonly observed in the other industrialized countries, has not happened in Italy where a dramatic fall of the fertility rate has caused only a modest increase in female labour force participation. Italy, together with Spain, Portugal and other Mediterranean countries, shared a severe situation regarding the internal population dynamic. Those countries were considered part of the so called "Southern model" (Del Boca, 2002), characterized by a low level of social protection for families and children and by strict employment regulations. They have also experienced a "lowest low" fertility situation in which the fertility rate was below the natural replacement rate and under the value of 1.3 children per woman (Billari and Kohler, 2004). In Italy, where the low fertility rate meets a stagnant level of productivity and a low female employment rate, the combination of these variables will probably create an explosive situation for the Italian future generations with a fall of the working population and an increase of the dependency ratio (measured as the ratio between the share of the population in the labour force and the retired one).

Table 2 illustrates as in 2018 Italy ranked low on the female employment rate compared with other European countries, not only below Germany and France, but also below other Mediterranean countries like Portugal and Spain, with a gap with respect to the Lusitanian country of 17 percentage points (OECD, 2018). Del Boca (2002) advocated the low employment level, that in 2002 put Italy also below the Greece by OECD calculations, to the rigidities and the imperfections of the internal labour market and the characteristics of the public child care system. One of the peculiarities of the Italian labour market, for the authors, seemed to be the lack of part-time opportunities

Table 2: Female Employment Rate by country, 2018.

Country	Female Employment Ratio 15-64 yo
Sweden	75.9%
Netherlands	72.8%
Norway	72.7%
Germany	72%
Denmark	71.3%
Finland	70.6%
United Kingdom	70.3%
Austria	68.6%
Portugal	66.8%
Ireland	63.3%
France	61.9%
Poland	60.8%
Belgium	60.8%
Spain	56.9%
Italy	**49.5%**
Greece	45.3%

Data warehouse OECD.

for women, part time employment that in 2002 accounted only for the 22.7% of the total by OECD calculations. This kind of contract could allow mothers that have not been able to enroll the child to a formal child care system (due to the lack of available places), to keep working and entrust, if possible, other family members to the care of the child during the working time, working time that is in this case shorter than the full-time situation. The results they presented demonstrated how the decisions to work and to have a child are both positive related to the formal child care supply and the availability of part-time jobs. A fundamental role is then played by the number of available places in the formal child care system, an increase in the total coverage ratio would have a double impact simultaneously increasing job opportunities for women and reducing the cost of taking a full-time job (Del Boca, 2002).

Before the start of the decade it was difficult to find the data necessary for a complete analysis of the child care dynamics because there did not exists a unique survey that comprehended all the variables

of interest. Del Boca et al. (2005) overcame this problem matching different surveys from different sources, Multiscopo ISTAT for the use of formal and informal child care and the Bank of Italy Survey for the income and earnings characteristics of the families. Analyzing the mechanism that determines the demand of the service, they found that for the public sector the demand is driven by the availability of places. Furthermore, they noticed how the public and private sectors are not perfect substitute; in the econometric model adopted they found that when the cost of the public sector increases, mothers prefer to move to an informal solution rather than switching from the public to the private sector.

Given the scarcity of available places, municipalities ration the excess of demand for the public early child care system, introducing eligibility criteria to favour the families that respect determined characteristics. As explained in Del Boca and Vuri (2007), the eligibility criteria are strictly related to the social objective that the central and regional government in primis, and the municipality, want to pursue to promote their idea of society. While is considered ethic to give a facilitated access to children with disabilities or from disrupted families, other criteria strongly depends on the final effect that the legislator wants to achieve. Generally the eligibility criteria influence both the access (position in the waiting list) and the final price once the child is enrolled in the structure; the variables from which are determined depends on family size, family income, family composition, family residency and the employment situation of the parents. Giving a higher score to the children of working mothers influences for example the probability that these mothers will maintain the job during the years in which the child attends the service; favouring the child from a family that lives in the same municipality in which the public structure is located, can create a specific social welfare area for the inhabitants and so on. It is also important to consider that the availability is also

rationed in the number of hours of care supplied and that this specific parameter strongly influences the maternal employment choice and the type of contract of work that the mothers look for. In a context characterised by a lack of territorial homogeneity in the diffusion of the service and in the flexibility of the opening hours, the coverage ratio appears to be one of the most relevant variable to measure the early child care system; this variable does affect both the child care decision of the family and the employment one. Del Boca and Vuri (2007) stated that this variable should at least reach the 40% to lead an increase in the female labour market employment up to 60%.
The crucial role of the quantity and quality of the formal child care supply, emerges in almost all the Italian literature. The level of early child care coverage has been demonstrated to have the strongest and least ambiguous effect on the employment decision of the mother compared to other variables (Del Boca et al., 2008).

The Government decisions during the last decades regarding the formal early child care system, did not promote a structural well-defined plan, preferring to give a direct monetary help to the families with a child or to help the families whose children are already enrolled to public or private structures, not considering the huge number of families that, for many reasons, do not have the possibility to access to the service. Moreover, the institutional subsides appear to be insufficient to support the expenses related to the child-rearing of the child. The long-term instability of the Italian politics, with a very short lifespan of the Governments, creates a fragmented regulatory system with frequent changes in the legislation, exacerbating an already complicated situation. The lack of a long-term objective is reflected in the absence of an effective plan to increase the number of structure and to level the differences across the regions. While it will anyway result relevant a strong reduction of the fees for enrolled children (as it seems to be implemented in 2020), the attention should also be moved to the

level of coverage ratio and to the correct location of the structures.

2.2 Early child care provision in Italy

In 2013, the reference year of our research, the number of structures in Italy were 13,459: the 35% of them were public while the 65% private, for a total coverage equal to the 22.5% of the population, far from the European target (ISTAT, 2016). Not only Italy performs badly on average, but particularly considering different areas of the country. In fact, as described in Table 3, if in the Northern regions the child coverage ratio states around the 28%, in the Southern ones the level is mostly under the 15%, being then almost ineffective for the development of a widespread support network for young families. This situation remained unaltered in 2017, the last year for which we can have the data, Table 3 shows that the Southern regions increased the gap with the Northern ones. In particular in the Islands we observe a decrease in the child coverage ratio indicating the absence of attention, given by these regions, to the early child care system (ISTAT, 2019a).

Table 3: Child coverage ratio per macro-area, 2013 and 2017.

Area	Coverage ratio 2013	Coverage ratio 2017
North West	27.4%	29.2%
North East	28.4%	32.5%
Center	29.2%	32.4%
South	10.1%	12.3%
Islands	14.5%	13.5%

Data warehouse ISTAT.

Only two regions in 2013 had a coverage ratio higher than the European threshold, Umbria and Emilia-Romagna. In Table 4 we observe as in 2013 the top three regions by level of child coverage ratio were located in the North and Center of the peninsula, while the worst three in the so called "Mezzogiorno" comprehending South and

Islands. Impressive it results to be the gap between the top scorer and the worst performer of 31 points. We can then identify as one of the

Table 4: Decreasing level of coverage ratio per region, 2013.

Region	Coverage ratio
Umbria	37%
Emilia-Romagna	35.2%
Toscana	32%
...	...
Sicilia	11.1%
Calabria	10.6%
Campania	6.2%

Data warehouse ISTAT

main issue of the Italian context to be the level of child coverage ratio that is still too low compared with the other industrialized countries of the European Union. Focusing on the internal regional heterogeneity, the differences reflect a long term trend of low availability of services in the Southern regions, while the Northern regions reach standards on average with other European countries.

The formal public early child care system in Italy is decentralized: the main decision makers are the municipalities with regions empowered only with a management control. The government responsibilities rely on the definition of common objectives and on the allocations of the resources to sustain the municipalities and, directly, the family that decides to use the service; either public or private. Even if there is a difference in the number of structures between the public and private sector, the number of infants that attended the formal child care system in 2013 resulted to be equally divided between them, the number of spots available in the public sector accounted for the 50.5% of the total supply. In fact the number of places available per structure are, on average, higher for the public sector than for the private. As mentioned in Brilli et al. (2016), there exists a positive relation between the availability of public and private child care facilities across Italy;

areas in which the public early child care system is more developed are also those with an higher number of private centers. This correlation may reflect the propensity of the public authorities to outsource the service to private structures to manage the high level of demand from the population, creating agreement between the municipalities (the owner of the public structures) and the private counterpart to increase the number of total slots, maintaining the control under the public administration.

The decision regarding the quantity supplied by the municipalities depends on the budget constraints that they have to subdue (monetary resources, presence of qualified childminders, rooms available to create the center and so on) and on their preferences. Each municipality is then free to determine eligibility criteria to manage the excess of demand giving advantages to certain families instead of others. While children with disability and orphans are usually the most favoured, each municipality can decide to give more importance to different social aspects. The municipalities can decide to use eligibility criteria to encourage the female labour force participation (priority given to family where the mother or both parents are employed), to help the economically distressed family (low income), to favour the families that are resident in the city where the structure is located or, using a combination of criteria, to specifically select families that for the policy maker are those that would benefit more from the service. Municipalities are also free to cooperate one another to create a wider network to meet more efficiently the different level of demand in the area (usually between small villages). Differently from the public sector, the private one can decide eligibility criteria that are unbounded from the social condition of the family or from its residence; for example the access to the service can be restricted only to children with at least one parent working in a certain firm (corporate welfare). The public and private systems are then complementary in the supply of

the service, mostly working together to respond to the different needs of the demand.

In 2002 the European Council, after the meeting of Barcelona, pointed out the importance of having an efficient formal early child care system and preschool child care system, as one of the key determinants for the development of employment strategies in order to remove disincentives to female labour force participation. The document encourages the Member States to remove any kind of barriers to female participation improving the provision of child care facilities, and stated as thresholds of coverage ratio a level of 90% of attendance for children between 3 years old of age and the mandatory school age, and of 33% for children under 3 years old of age. The gap between these two thresholds underlines the role of the early child care system, in particular in the countries with a not efficient labour market and a high unemployment level. Not having the possibility to access to the early child care system forces the mothers, when it is not possible to use an informal child care provision, to exit the labour market for taking care of the child. In particular in Italy, considering the high level of preschool attendance (more than 90% for children 3-5 years old of age) and the high female unemployment level; it is straightforward to understand how risky is for the Italian mothers to leave their job positions to take care of their newborns and maybe not being able to re-enter to the labour market when the child gets 3 years old of age and can be enrolled in the preschool system. For the development of a fairer society is then necessary to implement policies to increase the availability of early child care facilities to give women the possibility to follow their preferences, be they to keep working or to leave the labour market and take care of the child.

In 2013 the female employment rate ranged below the 50% indicating a scarce participation in the labour market by women; the rate decreases when we consider only women between 15 and 34 years old

of age with a level of 34.6%, with the male counterpart of 45.1% (ISTAT, 2019a). The difficulties that women face when they try to enter to the labour market, influence also a set of related variables as the fertility choice. The Italian fertility rate has reached, in the last decades, very low level ranging well below the replacement rate of 2.1 children per women. The uncertainty related to the labour market and the lack of solutions for making the coexistence of work and child-rearing possible, as the absence of a widespread early child care system, are probably some of the many causes of the current Italian situation.

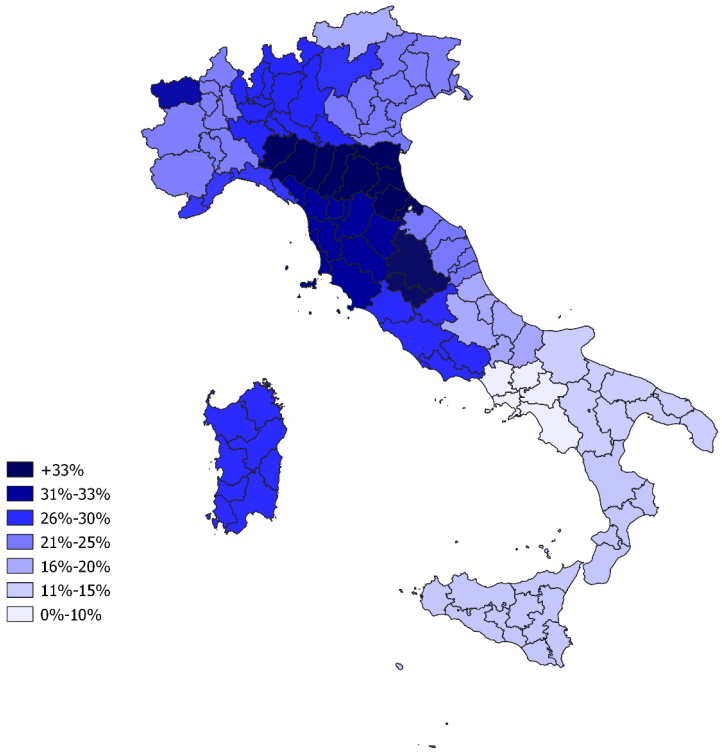
2.3 Spatial provision of early child care and policy implication

While the implications related to the affordability of the early child care service and the child coverage ratio have been largely discussed in the literature, resulting in a well-defined framework, the spatial problem related to this topic has been investigated only by a bunch of researchers and, notably, a spatial research analysis has not been conducted in Italy (as far as our knowledge). For spatial problem and spatial analysis we intend a methodology built to investigate the role that plays the distance between the location of the families demanding the service and the location of the supply (the child care structures), on the child care and employment choice of the family.

As stated in the previous paragraphs, one of the main concern in the Italian system, is the heterogeneity in the distribution of the service across the country. In Figure 1 we map the value of the child coverage ratio per region, in 2013, not considering the differences that exist between the provinces belonging to the same region just for clarity. The policy maker should first of all try to force the regions with the lowest values to catch up the more virtuous one through an increment in the number of the facilities and spots available, and to plan a long-term project for the increase of the national average at least as high as

the Barcelona target of 33% of coverage (in 2013 the level was 22.5%).

Figure 1: Child coverage ratio per region, 2013.



Own elaboration from ISTAT.

Even if the child coverage ratio has been demonstrated to be a significant variable in the analysis, we must observe how imprecise it is when used in relation with samples from surveys on familiar level. In fact it may happen that a family would not participate to the formal child care system not only because of availability rationing, child coverage ratio, but also because of spatial rationing, namely the closer structure to the house is located too far. It may then happen that in a province with a certain amount of places, the families that live in the bigger towns, are situated in an area with a high level of accessibility to the service due to the high concentration of facilities, while families that live far from those cities may have severe difficulties to find a facility close to the family location. Even if they belong to the same province, the aggregate absolute number of available spots, namely the child

coverage ratio, does not tell us enough information, we should then also focus on how those facilities are distributed in the administrative units.

To correctly understand the importance of the child coverage ratio we should then ask ourselves how effectively spatially available is the service. Another important consideration is that, even though one of the eligibility criteria that the municipalities use to ration the demand is the residency of the family, this consideration is not relevant when we look at the private sector. Due to the lack of provision of the service, it is also possible for many municipalities to make agreement in order to supply it to families that live in towns where the service is not present at the cost, for the family, of a longer travel to reach the facility[1]. In many cases the public sector allows also for special regulations, for which it is possible to enroll the child in a public structure located in a different town from the origin one, when both or one of the parents work in that town with a penalization in the position of the waiting list of the structure. As Del Boca and Vuri (2007) stated in their paper, the distance between demand and supply is an important component of child care cost and availability; so we should consider inside the expected cost function for the attendance of the formal child care system, also the transportation cost, both in terms of monetary and opportunity cost:

$$E[C_{if}] = p_{if} + z_{if} + \alpha * transport_{if} + \beta * time_{if}$$

The expected cost for a family i to enroll the child in a formal structure f -we do not control now for the child coverage ratio- will depend on: the price, or daily fee p, that the structure charges to the family; other general monetary and non monetary cost and subsides z and the cost of transportation, *transport*, that the family would spend

[1]Agreements are stipulated at the municipality level, it is not possible to define a common regulation. As an example "Unione dei comuni del Medio Brenta" is a case in which more municipalities decided to cooperate in the management of the early child care system

to take the child to the structure adding also the opportunity cost of the time spent for the travel, *time*. The family, before knowing if the child will be accepted in any of the structures, creates a set of "acceptable" and "not-acceptable" facilities. In the acceptable set we find structures located close to the family, the transportation cost and the opportunity cost are not too high and they do not affect strongly the expected cost function, families would then try to enroll the child only to structures belonging to this set. In the set of the not-acceptable, we find facilities located too far from the family, the parents would never consider one of them for the enrolment of the child in the formal early child care system because the costs related to these structures are too high (Del Boca and Vuri, 2007). To discriminate if a facility belongs to the acceptable or to the not-acceptable set, it is common to use a threshold; assume for example that a family considers in the acceptable set only the facilities located within 15 kilometers from the house, or the facilities reachable in 10 minutes by car. It is important to stress that, to be complete, this analysis would also have to take into consideration the distance between the workplaces of the parents and the child care buildings, not just the distance from home. It is known in fact that while the mothers work usually close to the house, fathers tend to work further and could for this reason bring the child to a child care structure located even in different provinces than the one of residence (Van Ham and Mulder, 2005).

We should also take into account that, when a family has the possibility to choose, ergo has many facilities that belong to the acceptable set, parents can rank these structures from the most favoured to the least one and then choose in which one to try to enroll the child. This situation would be always better than the one in which the options inside the acceptable set are limited to one or few, that is when only one or few child care centers belong to the acceptable set. Considering the spatial relation and the love for the variety, we should create

a framework in which we reward the families that not only have the possibility to enroll the child to a facility but that can also choose in which facilities to do it. The rationing system taken into consideration in the cited literature, the child coverage ratio, will display its effect only after the rationing procedure that the families adopt to determine which facilities belong to the acceptable set, according to the spatial relation.

As stated before, the spatial analysis allows for the control of the variation in the provision of the early child care system among areas of the same province or region. We expect in fact a great amount of facilities to be situated close to the most populated centers, leaving the peripheral areas with only a limited number of structures. The spatial procedure would help to consider the heterogeneity inside a certain area, explaining that there could be not only disparity among Italian regions but also among areas of the same region or province. In a developed society in which it is possible to reach a wide number of destinations in a limited period of time, in which the behavioural component can have a strong impact on the family choices, it emerges to be necessary to analyze the spatial relations that can occur between the variables; in our case to assume that families that are living in areas with a lack of early child care facilities can move and enroll the child to structures located outside the municipality in which they live. Assume for example that a father lives with the family in the city "A", where there are not early child care facilities, and that there exists another relative which lives in the city "B". The father works in a third city, "C", and to reach the workplace he has to pass through the city in which the relative lives. Accordingly to the economic theory, not having facilities close to the home and the workplace, he decides that for the care of the child, the right solution is to leave the child to the relative while going to work. The relative in city "B" agrees, but being aware of a child care facility in another city "D" close to his

home, suggests to the father the option to be the one to bring the child to the formal child care in city "D" and to bring it in its home until when the father would come back home from job. In this example we highlight two aspects: the interactions that can occur between formal and informal service, already known in literature, and the fact that it is possible for a family to enroll the child in a facility located far from the house under particular circumstances.

A useful tool in the description of spatial relations, arises from the development of more and more efficient and precise geographic systems able to map a wide number of different type of variables and to compute spatial analysis that are precluded to other statistical software. From being predominantly used by geographers, engineers, architects and other technical professions, the use of this system has began to take hold also in Economics as a support to the standard econometric analysis. The spatial analysis is not intended to substitute the economic one, but to reinforce it through the inclusion of new variables in line with the microeconomic models adopted. This system is commonly known as G.I.S., meaning Geographic Information System, which predominantly aim is to capture, store, analyze and manage geographic data to identify spatial relation between variables.
Only few researchers have used this tool to investigate the impact of spatial rationing related to the formal child care system. The procedures they adopted, share the idea to investigate the relation between the position of the families and the position of the facilities to determine which areas in a country are the most well served and in which instead the access to the service is scarce. The weak point of this analysis is the lack of the necessary data, due to privacy reasons, that forces the researchers to use approximations to overcome this issue. Before considering the specific GIS analysis that we adopted in our model in the following chapter, we describe two researches from the related literature as example.

Van Ham and Mulder (2005) adopted a GIS spatial analysis to introduce a distance related variable in the Netherlands, to study its impact on the employment choice of the mothers. They investigated the mechanisms by which geographical availability of formal child care influences mothers' labour force participation and tested this assumption in an econometric model. The attention is stressed on the time constraints that limit the set of daily possible movements of the mothers (travelling time to reach the workplace and the child care facilities, combined with the working time and the opening hours of the facilities); mothers that live close to those "destination points" have higher probability to both work and having the child enrolled in the child care service. The methodology used by Van Ham and Mulder (2005) is common to other researches as in Langford et al. (2018) and consists in three steps: firstly it is decided how to identify the households, it may seem obvious to consider the exact location of the families with a child in age 0-2 (or different) but this kind of data is usually protected by privacy constraints, therefore it is necessary to use a proxy for it; secondly it is introduced the supply side, namely the location of the facilities and thirdly it is computed the analysis.
Van Ham and Mulder (2005) used as sample the mothers with children in age 0-6 from the Netherlands Housing Demand Survey 1998, the dataset included information on the families and the locations of the houses according to the postcode area. For the supply side they used the number of slots of the child care system per postcode from the Monitoring Agency Child Care Provision. Both for the households and the child care facilities, the authors were not able to know the addresses of the locations to be considered, so they use as proxy the centroids of the postcode areas in which they belonged to; for centroid we intend the geographic centroid of the area in the geometric assertion. For each centroid they gathered the data related to the total number of children living in the area, the number of spots available

and other information. They then built a time travelling matrix to investigate how many connections between the centroids were possible in a 10 minutes travel by each of them. Knowing the aggregate number of children and the aggregate number of slots per postcode, they calculated how many slots were available per child per postcode, considering that families could move from the origin centroid to other centroids closer than a ten minutes travel. The final result they obtained was a variable measuring the number of child care slots per 100 children in each postcode area. They introduced this variable in an logistic model obtaining a positive significant relation between the probability of being employed and the geographical variable.
Even if aware of the approximation made, this model represents a way to address the spatial problem trying to understand the interactions between the family location and the facilities of the child care system.

A more didactic paper has been presented by Langford et al. (2018); analyzing the Welsh framework they presented a detailed procedure with an accurate description of the steps taken to create the geographical variable. Differently to Van Ham and Mulder (2005), Langford et al. (2018) accurately explained how using administrative units to define an accessibility variable may lead to an omission of the description of the internal variation of the accessibility to the service. It may happen in fact that in a large region, province or municipality, the provision of the service would result to be not homogeneous, with areas in which there could be a completely lack of the early child care service. So having a high value of child coverage ratio, does tell us only a part of the real availability because it does not consider the spatial dimension; if the facilities resulted to be far from the location of the households, it would be more difficult for the families to reach them. Using administrative borders to define the accessibility variable, results to be a limit because families (under specific circumstances) may choose early child care centers that fall outside the administrative zone

in which they live.

The accuracy in the resulting spatial variable would depend on the precision of the data in possess of the researchers, the privacy policies strongly limit the use of the geographic systems, reducing the power they would have and forcing the researchers to adopt approximations that, even if difficult to implement, represent good starting points for a first comprehension of the spatial relation. Having more detailed information would not only help the researchers but also the policy makers to correctly operate to obtain social optimal solutions. It is not only the privacy the cause of the absence of the data, sometimes those type of data are not even collected by the public institutions.

3 GIS software and early child care analysis

In this chapter we are going to discuss the use of a Geographic Information System (GIS) software for the description of the formal early child care provision in Italy, all the details on GIS can be found in Appendix A. We describe the procedure adopted to build the spatial accessibility variable and the results obtained at a regional level. Even if the child coverage ratio has been demonstrated to be a variable able to depict the effects of the formal early child care system, the use of a spatial approach could help to better characterize the impact that this service has on the families and the mothers. Using a spatial variable we are able to identify the spatial distribution of the early child care facilities investigating the effects that distances between the structures and the location of the families have on the familiar context. The child coverage ratio, being the ratio between the available spots of the early child care system and the number of children between 0-2 years old of age, represents a value considered homogeneous in a certain administrative unit; consider a region, this variable assumes that the provision of the service is equal among all the villages and cities of the region. We consider this assumption to be excessively strong because in an administrative unit it is possible to find areas that are well served by the service and areas in which there is a substantial lack of it. Assume a region with a medium high level of child coverage, if all the facilities are located in the biggest cities, families that live far from them are not affected by the provision of the service and the child coverage rate is not able to capture this situation. Using a spatial approach we are able to measure and compare the accessibility to the service of different areas of a region with respect to the spatial distribution of the early child care facilities and to the spatial distribution of the population, considering also the possibility for the families to move to reach the service from the municipalities in which they live to other close municipalities.

The use of GIS to measure the potential accessibility to a service, is something common in the literature and represents one of the main issue in physical planning. As explained in Geertman and Van Eck (1995), accessibility issue can be easily investigated using GIS features; with accessibility we define the possibility to reach a location, a destination point, from the origin one at suitable time or considering a set of characteristics from the dataset that can favour the access or impede it (distance, transportation mean, ability of the person to move and so on). The accessibility will then depend on subject specific preferences, on person's mobility and on the spatial environment surrounding the observations. For example one location can be more accessible than another because it is more close to the origin point or because is better served by the transportation system. In the early child care context, accessibility has been widely defined using the number of places over the target population, in the spatial framework we are going to combine this variable with a spatial accessibility variable, to create synergies between these two sources of information and better characterize human behavior considering people's daily movements and preferences. How to assign the final accessibility value to each family it is up to the researcher and to the model he has decided to use; one of the most used model is for example the gravity model, modified by the different authors according to the case of study, for which the observations can be strongly influenced by their relative position in a map and by the distances they have with each other (Geertman and Van Eck, 1995). The use of GIS for the analysis of accessibility to the service has been widely adopted also in the Health Care literature (Love and Lindquist, 1995; Foley, 2002; Hyndman and Holman, 2001; O'Dwyer and Burton, 1998; Philips et al., 2000). Many researchers exploited the geographic system to understand the relation between the providers of the service and the users, investigating the number of potential users of the hospitals in a country, the average

time to reach the facility considering the dispersion of the population and so on.

When introducing an accessibility model, many issues arise regarding the correct location of the input data and the method to use to build the accessibility variable. For the definition of the accessibility variable we first need to identify how to represent the spatial provision of the service and the distribution of the population. Limitations due to the privacy policy force the researchers to use approximations to spatially locate these data. In particular it appears almost impossible to acquire information on the precise location and composition of each family in a country, researchers have then to identify valid proxies for the description of the data at the cost of reducing the degree of precision of the model. One of the most used method in the literature provides for the use of a point-to-point analysis, a calculation accounting for the distances between all the possible combination of points within a fixed search radius (may it be in spatial or time unit). The population is usually defined using a set of points, the centroids, representative of the administrative units considered or of a regular grid for which we have the information about the number of inhabitants. Each centroid incorporates the list of attributes of the polygon (the administrative unit, the regular grid and so on), namely the average values of the variables we are interested in: number of total population, number of child care facilities, name of the municipality and so on. The supply side instead may be defined in various ways according to the kind of data in possess of the researchers; in fact while it is almost impossible to know the address of the families that are part of the sample of the research, that is the reason of the need of approximation for the population distribution; for the supply side it could be easier to have access to database in which it is specified the precise locations of the early child care facilities. The spatial representation of the supply side would then depend on the degree of precision of the

data, if the reasearcher knows only in which municipality are located the facilities it would have to use an approximation at an administrative level, if it is in possess of more refined data it can improve the degree of precision of the location of the structures. The best dataset for the supply consists in the list of early child care facilities with the addresses and the number of spots available for each structure.

3.1 The spatial model

The first step in the process of construction of the spatial model for our research, regards the decision on the software to use and the search for the basic data to build the map of the Italian borders. Within the wide range of GIS applications, we chose to use QGIS, an open source desktop application that, even if less complete than the most famous but expensive counterpart ArcGis, it incorporates the features required for our spatial accessibility analysis, being the best solution for the GIS beginners (QGIS, 2019). After having decided the application to use, we started to look for the data to define the Italian boundaries to proceed with the implementation of the model and of the observations related to the early child care system. Looking for a vector containing the Italian administrative boundaries (country, regions, provinces and municipalities), we found in the ISTAT website, a dedicated section for "Territory and Cartography", with the information that we needed. In particular we have been able to download the information on the boundaries and on the population census of 2011 to represent the population distribution.

As reference year of our research we decided to use the 2013 for many technical reasons, the most relevant concerns to the availability of the data. The 2013 was one the most recent year for which it was possible to have specific data on the Italian child care system, in particular the child coverage rate level for all the Italian regions and provinces, and the list of the number of the child care facilities

in each municipality. While we could have preferred to use the 2014 or 2015, the 2013 was chosen also for the temporal proximity to the national population census of the 2011, whose data have been used in the spatial analysis for characterizing the population distribution of the spatial model.

Aim of this research is to address the spatial early child care issue trying to control for the geographic variation in the distribution of the population, the population in fact is not evenly distributed across a country, not even across a region or a municipality. The location of an early child care structure has a spatial influence on the surrounding environment, the distance between the structure and the family plays an important role in the familiar decision on the enrolment of the child to the formal early child care system due to the travel cost that the family would have to spend to reach the facility. Moreover, the parents when looking for a facility, do not know the level of child coverage ratio in the administrative unit, they firstly observe which facilities are more accessible considering the distance from home or from the workplace. Only after having decided which are the facilities in which they would try to enroll the child, the child coverage ratio and the eligibility criteria would become extremely relevant defining if the child would access or not to the service. In the early child care framework it is important also to consider other variables as the employment decision of the parents, the daily movements of a family and the social interactions that the members of the family can have outside of it (relatives and friends for example). In doing so we try to reconcile both the need to consider the political administrative boundaries, and the desire to be able to study the smallest possible aggregation unit of the population without making approximation using the centroid of a very large polygon as an entire municipality or province. In doing so we looked for geographic data from various sources and faced some severe limitations due to the impossibility to have access to specific

type of data as the addresses of the early child care facilities that are present in Italy, because protected by the privacy policy. Nevertheless we have been able to obtain a variable describing the accessibility level of the early child care service exploiting the spatial location of the population and of the early child care provision in the country.

3.1.1 The spatial distribution of the early child care facilities

Before talking of the spatial representation of the supply, we describe what we consider to be a child care facility in our research. We consider an early child care facility to be a structure dedicated to all the children between 0 and 2 years old of age, recognized by the national authority with qualified childminders. In the set of the analyzed facilities we gather also the child care facilities for children between 3 and 5 years old of age, so far as in these structures there exist also specific classes for the children in age 0-2 ("Sezioni Primavera"). We do not differentiate the supply between the public and the private sector because we are interesting in the whole early child care system. The early child care facilities we considered were 13,459, and represented the total supply present in Italy in 2013.
The information we needed for the spatial model, the distribution of the early child care structures in the Italian territory, were not present in the official database and, even after a formal request, it was not possible to access to the information on the addresses of the early child care facilities; knowing the exact location of the structures would have allowed us to compute a point-to-point analysis between the families and the facilities. Unfortunately, this kind of data are considered sensible data, so at the moment of writing it appears to be extremely difficult to have access to these information. Although this issue strongly influenced the construction of the spatial model, we solved the problem using an approximation. In the ISTAT dataset on the early child care system, it was possible to know the number of

Table 5: Number of facilities per region in 2013. Trento and Bolzano provinces have been considered separated due to their legislative powers.

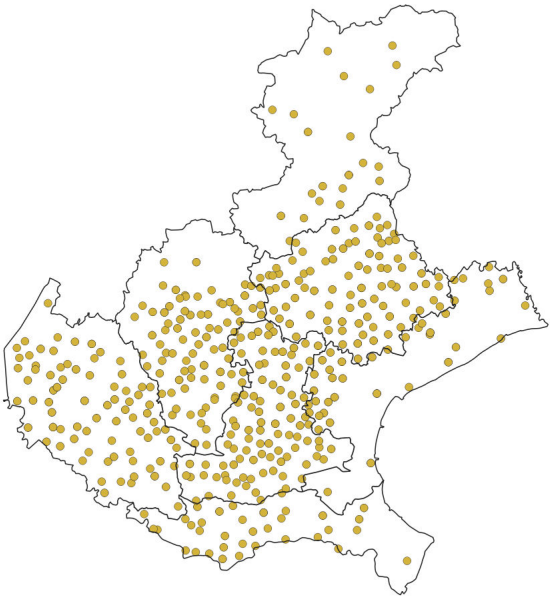
Region	Number of facilities
Valle D'Aosta	70
Piemonte	1056
Liguria	395
Lombardia	2993
Bolzano	236
Trento	229
Veneto	1312
Friuli Ven. Giul.	298
Emilia Romagna	1232
Toscana	1073
Umbria	307
Marche	327
Lazio	1310
Abruzzo	225
Molise	87
Campania	461
Puglia	532
Basilicata	80
Calabria	327
Sicilia	506
Sardegna	403

ISTAT.

facilities for each municipality so we decided to use this data as the approximation for the distribution of the early child care facilities. We built a dataset in which we listed all the municipalities and the number of facilities for each administrative unit, we converted the dataset in a GIS format and we mapped the outcome.

In Figure 2 we mapped the early child care supply for the Veneto region. Using this figure we can describe the approximation we made in order to map the early child care facilities. Because the extension of the dataset that we created for the supply side was not compatible with QGIS, QGIS elaborates external data only through the geographic coordinates of the information we give to the software, we converted the dataset in a QGIS compatible extension using the Google Earth Pro

Figure 2: Early child care facilities in Veneto region.



Own elaboration.

(GEP) application (Google Earth Pro, 2019). This software used the name of the municipality to assign to each row of the dataset a geographic coordinate, but doing so, not knowing where to place the early child care facilities of each municipality, it located them in the centroid of the municipality they belonged to. Instead of having the precise positions of the structures, we obtained a map where the facilities have been located in the same place of the centroid of the corresponding municipality. Another issue arose when for a municipality we had more than one early child care structure, in fact in this case the GEP application stacked in the centroid position a number of points equal to the number of structures and, as a consequence, we obtained the identical situation in QGIS when we mapped the converted dataset. We can not accurately describe this issue using the map produced by QGIS because, being be-dimensional, we see the vertical pile of points

as a single point. The supply side of our model is then defined as a set of points representative of the early child care facilities present in Italy in 2013, located in the centroid of the municipality they belong to.

We are aware of the inaccuracy that this representation of the supply side leads to the model, but the absence of the addresses in the official national dataset and in the regional ones, forced us to adopt this strong approximation. We are going to dedicate an analysis of the measurement error that we introduced in the model using this approximation at the end of the chapter, after having described the spatial accessibility variable.

3.1.2 The spatial accessibility procedure

In the economic literature, the variable that better describes the interaction between the quantity demanded and the quantity supplied is simply the ratio between the number of spots and the total number of children between 0 and 2 years old of age in the population. This variable is then used to describe the effects of an improvements of the early child care system on other variables as the maternal labour force participation. In the spatial context instead, we want to observe what would be the effect of variations in the actual location of the service and the effect of an increment in the number of the facilities accordingly to the spatial distribution of the population.

The accessibility variable we want to find should be able to describe the decreasing influence that an early child care facility has on the surrounding space as we move away from it. The variable can then be interpreted in two different ways; if we look only at the spatial provision of the service we can observe which areas in a region result to be the most accessible, if we introduce the distribution of the population, dividing between inhabited and uninhabited areas, we can describe the average accessibility level for the families living in a

certain municipality, province or region, considering only the inhabited areas of the country. In this chapter we will initially describe the spatial provision of the service, meaning the location of the facilities and their spatial influence, and only after we are going to introduce the distribution of the population to observe the spatial accessibility value for the families living in a region.

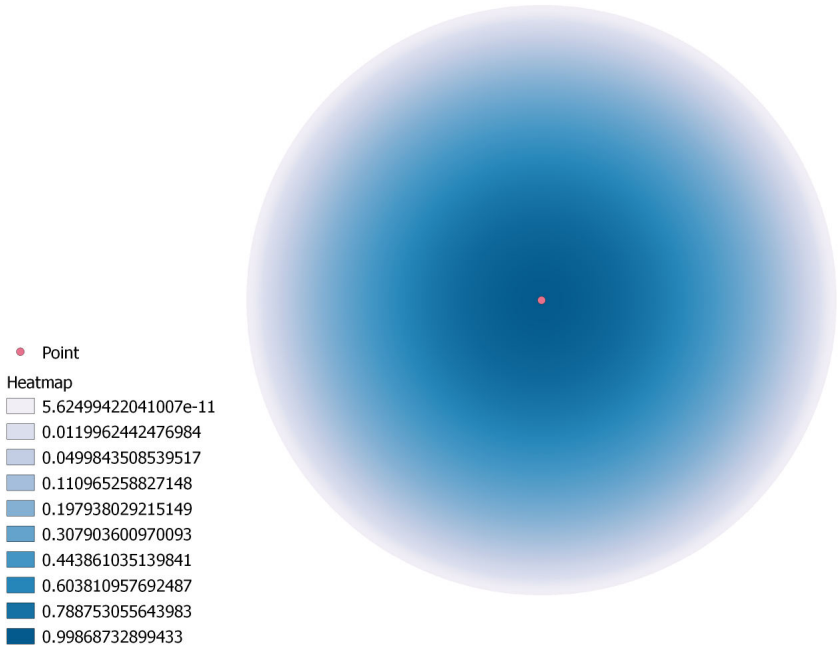
The variable we want to define has also to respect the norm for which if a family is located in a place influenced by the presence of more facilities, its corresponding accessibility value should be higher than the one of a family that lives close only to one early child care structure. In the specific, if a family lives 15 kilometers away from two different early child care structures, we should assign to it an accessibility value higher than the one we give to another family that lives 15 kilometers away from only one structure. The accessibility variable we are going to propose would not be weighted by the number of spots available for each facility for many reasons. The most important is that the information on the number of spots per facility could not always be considered reliable because, as the provider of the data (ISTAT) stated, the collection of the information about the number of spots of each facility in each municipality has required the introduction of an estimation component due to the decentralized framework of the early child care system; moreover, in the analysis we would always consider the child coverage ratio to be the variable that better describes the relation between the quantity demanded and supplied, any further econometric analysis will then consider both the child coverage rate value and the spatial accessibility value.

The method we decided to adopt is the kernel density estimation (KDE), an interpolation procedure. QGIS has a specific feature for this calculation and the final outcome consists in an heatmap describing the intensity of the supply proportionally to the distance from it. For the description of the kernel density estimation procedure we refer

to an important book of B. W. Silverman (2018), details in Appendix B.

The algorithm that QGIS uses to elaborate the kernel density estimation does not allow us to specify the kernel estimator but only its kernel function, that is the function that defines the shape of the density function. Using the KDE procedure we are able to investigate the influence that each early child care facility (represented by a point) has on the surrounding space. What we obtained were smoothly curved surfaces centered over each facility, in which the surface values (the density values) are highest at the location of the point and slowly decreases proportionately to the distance from it, reaching the value of zero at the limit of the surface defined by the search radius.

Figure 3: Heatmap generated by QGIS.



Own elaboration.

In figure 3 we observe a general example of an estimated density function associated to a single point elaborated by QGIS. The formula

of the kernel function defines the shape of the surface; if we consider a random point on the surface, its density value would change with respect to the formula used for the kernel function. Some functions are more concentrated around the point, leading to a steeper slope, others assign higher values to the tails producing smoother curves. If we would have wanted to move from a flat graph to a graph representing the heights of the curved surface we would have seen the peak of the function to be on the vertical line with respect to the point, the unique observation, with decreasing symmetrical values as we move towards the tails. In QGIS the outcome consists in a layer called heatmap, by its ability to display the area in which a distribution of points is more clustered, just varying the intensity of the colour or the colour itself according to the density values of the surface. In the case of a single observation, QGIS assumes the values of the heatmap to be between 1 and 0, with the highest value close to the observation and the lowest values on the borders of the heatmap.

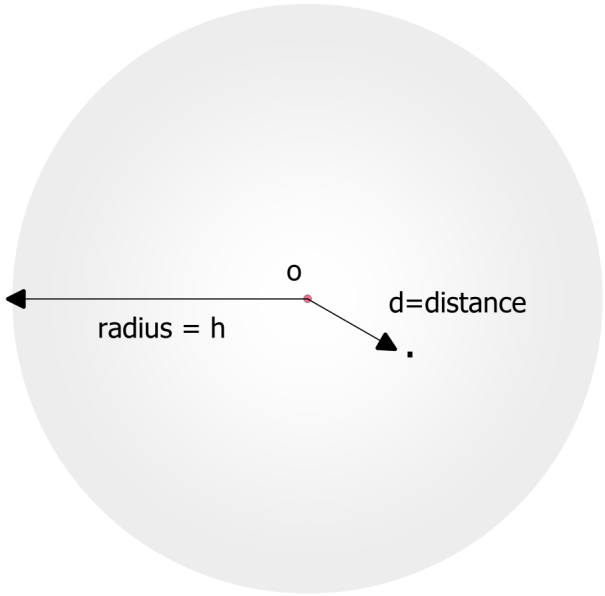
Each kernel function has specific characteristics with respect to the way in which it defines the probability density function; for example among the most efficient kernel density function, we find the Epanechnikov function and the Quartic (Biweight) function with the latter that assigns higher values to the points located close to the observation than the former. For this reason we decided to use the kernel function suggested by QGIS, the Quartic function, in particular because we do want to assign higher values to those points that are closer to the early child care facility[2].

In figure 4 we described the parameters that characterize an heatmap, parameters that strongly influence the extension of the curved surface and, as a consequence, the values assigned to it.

The bandwidth of the kernel density function is described by the search radius h, the threshold distance outside of which the influence

[2]QGIS does not explain the formula of the kernel function used. We assume the quartic kernel function used in QGIS to be equal to $K(x) = \frac{15}{16}(1 - x^2)^2$.

Figure 4: Parameters of the heatmap.



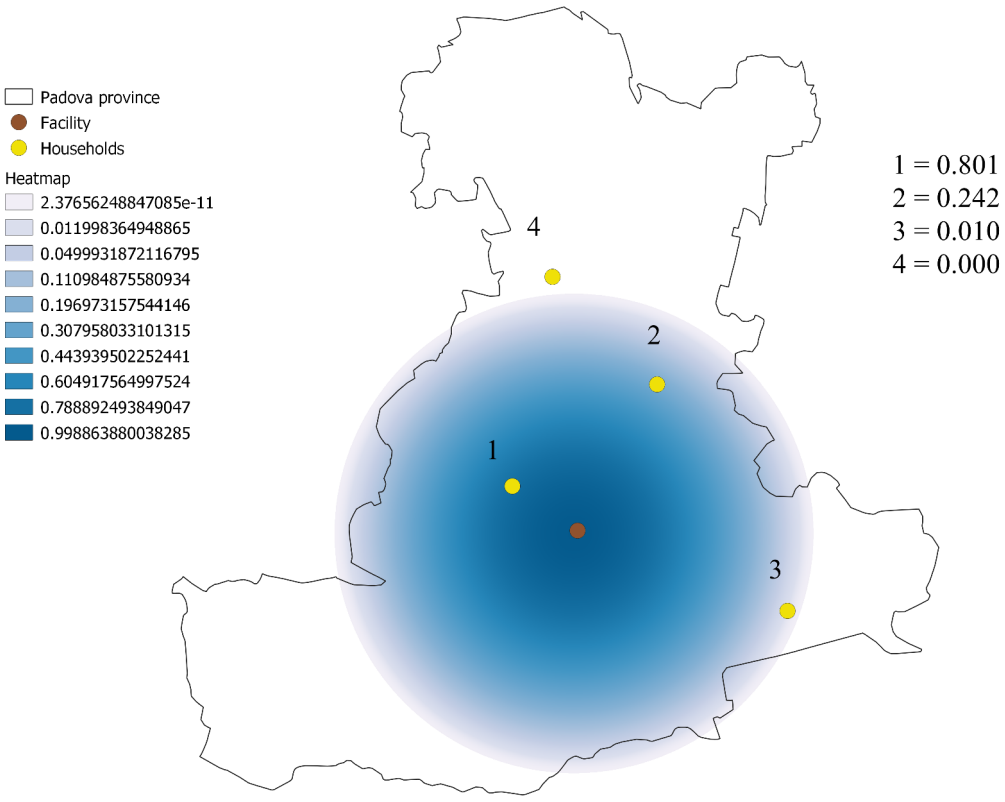
Own elaboration.

of the point in the surrounding space is equal to zero; the reference point o, the observation that is the center of the heatmap; the distance d, the criterion through which we define the surface values of the curve. For the aim of the research, we decided to use a large bandwidth fixing $h = 20,000$ meters. We considered this value to be the compromise between a too tight radius, meaning the impossibility for a family to enroll the child in facilities located close to the house, and a too large radius, that would have led in the acceptable set of facilities, also structures that a family would never choose simply because too far from its location. Even if arbitrary, a search radius of 20 kilometers is not too large to completely obscure the distribution structure of the density function, because its relative value with respect to the Italian territory is minimal.

In our research we allow for the possibility for a family to enroll

the child not only in any structure located in the same municipality in which the family lives, but also in facilities located in different municipalities and different provinces. This constraint defines that an early child care facility can be chosen by any family situated in the same region of the facility and no more far than 20 kilometers from it. In the following hypothetical examples we are going to discuss all these cases to understand the functioning of the heatmap procedure.

Figure 5: Single early child care facility and households belonging to the same province.



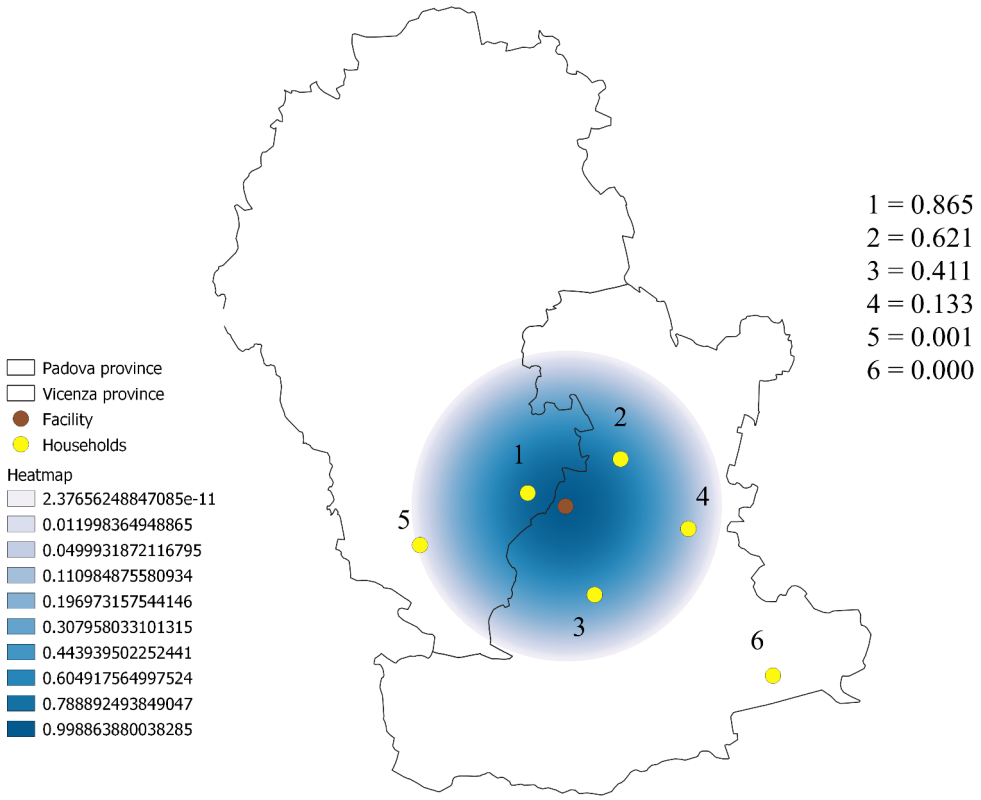
Own elaboration.

We begin by analyzing, in Figure 5, the case in which a single facility and a set of points representing the households are both located in the same province. We do not consider the case in which these points are located in different municipalities because we explicitly stated previously that we allow for the possibility to enroll the child in facilities not located in the municipality in which the family lives. In the map

it is possible to observe the heatmap generated by the early child care facility and its interaction with the locations of the families. A specific algorithm in QGIS is then able to assign to each point the corresponding density value with respect to its location on the map. Visually we can observe the variation of the values of the heatmap just looking at the variation of the intensity of its colour; a darker blue represents the area with the highest values, that are correctly in the closest neighborhood around the facility, while the lighter shades that progressively turn into a grey represent the decrease in the surface values. On the right of the figure we ranked the households, that we locate randomly, by the values that the algorithm assigned to the points representing them. Correctly we have that point 1 accounts for the highest value being the closest to the facility, point 2 an 3 have both a positive value lying inside the influence area represented by the heatmap (bandwidth 20 kilometers), with point number 3 located on its border to which has been assigned a very small value. Point number 4 lies outside the heatmap and this is the reason why the algorithm assigns to it a null value. QGIS and the heatmap procedure do not differentiate points that are beyond the border defined by the search radius. Suppose to have more points outside the heatmap, the system would treat them in the same way assigning a value of zero to all of them, being a specific point very close to the border or kilometers away from it. In the early child care system we interpret the value assigned to the points in which families live as the accessibility value they have with respect to the early child care supply, that in this case is represented by only one facility. The facility appears to be more accessible for point 1 than for the points 2 or 3, even if for all of them the facility belongs to the acceptable set, meaning is a structure that, being positioned at a distance that is closer than the threshold of 20 kilometers, can be selected as the child care facility in which to try to enroll the child. A family living in point 4 would not be influenced at all by the pres-

ence of this structure, being it too far from its relative position (not acceptable set); the transportation cost and the opportunity cost that the family would have to spend to reach the facility are too high.

Figure 6: Single early child care facility and households belonging to different provinces.



Own elaboration.

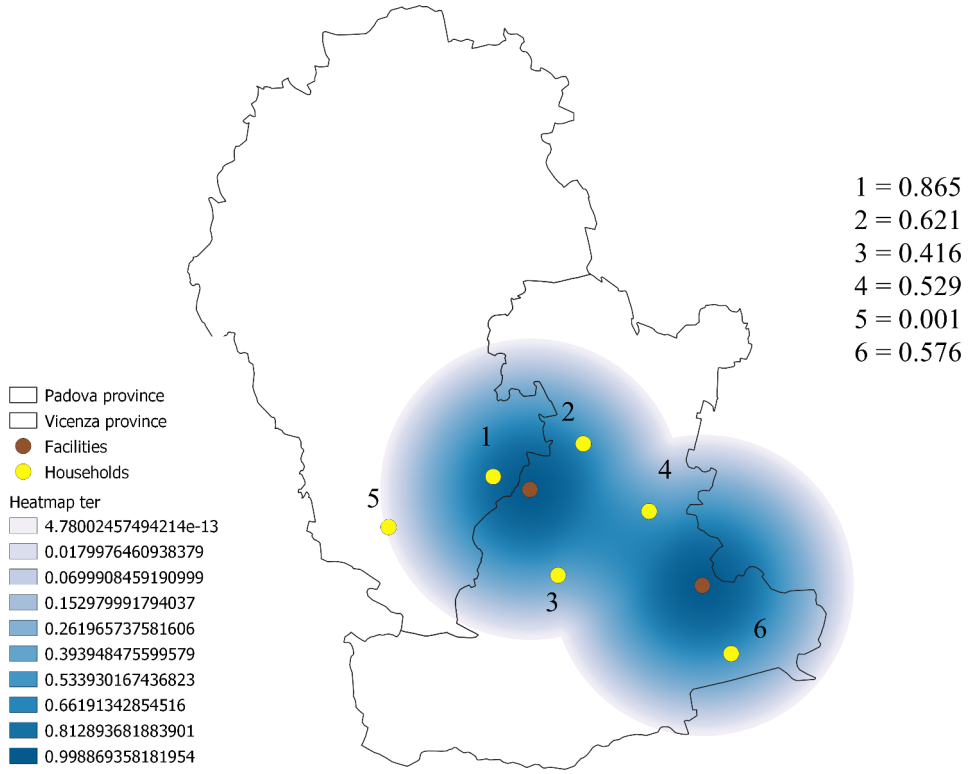
An interesting situation is when a facility is situated close to the boundary of a province, as in Figure 6. The two provinces are delineated by the black lines representing the different boundaries. In this case, a family living in a yellow point, can be located in a province that is different from the one in which the facility is situated but still extremely close to it, as point number 1. For situations like this, we adopted the same reasoning that we used for points belonging to different municipalities. Due to the existence of a widespread private sector, that is not constrained by eligibility criteria that bound the possibility to enroll the child only to the family living in the same

municipality in which is located the early child care facility, and due to the possibility for the public sector to receive infants from different municipalities (under particular circumstances); we allowed for the possibility for a family to consider in the acceptable set of early child care facilities also the structures that are located in other provinces, although below the distance threshold. In the map we observe as the facility is located in the right province, the province of Padova, while the family with the highest value is in the left province, the province of Vicenza. As in the previous case QGIS assigns to each point a value corresponding to its accessibility value with respect to the supply of the early child care system.

Until now we have described the case in which the supply side consisted of only one point, one facility. In the following case, Figure 7, we increase the number of early child care structures to two and we analyze the new shape that assumes the heatmap. When we have more than one point, the heatmap generated by the early child care provision accounts for the possible spatial interaction that can occur between them, in particular when the distance between the points is $d < 2h$.

In this case in fact the algorithm creates a unique density surface. All the points representing the families that are less distant from the supply than the search radius, can then be influenced by the simultaneous spatial presence of both the facilities. Exploiting the case in Figure 7, we identify the point number 4 to be less than 20 kilometers distant from both the two early child care facilities, a family living in this position would then have the possibility to choose between the two facilities, firstly trying to enroll the child to the most preferred (the closest) than to the least preferred (the furthest). The density values that are assigned to the points that lie in this particular area are then derived by the spatial interaction due to the distribution of the facilities. Taking a closer look at the map we observe as the points

Figure 7: Early child care facilities and households belonging to different provinces.



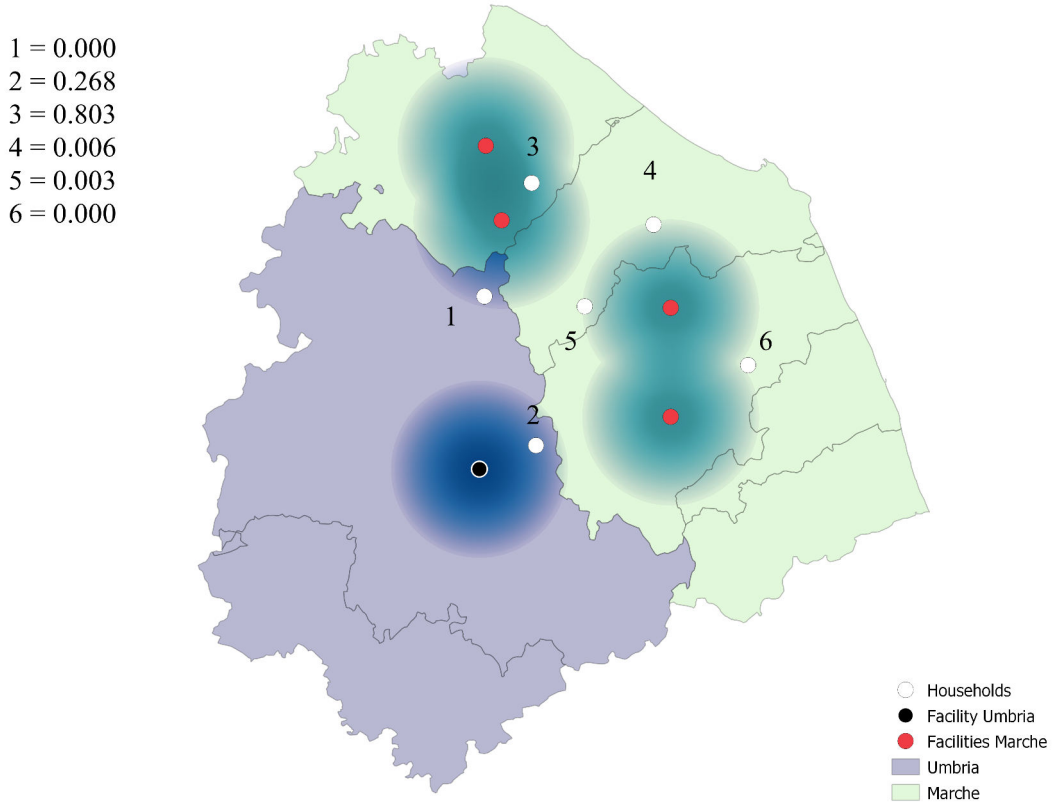
Own elaboration.

3 and 4, even if almost at the same distance with respect to the facility on the left, have different accessibility values; point number 4 in fact benefits from being influenced also by the presence of the structure on the right, while point number 3 interacts only with the part of surface generated by the one on the left. When we rank the points by their accessibility values we have that even if point number 4 has the possibility to choose between two facilities, its value is anyway below than the one of the point number 1 that is extremely close to one of them. This situation follows our interpretation of the early child care system, in fact even if a family located in the point 4 has the possibility to choose between the two options, it is extremely far from the two facilities so its level of accessibility has to account for both the high distance and the possibility to choose, and that is why the accessibility value of point 4 is higher than the one of point 3 but lower

than the accessibility of point 1. While the lower bound of the spatial accessibility value is 0 for construction, the upper bound in presence of more facilities does not correspond to 1; in particular if we have many facilities very close one another, the density values corresponding to the area in between can be higher than one depending on the number of facilities and the relative distance between them.

The last case we define is when an early child care facility is located close to the regional boundary, Figure 8. In this case we observe two Italian regions, and two different vector of points representing the supply, the red dots and the black dot. Considering the early child care system and its decentralization, we know that there exist different regional regulations and that for this reason it is not immediate to allow the possibility for a family to enroll the child in a facility located in a different region.

Figure 8: Facilities and households located in different regions.



Own elaboration.

In particular for the public sector, where it may result complicated even the enrolment of the child in a different province, we exclude the option for a family to enroll the child in a facility located outside the region in which it resides. Referring to the private sector we recognize that this case may happen, but being entities independent one another it is almost impossible to discern a general criterion to adopt to include this possibility, so we preferred to exclude this option also for the private sector. The consequences of this constraint are described in the map; observing the location of the point 1 we note that, even if it is influenced by the presence of a facility situated in a different region, due to the constraint we adopted, the accessibility value of this point results to be equal to 0. If we would not have considered the regional boundary, the accessibility value would have been positive and high. Giving a general description of the accessibility variable in this scenario, we define the point number 3 as the one with the highest level, 0.803, being influenced by the presence of two early child care structures, followed by the point number 2 that is located in the other region and is influenced by the presence of a facility located in the same region. The other values reflect the various scenarios we described in the previous examples.

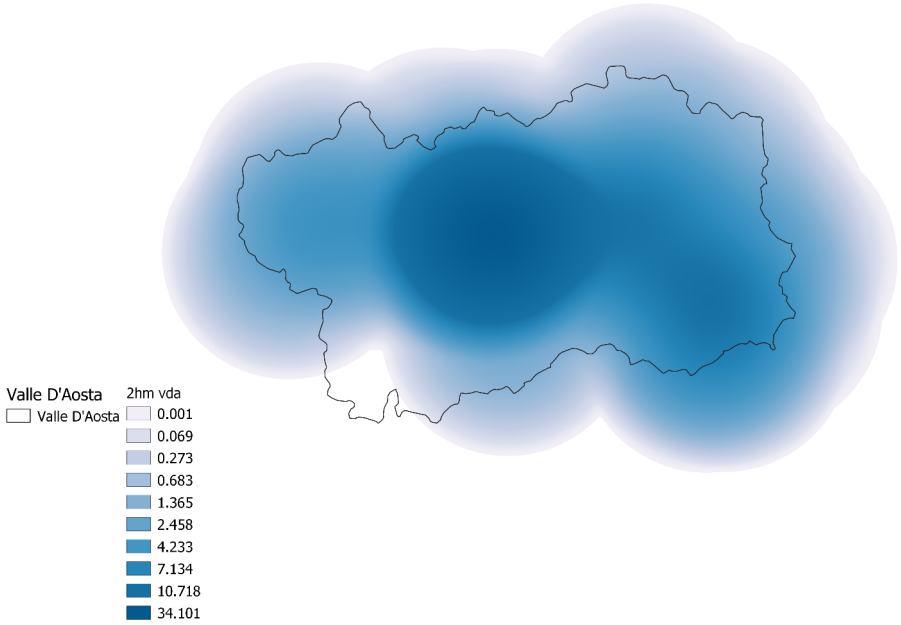
3.1.3 The regional analysis

In order to map the effective spatial availability of the early child care supply present in Italy in 2013, thanks to the constraint that limits the spatial influence that a facility has, only to be inside the regional borders; we built for each of the Italian region the corresponding heatmap to visually analyze the regional supply of the early child care service. The data on the number of facilities per municipality has been elaborated by the database ISTAT and then mapped with the use of QGIS. To facilitate the comprehension of the differences existing between the regions, we grouped them in macro areas:

North-West, North-East, Center, South and Islands. Each region has been assigned to the corresponding macro area exploiting the division used by ISTAT. It is important to remember that, being the heatmaps independent one another, the palette of colour used, that is the same for all the heatmaps, calibrates the intensity of the colour accordingly to the distribution of points specific of the region. That is why if we look at the map without considering the legend, we would wrongly state that the highest availability level in Valle D'Aosta (Figure 9) is the same of the one in Lombardia (Figure 11) because the colour that defines the maximum values is the same; instead, using this regional representation, the colours only describe the availability level relative to the treated region. To compare different regions we should observe the values of the deciles that appear in the legend of each region. The aim of these maps is to show the distribution of the early child care facilities to help a visual comprehension of the supply side, these heatmaps do not represent the accessibility variable because they do not consider the distribution of the population, but they can be used to visually understand the spatial influence of the early child care system. These heatmaps permit us to appreciate the intra-regional variation in the provision of the service and to visually identify the existence of areas where there is a lack in the diffusion of the service.

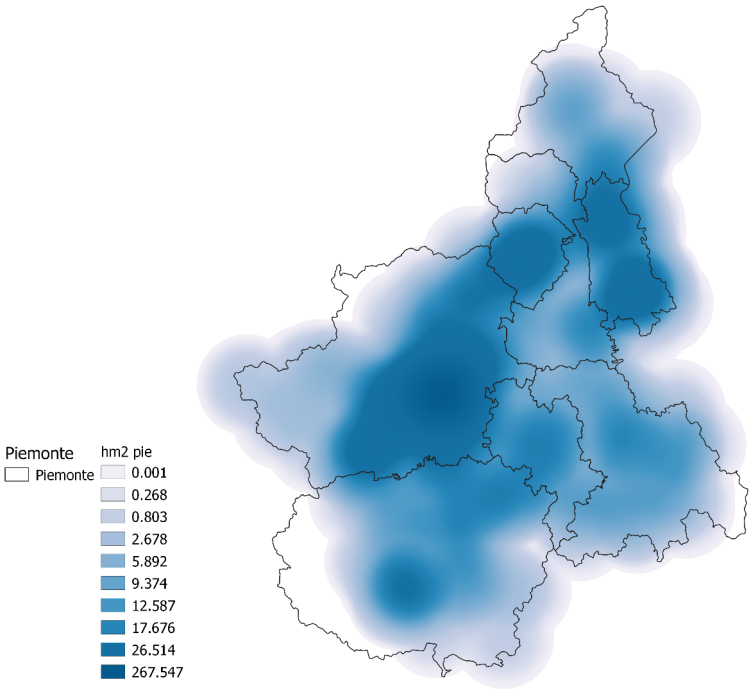
In the North-West of Italy we count four regions: Valle D'Aosta, Piemonte, Liguria and Lombardia.

Figure 9: Valle D'Aosta, heatmap.



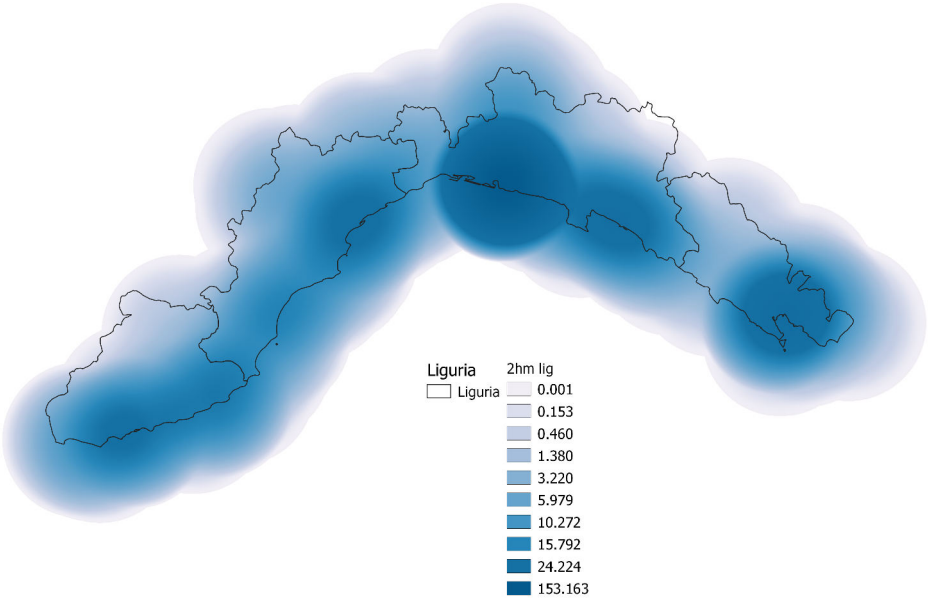
Own elaboration.

Figure 10: Piemonte, heatmap.



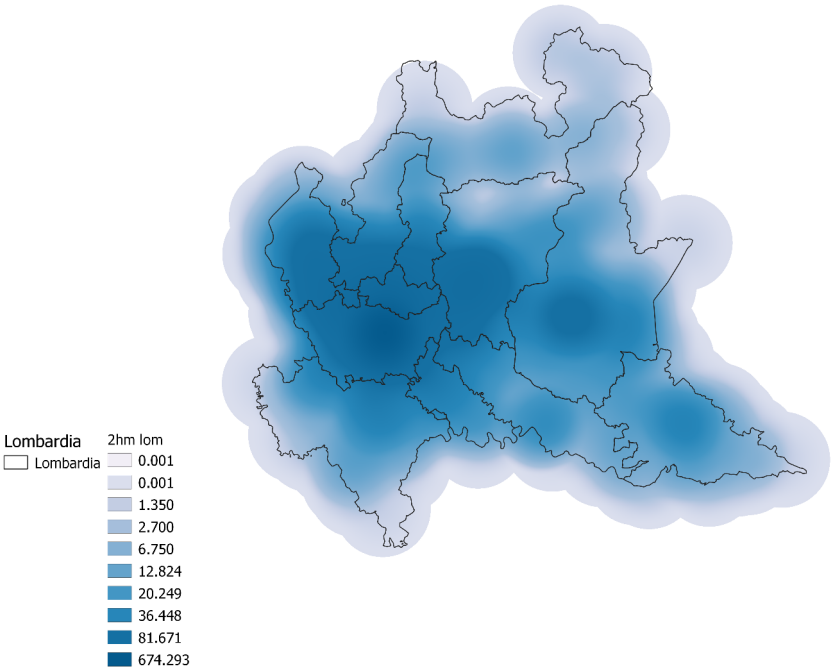
Own elaboration.

Figure 11: Liguria, heatmap.



Own elaboration.

Figure 12: Lombardia, heatmap.

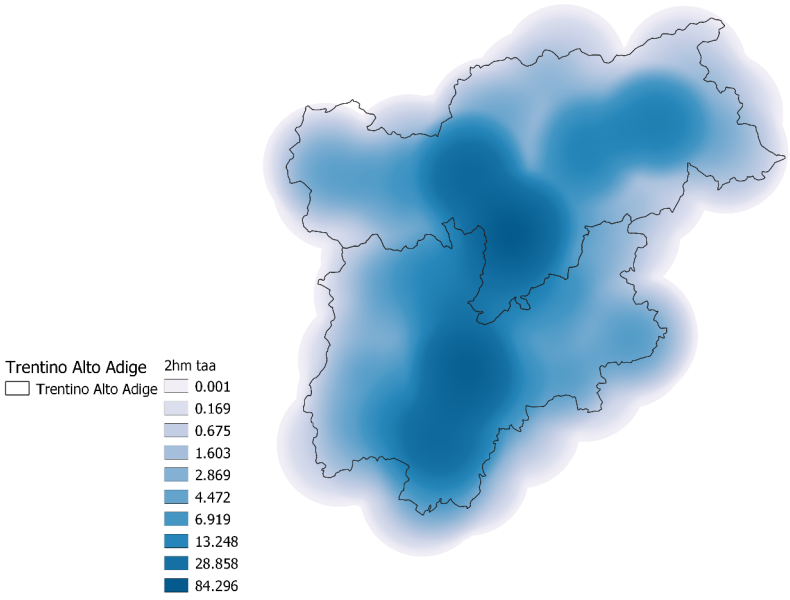


Own elaboration.

It is interesting to analyze the provision of the early child care service in these regions, in particular we observe as in Valle D'Aosta (Figure 9) and Piemonte (Figure 10) there are areas that are more than 20 kilometers away from the closest facility; the reason of this absence in the early child care system, may due to the territorial morphology of these regions that account for wide areas of mountains that are scarcely inhabited. In Liguria (Figure 11) and Lombardia (Figure 12) instead, almost the totality of the regional surfaces are covered by the relative heatmap, even if with different intensity.

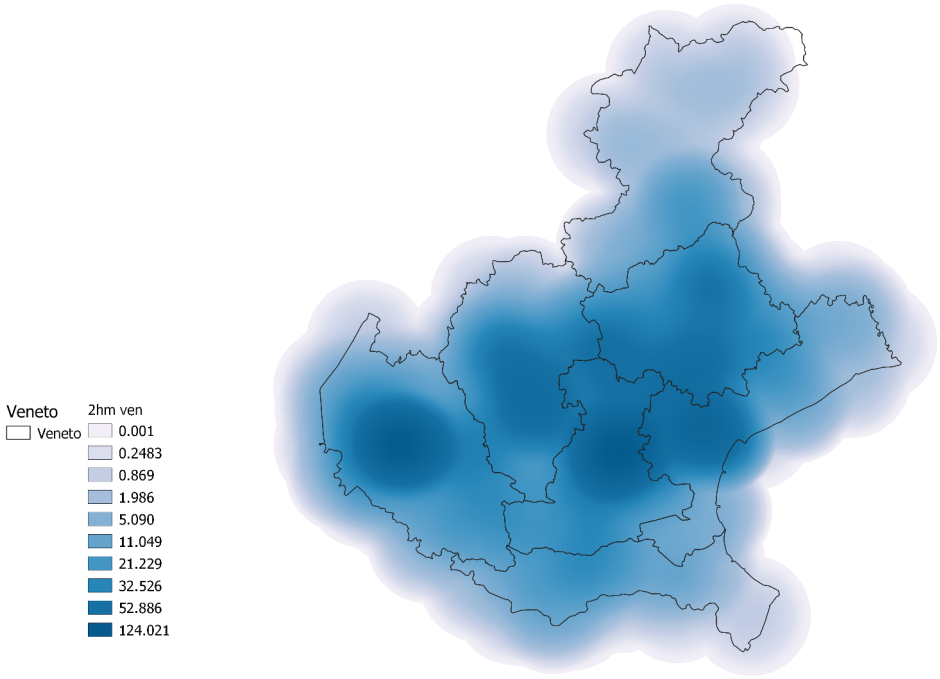
We consider now the North-East part of the country: Trentino-Alto-Adige, Veneto, Friuli-Venezia-Giulia and Emilia-Romagna.

Figure 13: Trentino Alto Adige, heatmap.



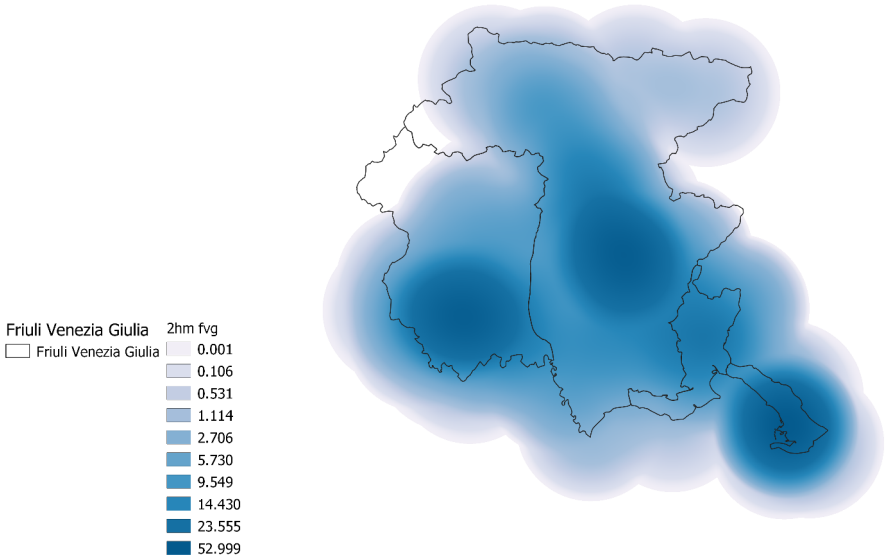
Own elaboration.

Figure 14: Veneto, heatmap.



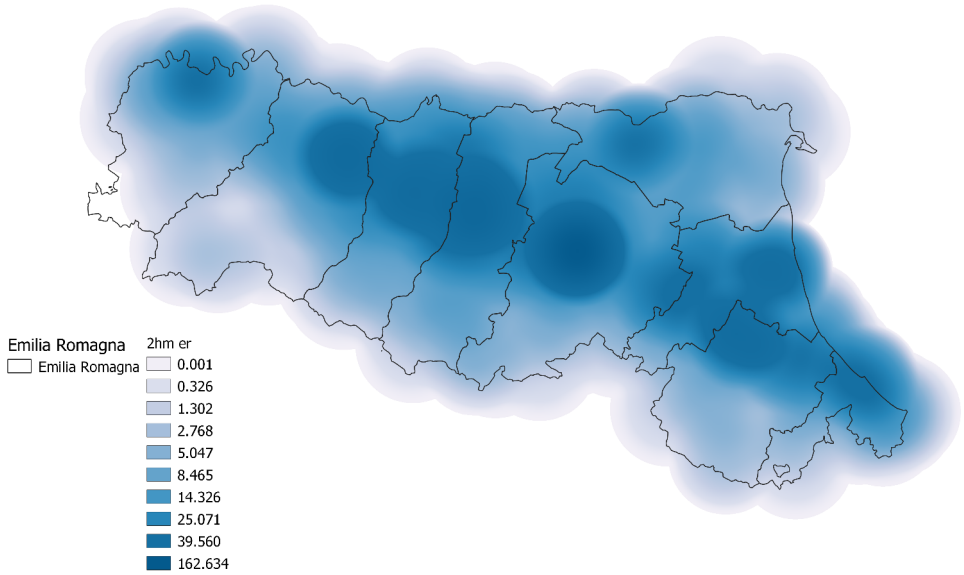
Own elaboration.

Figure 15: Friuli Venezia Giulia, heatmap.



Own elaboration.

Figure 16: Emilia Romagna, heatmap.

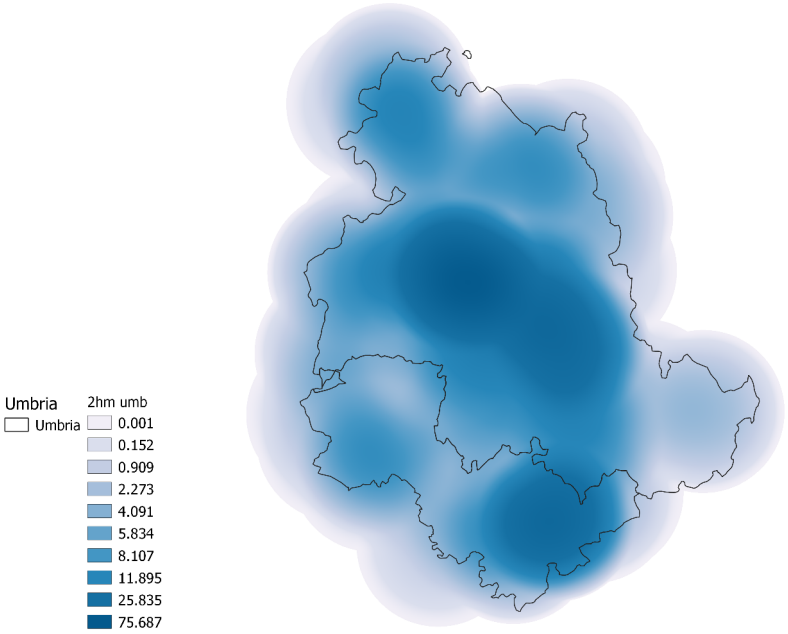


Own elaboration.

Looking at the Emilia Romagna (Figure 16), a region with an elevated child coverage ratio, we can see that the distribution of the early child care facilities results to be more concentrated in central part of the region, where the biggest and most populated cities of the region are located, while the accessibility level is lower in the areas close to the Southern border. In the Southern part we find in fact a mountain chain (Appennini), with a urban pattern characterized mainly by small, scarcely inhabited villages. Usually these villages are characterized by small highly populated centers, where it resides the wide part of the population of these municipalities.

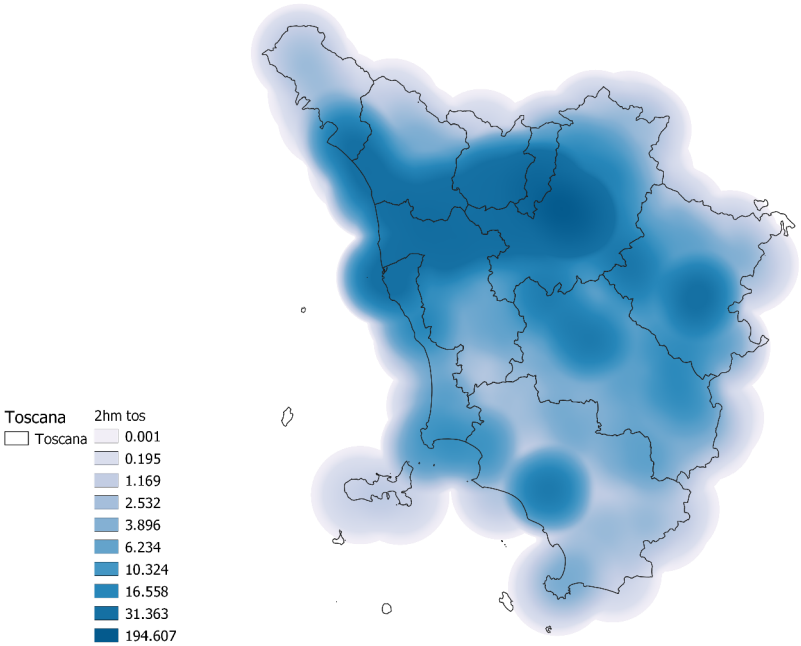
The regions of the central part of Italy, are: Umbria, Toscana, Marche and Lazio

Figure 17: Umbria, heatmap.



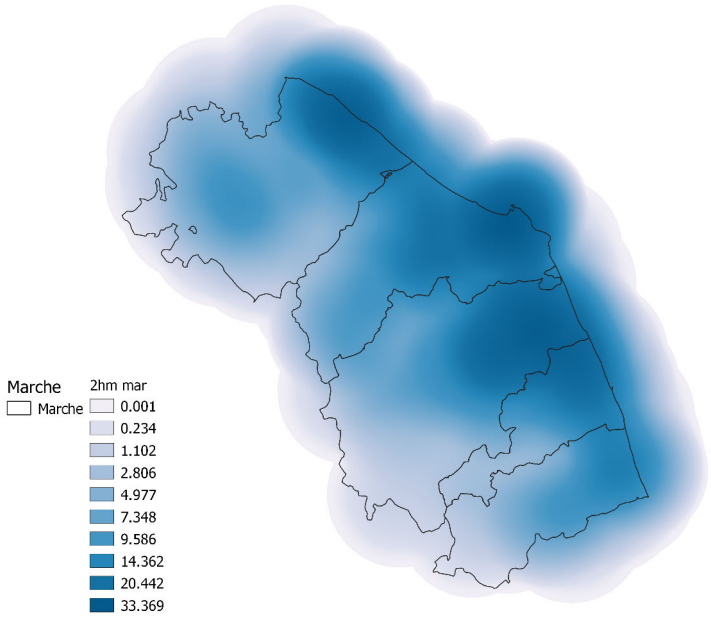
Own elaboration.

Figure 18: Toscana, heatmap.



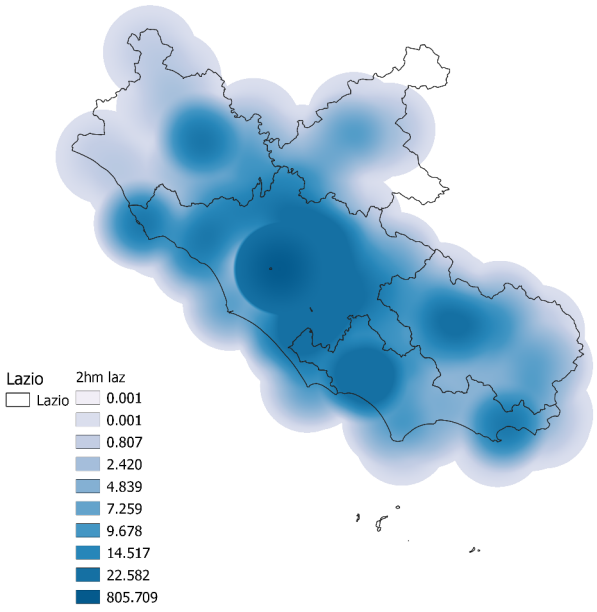
Own elaboration.

Figure 19: Marche, heatmap.



Own elaboration.

Figure 20: Lazio, heatmap.

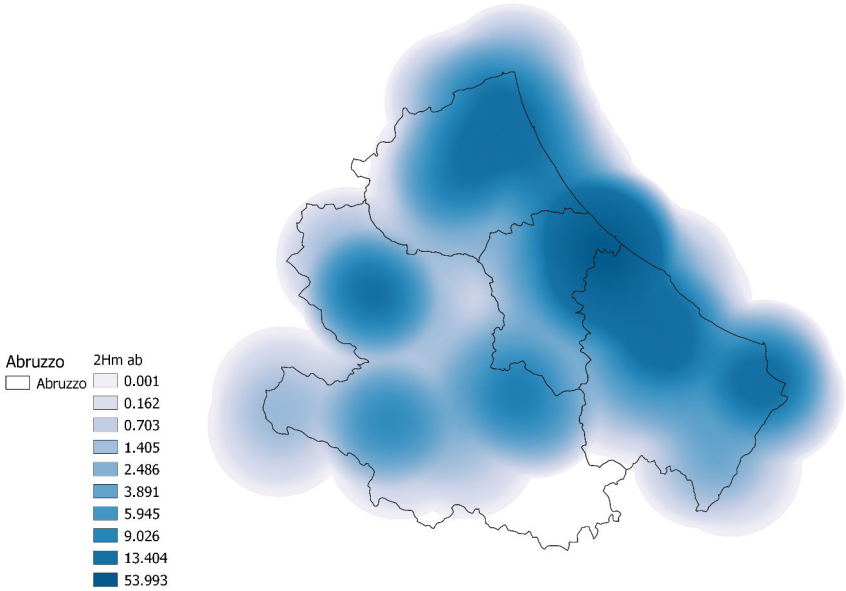


Own elaboration.

Umbria (Figure 17) and Toscana (Figure 18) are the regions that, in 2013, registered the highest values for the child coverage ratio. The relative heatmaps describe instead a totally different regional distribution of the facilities; Umbria appeared to have a more homogeneous density distribution, while Toscana a more heterogeneous one. Analyzing the deciles we observe an interesting situation, while the maximum values are different, due also to the presence in Toscana of high populated cities with an higher number of facilities, the values of the other deciles are very similar indicating a less dispersed supply in Umbria than in Toscana. This pattern is accentuated in Marche region (Figure 19) where we notice an even less dispersed supply.

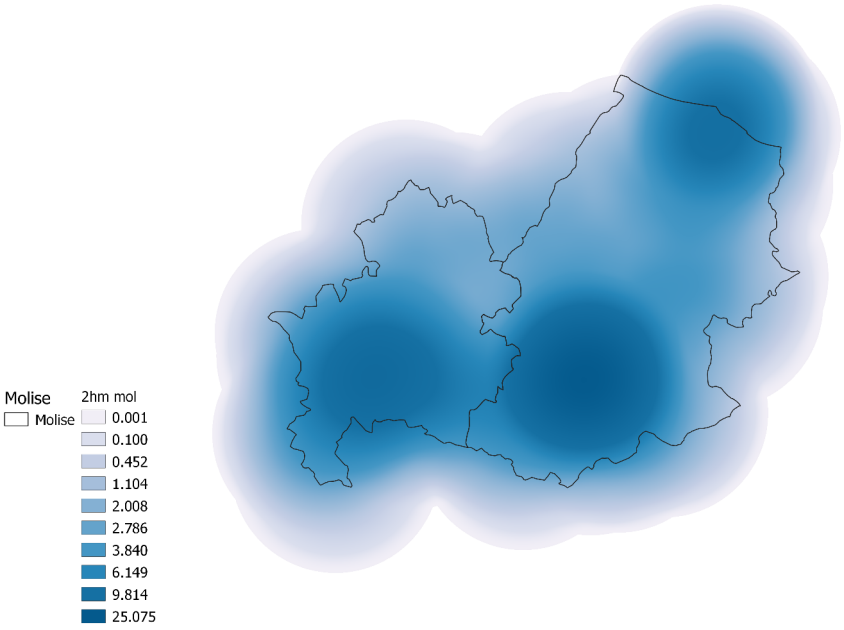
The regions belonging to the South macro area are: Abruzzo, Molise, Campania, Puglia, Basilicata and Calabria.

Figure 21: Abruzzo, heatmap.



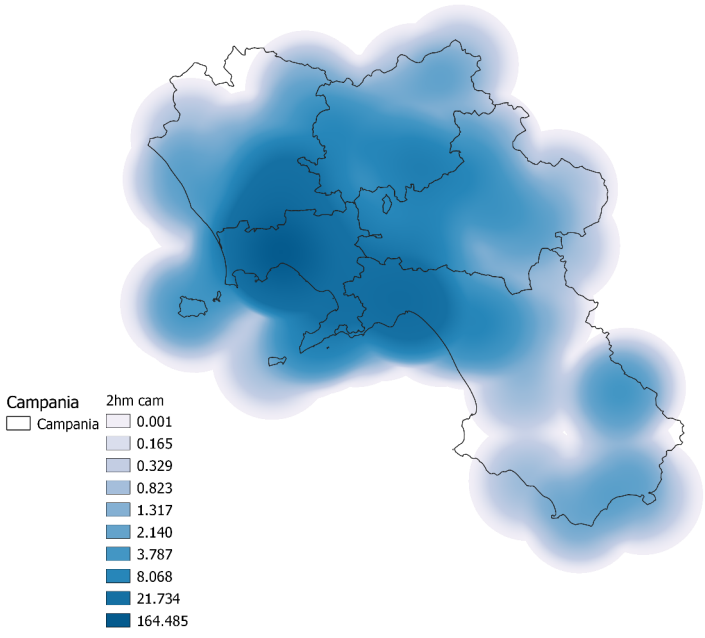
Own elaboration.

Figure 22: Molise, heatmap.



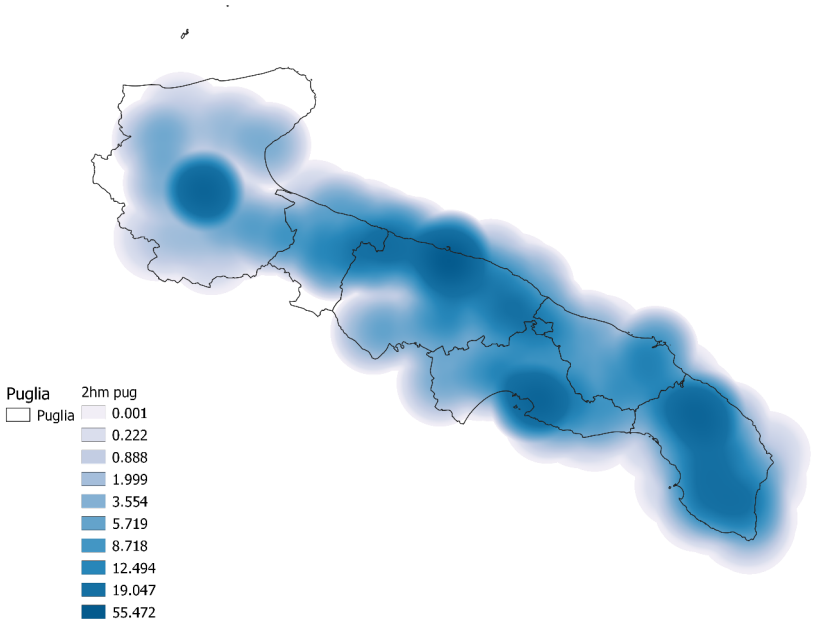
Own elaboration.

Figure 23: Campania, heatmap.



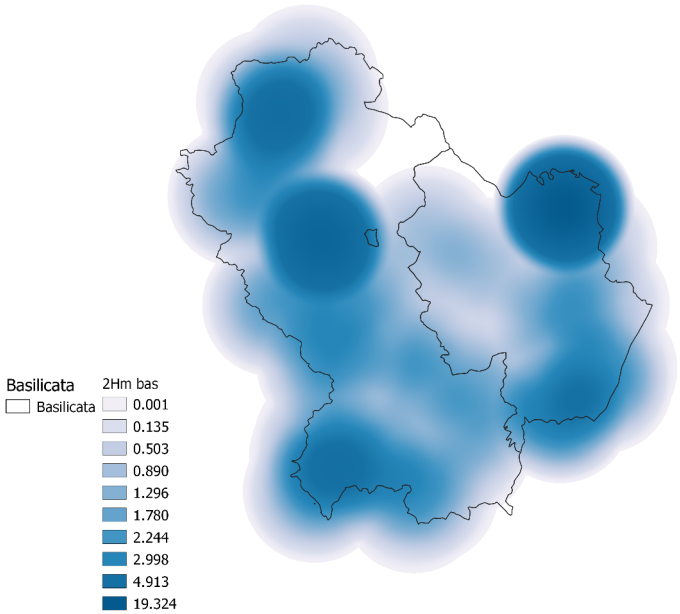
Own elaboration.

Figure 24: Puglia, heatmap.



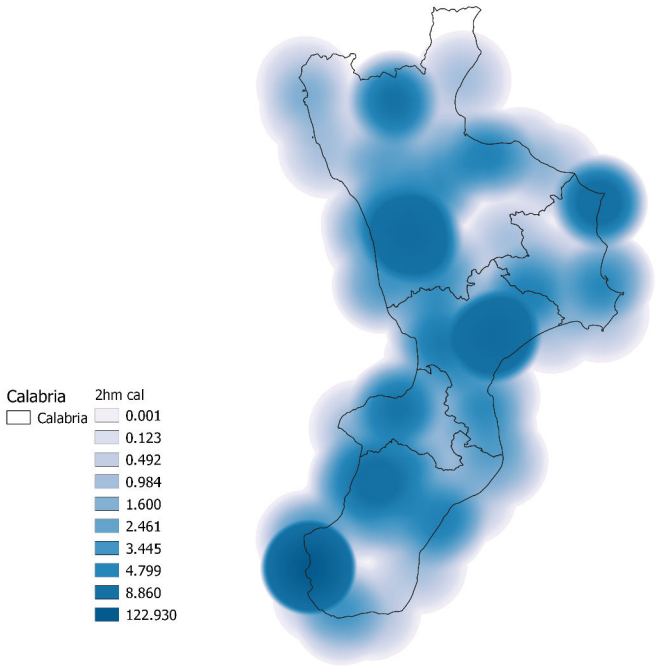
Own elaboration.

Figure 25: Basilicata, heatmap.



Own elaboration.

Figure 26: Calabria, heatmap.

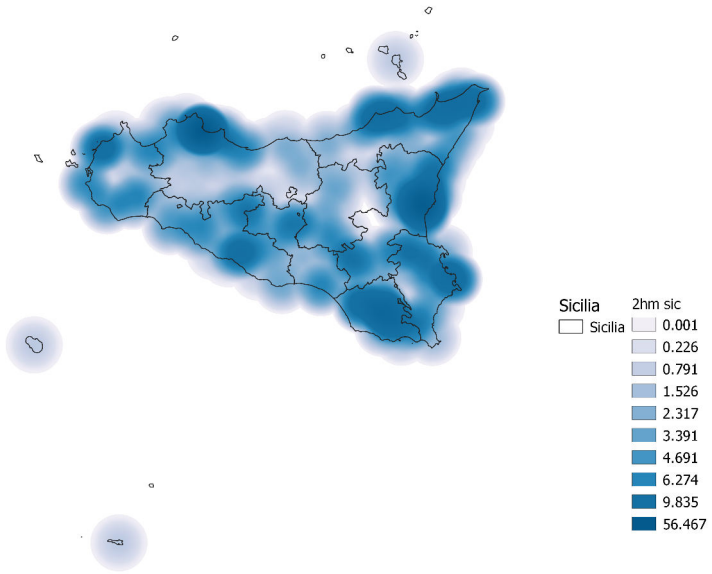


Own elaboration.

In the South of Italy, as for the Islands macro area, the regional heatmaps describe a completely different situation if compared with the regions of the previous macro areas. The values of the deciles, even with high maximum values related to high populated cities as Napoli in Campania (Figure 23), are particularly low indicating a wide gap between the urban areas and the rural areas. Entire provinces have very low average levels of supply, in this case the provision is concentrated only in the most populated cities leaving the remaining part of the province with few facilities. In Calabria (Figure 26) the ninth decile has a value less than $\frac{1}{10}$ of the tenth decile, describing an high intra-regional variation.

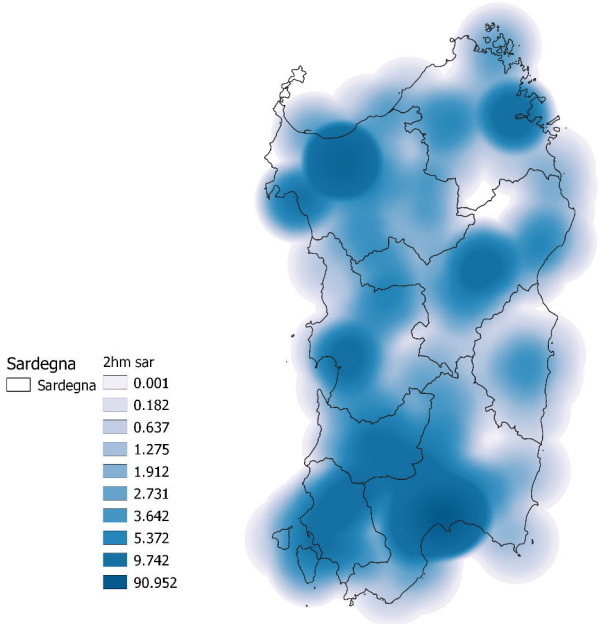
The last regions, Sicilia and Sardegna are part of the Islands macro area.

Figure 27: Sicilia, heatmap.



Own elaboration.

Figure 28: Sardegna, heatmap.



Own elaboration.

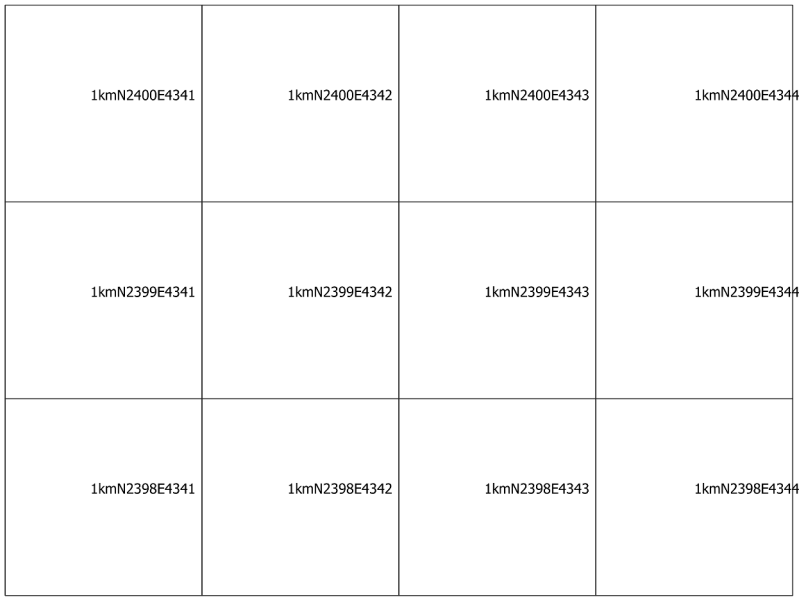
These regions share with the ones in South the strong intra-regional variation in the provision of the service and the concentration of the population on the coasts where we find the highest number of facilities.

3.1.4 The spatial distribution of the population

After having graphically displayed the early child care provision for the Italian regions, we are going to introduce the distribution of the population to define the final spatial accessibility value considering the effective location of the families. Being aware that it was impossible to know the composition of each family of the country and its exact spatial location, for privacy reasons, we decided to use the available information from the dataset of the population census of 2011. The most precise data that was possible to use from the ISTAT population census, was a regular grid of all the square kilometers of the Italian territory, with specific attributes from the census dataset, meaning a regular grid of squares of dimension 1KM*1KM covering all the peninsula.

In Figure 29 we can observe 12 of the total 310,980 squares that composed the final outcome. Each cell has a list of attributes assigned to it specifically linked to the spatial collocation of the cell: the length of the square and the surface of the cell, that are values common to all the cells (4,000 meters for the perimeter and $1KM^2$ for the area), the cell ID to distinguish them by a geographic criterion (e.g. 1kmN2399E4341) and the information on the total number of inhabitants, the population, that lived in the square kilometer considered. This last value turned to be extremely important in the definition of the potential demand of the early child care system. The uninhabited areas of the country were then immediately excluded in the definition of the population distribution, logically in an efficient situation we do not want to waste resources to give accessibility to the early child care service to places in which no one lives; trying to better characterize

Figure 29: Regular grid from the Italian population census of 2011.

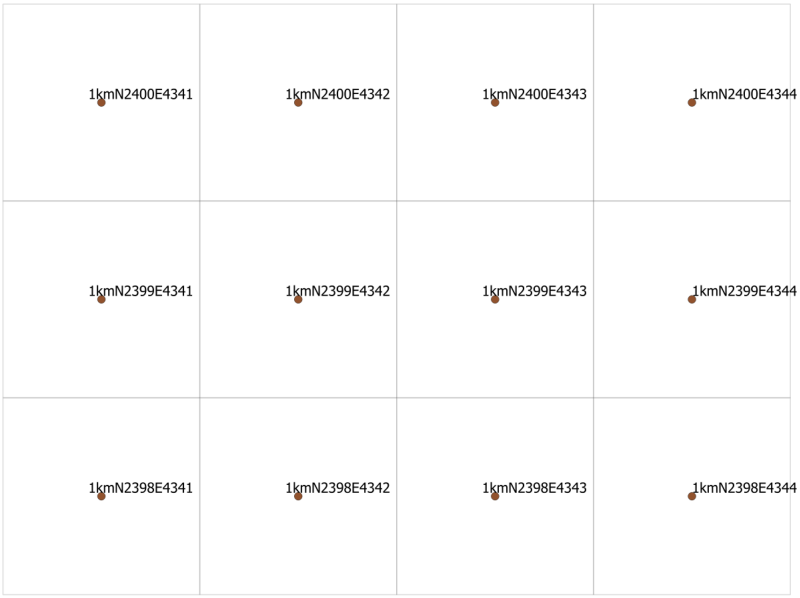


Own elaboration from ISTAT data.

these excluded cells, we know that a great amount of them are situated in the Alps and Appennini mountain chains, the two longest and highest mountain chains in Italy. Reducing the total number of cells have helped us to not consider areas like glaciers, rocks, swamps, lakes, parks and all those places, in particular in the South of the country, that suffer a strong depopulation phenomenon. The total number of cells from 310,980 reduces to 172,216, the final set accounts only for the 55.38% of the initial one; this strong reduction suggests that an analysis with the complete set would have been incorrect because almost the half of the territory we would have considered, would have not been of interest for the research. As stated in the literature, a common solution to compute the accessibility analysis is the point-to-point procedure, meaning to reduce the population distribution to points, each of them representative of an area for which we know the

total number of inhabitants. In our research this means to calculate for every cell the relative centroid and use them as points representative of the cells they belong to. This procedure is done also to ease the comprehension of the final result and to facilitate the spatial analysis. Note that we are not reducing entire heterogeneous surfaces to single points (as postcodes areas or municipalities with not regular boundaries), but squares with a fixed and common side of 1KM; what we expect is to have more points for each Italian municipality and to be able in this way to reproduce the differences in the accessibility level of the early child care system inside the polygon representing the administrative unit.

Figure 30: Centroids of the regular grid.



Own elaboration.

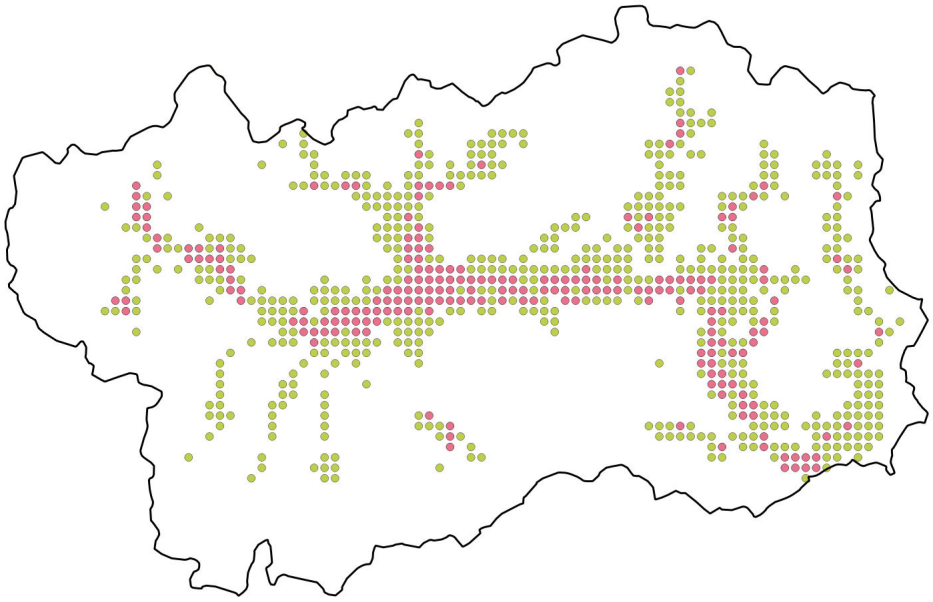
In Figure 30 we reproduce the regular grid with the centroids, one for each cell. QGIS (our GIS application) computes this operation automatically and assigns to each point the list of attributes of the

corresponding cell without losing the information. The transformation has to be considered only as a graphic elaboration of the initial data, in fact the list of attributes remained the same with also the information of the length and area of the cell of each corresponding centroid. With these two steps we obtained the graphic representation, a map, of the distribution of the Italian population. After having precisely identified the set of points with positive population, being interested not only in the number of people living in a certain area but also in knowing to which administrative unit a specific point belongs, we used a specific geoprocessing tool of QGIS to assign to each centroid the list of attributes of the municipality in which was located. The result was a new vector with an union of the list of attributes of the centroid and of the political boundaries by their relative position in the map. Summarising, QGIS counted for every municipality the number of centroids belonging to its surface and assigned to them the same list of attributes of the municipality; for example, the centroids that belong to Milan municipality have now as list of attributes the list of attributes of the centroids (length of the cell, area of the cell, ID number and population) and the attributes of the administrative unit: an ID representing the macro area, an ID that identifies the region (Lombardia), an ID for the province (Milan), the name of the municipality (Milan), the perimeter of the boundary of the municipality, the area of the municipality and other information. Doing so we move from a geographic identification, computed by assigning a geographic code, to an administrative specification for which we are able to distinguish the centroids by administrative means. In the definition of distribution of the population, it is crucial to remember that we do not know the number of children between 0-2 years old of age that belong to each centroid, so that it is impossible to further exclude all those points with no children in that age. Even if this set of points could be considered representative of the distribution of the

Italian population, because it is composed by all the square kilometers in which at least a person lives, we note that there may be centroids with a very low population that should be excluded by the final set because they are unlikely to represent points in which families with children between 0-2 years old of age live. While it appeared to be obvious to eliminate all the centroids with a population equal one or two, because it is difficult to imagine an infant outside a familiar context, it has been more difficult to decide the level of population threshold to use as criterion for this further reduction. The value we decided to use as threshold was set to a hundred inhabitants for square kilometer. Even if arbitrary, this level was thought to be both high enough to exclude the centroids scarcely populated, and not too high to exclude the rural areas with a lower population density. A value of 100 inhabitants for square kilometer was also well below the average population density of Italy that is close to the value of 200 inhabitants per square kilometer (ISTAT, 2017). This reduction has been done also considering the scarce availability of the early child care system, having a very low coverage ratio we expect to observe the service to be located primarily in the highly populated areas. We are anyway aware of the importance that a research on the scarcity of spatial accessibility of the early child care system for the less populated areas would have, but for the aims of our research we preferred to propose the general idea of the model to demonstrate the validity of the spatial approach. The final set of centroids, the points that compose the demand side of our model, accounts for 57,839 observations, the 18.60% of the initial grid of the Italian territory and the 33.59% of the set of the centroids with a positive population.

In Figure 31 we can graphically observe the result obtained for the Valle D'Aosta region. The green dots represent the centroids with a positive value of population, while the pink dots the centroids with a population of at least 100 people. Firstly we highlight how this

Figure 31: Valle D'Aosta region, centroids.



Own elaboration.

threshold has excluded all the points that resulted to be isolated from the others and for this reason with a small likelihood to be the places in which infants live, secondly that even if we may have lost a part of the population (if we consider the bottom left part of the map), we have conserved the municipalities with the highest number of inhabitants that are located in the central part of the map.

The final distribution of the Italian population adopted in the model, represented by the centroids in which we assume that the majority of the families with children between 0-2 years old of age live, is then defined as the set of square kilometers of the Italian country (represented by their centroids) with a population of at least one hundred inhabitants in 2011. This formulation allow us to create heterogeneity inside the municipalities through the exclusion of the uninhabited and less populated areas. Differently from the association of an en-

tire municipality to a single point, its centroid, with our procedure we can analyze the variations in the accessibility level of the early child care service not only between the administrative units but also within them.

3.2 The regional spatial accessibility value

After having described the regional supply of the early child care system, we introduce, for each region, the vector of points representing the centroids with more than a hundred inhabitants and we calculate the accessibility value for each of them. To calculate this variable, we used the regional heatmaps described previously and we assigned to each centroid the density value corresponding to its geographical location, with respect to the spatial diffusion of the early child care service. After having calculated the accessibility value of each centroid, we could either calculate the average value of the variable at a provincial level or at a regional level; for the use that we are going to do in our research we are going to consider the regional value. The accessibility value represents the average number of facilities that influences a family located in the considered centroid, weighted by the distance that each of the facility has from it. High values of the variable corresponds, for a family living in a centroid with more than a hundred inhabitants, to an high potential provision of the early child care service. Thanks to this variable, we want to investigate if a widespread diffusion of the service has a positive effect on the familiar choices. Generally a family does not know what is the child coverage ratio in the area in which it lives, but it considers firstly the number of facilities that can be easily reached from the place in which it lives; if the closest facilities are located too far from the house or workplace of the parents, the family would behave consequently and the mother may decide to exit from the labour market when it is not possible to use the informal system.

Table 6: Regional level of child care coverage ratio and accessibility value. The coefficient of variation and the GINI index describe the intra-regional variation in the accessibility variable, considering the sampled centroids. Bolzano and Trento, the provinces of Trentino Alto Adige have, been considered separately due to the special legislative powers they have.

Region	Coverage ratio	Accessibility value	Coef. of variation	GINI
Valle D'Aosta	31.5%	16.61	0.6257	0.3465
Piemonte	25.1%	36.39	1.4208	0.5579
Liguria	29%	32.32	1.0563	0.4744
Lombardia	28.1%	99.68	1.2523	0.586
Bolzano	17.2%	25.60	0.9656	0.517
Trento	30.1%	27.94	0.756	0.4254
Veneto	24.2%	43.98	0.6231	0.3471
Friuli Ven. Giul.	25.1%	22.03	0.6189	0.3486
Emilia Romagna	35.2%	36.75	0.8056	0.4126
Toscana	32%	39.66	1.0455	0.5087
Umbria	37%	25.23	0.8162	0.4454
Marche	23.7%	16.44	0.4964	0.2855
Lazio	27.9%	74.29	2.1922	0.7596
Abruzzo	19.4%	14.45	0.8521	0.4037
Molise	19.8%	10.40	0.6581	0.3665
Campania	6.2%	24.81	1.4546	0.6552
Puglia	12.1%	17.07	0.7154	0.3875
Basilicata	12.9%	4.38	0.9069	0.45
Calabria	10.6%	12.08	1.7153	0.6517
Sicilia	11.1%	13.29	1.017	0.5059
Sardegna	26.8%	14.89	1.4141	0.6309

Own elaboration.

In Table 6 we list the regions and the corresponding values for the child coverage ratio, the spatial accessibility variable and the coefficient of variation and the GINI index calculated from the accessibility values of the centroids. The two indexes of dispersion would then describe the intra-regional variations in the provision of the service that we can not appreciate from the spatial accessibility variable when aggregated at the regional level; to calculate these indexes we considered the accessibility values of all the centroids belonging to every region. Table 6 describes the two dimensions that are related to the early child care service, availability and accessibility at regional level. The two indexes of dispersion better specify the distribution of the service

and analyze the inequality level in its provision. Consider for example two regions, Piemonte and Friuli Venezia Giulia, they both have a child coverage ratio of 25.1% but different average accessibility values. Piemonte results to be on average more accessible than Friuli Venezia Giulia even if in this region the distribution of the facilities is more equal. The reason of these differences mainly regards the distribution of the population in the two regions and the presence of cities with an high number of inhabitants. While the child coverage ratio indicates the total number of spots available, the spatial accessibility variable describes if these spots are effectively accessible, meaning if the location of the structures is optimal considering the distribution of the population.

Table 7: Italian regions listed by population on 2011 with relative child coverage ratio and accessibility value referring to 2013. Bolzano and Trento, the provinces of Trentino Alto Adige have, been considered separately due to the special legislative powers they have.

Region	Coverage ratio	Accessibility value	Population
Lombardia	28.1%	99.68	9,704,151
Campania	6.2%	24.81	5,766,810
Lazio	27.9%	74.29	5,502,886
Sicilia	11.1%	13.29	5,002,904
Veneto	24.2%	43.98	4,857,210
...	...	...	...
Basilicata	12.9%	4.38	578,036
Trento	30.1%	27.94	529,457
Bolzano	17.2%	25.60	507,657
Molise	19.8%	10.40	313,660
Valle D'Aosta	31.5%	16.61	126,806

Own elaboration.

Table 7 lists instead the most and least populated regions, we observe as the accessibility variable seems not to be influenced by the total population of a region; analyzing Basilicata and Bolzano we note for example that even if they have similar population and child coverage ratio, the accessibility value results to be extremely low for the Basilicata region and quite high for Bolzano. Neither it seems that

the spatial variable is directly linked to the child coverage ratio, looking at Campania, whose child coverage value is the lowest in Italy, we observe that this region has an accessibility value higher than the one of Sicily even if the child coverage ratio is lower.

The regional aggregation does not permit anyway to appreciate how the accessibility values vary according to the population of the municipality; to avoid this inconvenient we implemented an additional specification thanks to QGIS tools. For every municipality we calculated the total number of inhabitants starting from the information of the population census of 2011 and we divided the set of municipalities in four bands according to the size of the population of the municipalities. To calculate the population of each administrative unit we used all the centroids with positive population, summing the value of every centroid belonging to the same municipality we were able to obtain the total number of inhabitants of every municipality. Then we adopted the same previous threshold of a hundred inhabitants to exclude the centroid less populated and for each band we calculated the aggregate accessibility value. The bands used to divide the municipalities by the size of the population are four: first band, population less or equal than 2,000 inhabitants; second band, population greater than 2,000 inhabitants and less or equal than 10,000 inhabitants; third band, population greater than 10,000 and less or equal than 50,000 inhabitants; fourth band, population greater than 50,000 inhabitants. The division we adopted followed the one used in ISTAT and it would be also used in the econometric specification of the research. Except for the last band, whose interval has open upper bound, this characterization permits to compare regions more rigorously than in the previous tables.

In Table 8 we list for every band the number of observations (the centroids of the population used in the calculation) and the regional accessibility value of the band. It is important to remember that in the

Table 8: Accessibility value by region and number inhabitants of the municipality

Region	Regional		0-2,000 inhabitants		2,001-10,000 inhabitants		10,001-50,000 inhabitants		more than 50,000 inhabitants	
	Value	Observations	Value	Observations	Value	Observations	Value	Observations	Value	Observations
Valle D'Aosta	16.61	207	13.89	107	17.22	86	33.63	14	0	0
Piemonte	36.39	4875	16.45	1665	28.21	1875	56.07	921	110.05	413
Liguria	32.32	1280	15.16	282	27.74	565	19.34	218	80.01	215
Lombardia	99.68	7577	29.93	1410	73.80	3668	147.47	1864	263.70	635
Bolzano	25.60	727	23.57	146	18.87	461	45.18	91	81.31	29
Trento	27.94	781	18.67	358	26.58	281	40.93	76	69.08	66
Veneto	43.98	6941	17.83	427	36.62	3300	48.82	2732	90.21	482
Friuli Ven. Giul.	22.03	1706	10.95	267	21.08	947	23.30	356	47.07	136
Emilia Romagna	36.75	4809	5.69	150	22.23	1477	39.65	1597	56.00	1318
Toscana	39.66	3924	5.61	226	17.97	1210	48.51	1670	63.06	818
Umbria	25.23	1372	7.19	113	19.90	394	18.85	550	49.49	315
Marche	16.44	1973	8.84	268	14.27	792	19.31	680	24.21	233
Lazio	74.29	4638	7.02	403	11.99	1221	23.88	1701	218.19	1313
Abruzzo	14.45	1980	6.81	501	13.36	776	19.67	470	23.93	233
Molise	10.40	488	7.73	220	10.62	171	16.07	97	0	0
Campania	24.81	3870	4.97	502	9.67	1547	32.27	1280	68.88	541
Puglia	17.07	2045	5.53	68	14.44	524	15.21	1061	27.59	392
Basilicata	4.38	697	1.88	147	3.35	337	4.12	128	13.03	85
Calabria	12.08	2499	7.14	527	7.52	1201	8.78	470	44.06	301
Sicilia	13.29	3151	5.62	193	8.82	1023	13.19	1159	21.24	776
Sardegna	14.89	1534	4.17	428	8.99	582	19.22	302	45.12	222

Own elaboration.

case of two municipalities, one with more than 50,000 inhabitants and one with less than 2,000 inhabitants situated one next to the other, the smallest can be influenced by the presence of the facilities in the biggest one when the distance between them is less than 20 kilometers; so even if we consider the groups separately, the supply side remained built starting from the kernel density estimation at a regional level.

Table 8 describes the intra-regional variation of the provision of the early child care system; if we consider two regions, Valle D'Aosta and Toscana, we observe an higher regional value for the latter, but when considering the first band, it is the former the region with the highest accessibility value indicating that, on average, families that live in small municipality have a better access to the early child care system in Valle D'Aosta than in Toscana. The use of the bands allow also for a better specification of the effect of the distribution of the early child care system for the econometric analysis, this specification considers differently the families that live in low populated and high populated municipalities.

3.3 Robustness check of the spatial accessibility variable

The main issue concerning the definition of the supply side of the model, was related to the impossibility to locate the early child care facilities precisely on the map due to the lack of the relative addresses. For the reference year, the 2013, this information was not available in the database ISTAT due to the privacy legislation, and even looking at the websites of the Italian regions, it appeared not possible to acquire the addresses of the facilities for all the Italian regions. The measurement error we introduced in the model through the approximation, certainly affected the final result of the accessibility variable but in a way that it was not possible to measure with the data of 2013 in our possess. Searching for this information, we found in the websites of four regions: Lombardia, Friuli Venezia Giulia, Veneto and Valle

D'Aosta, the regional lists (for different years) of the early child care facilities with the addresses indicating the precise location. To check the validity of the approximation made in the model of our research, the location of the facilities on the centroid position, we used QGIS to compare the heatmaps generated by the distribution of the service in these four regions in two different situations. In the first case we estimated the accessibility value using the addresses of the facilities to obtain the real provision, in the second case we applied the same approximation used in the model, positioning the structures on the centroid location of the municipality they belonged to; we then calculated the regional accessibility value and we compared the results obtained to measure the error we introduced with the approximation.

The regional datasets used are relative to the year 2019 for Lombardia and Valle D'Aosta, 2018 for Friuli Venezia Giulia and 2017 for Veneto. Even if the data belong to different years, due to the regional construction of the heatmap and the type of analysis, the temporal discrepancy does not affect the result of the robustness check that aims only to measure the magnitude of the error that is created when instead of considering the exact location of the early child care facilities, we use an approximation. The population distribution would be represented by the same set of points used in the main model of the research, the centroids derived from the population census of 2011 (ISTAT, 2019b). It is anyway impossible to compare those data with the one of the principal model; not only the reference year is different, but also the sources are different, if in the main model the regional accessibility value has been built starting from ISTAT data, in this case we acquire the information from the databases of the regions.

The results obtained for the accessibility variable and the two indexes of dispersion, coefficient of variation and GINI, are reported in Table 9. The discrepancy in the measures is relatively small indicating that the approximation adopted in the main model can be considered

Table 9: Accessibility variable, robustness check

Region	With addresses			Without addresses		
	Accessibility value	Coefficient of variation	GINI	Accessibility value	Coefficient of variation	GINI
Lombardia	74.00	1.3618	0.5951	73.85	1.3826	0.5959
Friuli Ven. Giul.	23.13	0.6318	0.3587	23.30	0.6373	0.3617
Veneto	31.39	0.6780	0.3719	30.92	0.6883	0.3733
Valle D'Aosta	11.29	0.5828	0.3278	11.24	0.5900	0.3285

Own elaboration.

acceptable for the analysis. This small measurement error may due to the large search radius of the heatmap and to the presence of the child care structures mainly in the central part of the municipalities. A large search radius may in fact smooth the differences in the accessibility values in the smaller municipalities, moving the position of a facility of few meters has not a relevant impact for a family living close to it. In the bigger cities instead, in the case of the absence of the addresses the highest values of the variable were located in the center of the city leaving the peripheral areas with lower accessibility, but in the case in which we use the addresses, the accessibility to the early child care service results to be lower for the center of the city but a little higher for the peripheral areas because some of the total number of facilities are now closer to the external part of the city. The interaction between this variations leads to minimal differences.

Table 10: Variations in the values between the use of the precise location and the approximation, in percentage

Region	Accessibility value	Coefficient of variation	GINI
Lombardia	0.21%	−1.524%	−0.13%
Friuli Ven. Giul.	−0.74%	−0.871%	−0.854%
Veneto	1.50%	−1.522%	−0.372%
Valle D'Aosta	0.41%	−1.242%	−0.200%

Own elaboration.

As stated in Table 10 the use of the addresses in the model increases

the regional average accessibility value, except for Friuli Venezia Giulia, and decreases the inequality in the accessibility variable across the centroids representative of the demand side. These effects strengthen the empirical framework proposed in the research, and justifies the approximation used. Unluckily we were not been able to replicate this check also for the region with the lowest level of child coverage ratio and accessibility, due to the lack of information in the regional websites, so we cannot estimate the measurement error in regions in which the service is scarce.

3.4 Alternative methods

To conclude this chapter we are going to explain the reasons that led us to use the heatmap procedure in the construction of the spatial accessibility model, differently from other methods suggested in the literature.

In the literature we described two different spatial analysis for the early child care system computed using GIS. In both of them we find the use of the centroid approach to represent the population distribution and the early child care supply, and the use of the road network in the construction of the spatial accessibility variable. Centroids appear to be the ideal solution to overcome privacy issues, it is anyway important to stress that if the centroids are representative of the population or the facilities of wide polygons (as postcodes or entire municipalities) the measurement error that the researcher introduces in the model could be relevant and not marginal as in our case. Using the centroids solution for representing the families with children of an area, we assume that the total weight of each zone is concentrated in the centroid but this is not always true, in particular when the corresponding surface is large. Even if the operational limitation that researchers face using centroids is relevant, this approximation is sometimes the only possible when there is an absence of more precise data. Even in the case

of more detailed data, unless we do have the exact location of the observations, it is always needed to make approximations, using smaller administrative units, or regular grid of smaller size (e.g. 1KM*1KM squares).

After having mapped the data of the early child care provision and the population distribution, one of the most used approach to derive the spatial accessibility variable is to exploit the GIS features to represent in the map the road network system and use it to track the path that each family has to do to reach the surrounding facilities. Doing so it is also possible to set a spatial or temporal threshold to limit the set of facilities that can be reached by each family; for example we can impose that the families are interested in a facility only if the distance between the structure and the house is less than 15 kilometers using the road network, or that families would not want to spend more than 10 minutes to drive the child to a structure. The trade off using this approach is between the number of points to connect and the time used by GIS to calculate, for each point, the shortest route. More are the points to connect, more is the time spent by the system to calculate all the possible routes and then to choose which of them respect the threshold set by the researcher. A road network approach can be implemented when the points to connect are in a limited number, when we have the possibility to work with a very powerful machine and using GIS software more sophisticated than QGIS. While using the centroids and the road network to define the spatial relation between the observations, appears to be one of the easiest and fastest solution, it is not always the best one. This procedure, being difficult to implement with the QGIS application for its limited features, would also be partially inefficient because the families can choose not to follow the fastest route to reach the early child care facility. The use of the road network, even if it accounts for the limits of speed for the different type of road (information not always available), does

not consider the level of traffic that can be present during the day, in particular in the morning. The traffic could in fact induce a family to choose a longer route not to risk to waste a lot of time when choosing the fastest but most congested route. The road network method does not account also for the possibility to bring the child to the facility using a different transportation mean than the car; we are aware of the rarity of this situation, but we preferred not to exclude it. Finally, the fastest route can account for the use of highways which price can affect the final decision of a family, inducing the parents not to enroll the child due to an increase in the final cost.

4 The econometric analysis

In this last chapter we are going to introduce the econometric model of the research. Following the economic literature on the effects of the early child care system on the employment decision of the mothers, we firstly define a model similar to those reported in the literature, focusing on the role of the child coverage ratio, then we investigate the role of the spatial accessibility variable in the description of the familiar decision on the child care and the maternal labour force participation.

4.1 The data

For the development of the econometric model we analysed the survey called "Multiscopo-Aspetti della Vita Quotidiana", a survey conducted by ISTAT on year base. The survey gathers information on the habits, on the economic aspects, on the degree of satisfaction of the services provided by public and private organizations and on the quality of life of the Italian citizens. It is conducted on 25,000 families spread in 800 municipalities of different dimensions; each year a representative random subset is selected as sample for the survey. For the analysis conducted in our research, we focus on the survey of 2013, the reference year for which we have calculated the spatial accessibility variable, and on the families in the sample in which it was present at least one child between 0-2 years old of age. Even if the survey was conducted at a familiar level, the interview has been conducted asking to each member of the family the same questions, so the final dataset consisted in a number of rows equal to the total number of respondents. The database in 2013 accounted for 8,361 families and a total of 20,275 people interviewed. For the computational analysis we excluded all the families without a child of the target age and, after having excluded a small number of respondents due to some missing information, we defined a dataset of 439 families with at least one child between 0-2 years old of age. For each family we gathered the

relevant information in a row per mother to obtain a final dataset of 439 observations each representative of a different family.

The variables considered in the analysis recall those used in the literature, in particular on the characteristics of the mother and the family, controlling also for the macro area in which the family lives. For each family the survey gathered also the information on the region and on the type of municipality in which the family resides. The population bands adopted by ISTAT to identify the dimension of the municipalities are the same used in the previous chapter for the construction of the spatial accessibility value. The bands are four: 0-2,000 inhabitants; 2,001-10,000 inhabitants; 10,001-50,000 inhabitants; more than 50,000 inhabitants. Using this specification we can assign a spatial accessibility value to each family knowing the region in which it lives and also the characteristic of the total population of the municipality in which it resides. Exploiting this specification we are then able to study the impact of the spatial accessibility to the service considering a more precise location of the family than the information on the region. In Table 11 and Table 12 we define the variables and their descriptive statistics.

The final dataset consisted in a row per mother and not a row per child, meaning that there is a possibility for a family to have more than one child between 0-2 years old of age; in this case the dummy Attendance has value 1 when at least one of them attended the formal early child care system. Age is a variable describing the age of the mother, its value corresponds to the median value of the intervals that ISTAT used in the dataset of the survey to classify the age of an individual. The variable Couple has been built considering the differences that may exist between a family composed by only one parent and a family in which the parents are a couple, considering in the same way married and not married parents (dummy equal 1). For the variable defining the Economic condition of the family, we

Table 11: Definitions of the variables.

Variable	Description	Source
Work	Dummy equal 1 when the mother works	Multiscopo survey, ISTAT
Attendance	Dummy equal 1 when at least one child in the family is enrolled in the formal early child care system	Multiscopo survey, ISTAT
Age	Median values of the bands used by ISTAT to classify the age of an individual	Multiscopo survey, ISTAT
Other child	Dummy equal 1 when is present in the family at least another child older than 2 years old of age	Multiscopo survey, ISTAT
Education	Dummy equal 1 if the mother has a tertiary education	Multiscopo survey, ISTAT
Couple	Dummy equal 1 if the parents of the child are a couple	Multiscopo survey, ISTAT
Economic condition	Dummy equal 1 if the economic situation of the family is described as good by the parents	Multiscopo survey, ISTAT
NE	Dummy equal 1 if the family lives in the North-East	Multiscopo survey, ISTAT
NW	Dummy equal 1 if the family lives in the North-West	Multiscopo survey, ISTAT
C	Dummy equal 1 if the family lives in the Center	Multiscopo survey, ISTAT
Child Coverage	Child coverage ratio of the region in which the family lives	Datawarehouse ISTAT
Spatial Accessibility	Spatial accessibility value of the region in which the family lives according to the dimension of the municipality	Own elaboration from ISTAT data
Total Accessibility	Spatial accessibility value multiplied by the regional child coverage rate	Own elaboration from ISTAT data

Own elaboration.

Table 12: Descriptive statistics.

Variable	Mean	SD	Median	Min	Max
Work	0.5103	0.5005	1	0	1
Attendance	0.2118	0.4090	0	0	1
Age	33.81	5.3710	32	22	47
Other child	0.4784	0.5001	0	0	1
Education	0.2415	0.4285	0	0	1
Couple	0.9021	0.2976	1	0	1
Economic situation	0.5399	0.4990	1	0	1
NW	0.1617	0.3686	0	0	1
NE	0.3075	0.4620	0	0	1
C	0.139	0.3463	0	0	1
Child Coverage	0.226	0.0848	0.242	0.062	0.370
Spatial Accessibility	47.69	60.8703	23.93	1.88	263.7
Total Accessibility	12.1138	17.3991	5.4243	0.2425	74.0997

Own elaboration.

may have different answers at individual level between the mother, the father and the other components of the family. Because it is a variable that influences the choice on the care of the child at a familiar level, we decided to consider as representative of the family, the answer with the worst outcome between the answers of the parents. If the mother describes the economic condition of the family as good and the father as not satisfactory, or viceversa, the resulting value at familiar level is not satisfactory, dummy equal 0. The variable Total Accessibility has been built considering the product between the Spatial Accessibility and the Child coverage variable. The reason behind this choice relies on the fact that, considering together the two dimensions of the early child care supply, we can control for the case in which the child coverage is high but the facilities are located far from the location of the families (low value for the spatial accessibility), and for the case in which we have a good spatial location of the structures but the spots available are not sufficient for the population demanding this service. The Total Accessibility variable would then avoid to fall in these situations considering together the spatial accessibility and the availability in terms of ratio between quantity supplied and quantity demanded. The dummies for the macro areas are then used to control for the variations of the dependent variables due to the geographic position of family. The data for the child coverage ratio are derived from the ISTAT database (ISTAT, 2019a).

The regressions we are going to present have as dependent variables the employment situation of the mother, through the variable Work that is a dummy equal one when the mother works, and the dummy Attendance that describes if the child is enrolled in the formal early child care system. For all the regressions that we are going to treat in this chapter, we used robust standard errors to control for the heteroskedaticity.

4.2 Spatial accessibility and attendance to the service

From the final sample we derive in Table 13 the probability for a family to enroll the child in the early child care system according to the macro area in which the family lives. Even if slightly different from the child coverage ratio for the macro areas, we observe that, as we expected, the macro area with a substantial lack of facilities, South, it is also the one in which the number of children enrolled is the lowest in the sample.

Table 13: Attendance of the early child care system per macro area in the sample, 2013.

Macro area	Attendance
North West	25.4%
North East	22.2%
Center	21.3%
South	18.4%
Islands	19.4%
Italy	21.18%

Multiscopo survey, ISTAT.

It is important to remember that the attendance to the service may be driven also by the dimension of the municipality in which the family lives, information that we find in the Spatial Accessibility variable. The spatial variable consists in four values per region, each value represents the average spatial accessibility value by type of municipality, meaning the values found in Table 8. This specification helped us to create heterogeneity inside each region, increasing the precision of the variable in the econometric analysis.

We want to analyze whether the attendance to the early child care system depends on the spatial provision of the facilities, for this reason in Table 14 we represent Attendance as dependent variable and in the list of regressors our Spatial Accessibility variable with other variables describing the characteristics of the family.

Table 14: Attendance OLS, Spatial Accessibility. HC heteroskedasticity-consistent standard errors have been used in the regressions.

	Dependent variable:	
	Model I	**Model II**
	Attendance	
Age	0.009***	0.009***
	(0.003)	(0.003)
Other_child	−0.068*	−0.078**
	(0.039)	(0.038)
Education	0.136***	0.130***
	(0.050)	(0.049)
Couple	−0.026	−0.009
	(0.066)	(0.067)
Economic condition	−0.017	−0.017
	(0.039)	(0.039)
NE	−0.090	−0.155**
	(0.059)	(0.062)
NW	−0.129	−0.189**
	(0.084)	(0.085)
C	−0.150*	−0.203**
	(0.083)	(0.083)
Child coverage	0.858**	0.877***
	(0.337)	(0.330)
Spatial Accessibility	0.001**	
	(0.0004)	
log(Spatial Accessibility)		0.094***
		(0.022)
Constant	−0.236*	−0.478***
	(0.136)	(0.151)
Observations	439	439
R^2	0.074	0.096
Adjusted R^2	0.052	0.075
Residual Std. Error (df = 428)	0.398	0.393
F Statistic (df = 10; 428)	3.399***	4.542***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Own elaboration.

In both the specifications the Spatial Accessibility variable results to be statistically significant and with a positive effect with respect to Attendance, these results confirm the assumption that closer is a family to the early child care system, more probable is that the child of the family results to be enrolled in a facility. In Model II in particular, thanks to the logarithmic transformation, we can state that the effect on Attendance is greater for the area of the country that starts from low levels of Spatial Accessibility; in the areas where the provision is good the effect reduces. Focusing on the other regressors we note as the age of the mother has a positive effect on the dependent variable, we can explain it assuming that for young mothers it may be more likely to have a familiar support in the care of the infant; older is the mother, less probable is to have relatives in good health able to take care of the child, meaning that the family may be forced to enroll the child in the formal system. Interesting is the negative effect of Other child on Attendance; in a family with more than one child in which one of them is older than 2 years old of age, the probability to enroll the youngest child in the early child care system is less than in a family with only one child between 0-2 years old of age. An explanation may be given referring to the fact that, being a scarce service, the older child may have forced the mother to exit from the labour market before the birth of the second child and, due to the difficulties to re-enter in the labour market, to become an household. In this situation the newborn would be cared by the mother and it would not need the formal early child care system. Children born from mothers with a tertiary education have an higher probability to attend the early child care service probably because the mothers are more likely to work.
Living in a region with an high child coverage ratio increases the probability for a child to be enrolled in the formal system, the variable ranges between 0 and 1, so the coefficient represents the effect due to an increase equal 1 in the ratio; to better understand the impact of

the variable we can state that an increase of 0.1 in the level of child coverage, increases the likelihood to enroll the child in the early child care system of 8.58 percentage points in Model I and 8.77 percentage points in Model II. The two models investigate the spatial effect using different approaches, in Model I we use the value of Spatial Accessibility without transformations, in Model II we are interested to know if it may exists a logarithmic pattern. Both the specifications result to be statistically significant and positive despite the presence of the variable Child coverage in the list of regressors. In Model I an increase of a standard deviation in the Spatial Accessibility variable leads to an increase equal to 6.1 percentage points in the probability to enroll the child to an early child care structure, while in Model II the effect of a 1% increase in the Spatial Accessibility variable has as effect a 0.44% increase in the probability to enroll the child to the system. The importance of the Child coverage variable has been largely discussed in the literature and the coefficient in the two models corresponds to the expectations we had on it. This variable anyway, does not crowd out the effect of Spatial Accessibility, meaning that the Child coverage alone is not able to completely explain the effect that the early child care system has on the probability to enroll the child in the system. Not only is important the number of spots available in a region, but also the distribution of the facilities with respect to the population.
For this reason, it may be interesting for a researcher to consider the product between these variables to obtain a new variable, that we call Total Accessibility, able to combine the information on the number of spots and on the spatial accessibility of the service.

$$\text{Total Accessibility} = \text{Child coverage} * \text{Spatial Accessibility}$$

Using this variable we are able to differentiate cases in which the Spatial Accessibility is similar but the Child coverage different. Referring to Table 8, Molise and Umbria have an accessibility value for

municipalities with less than 2,000 inhabitants that is similar: Molise 7.73 and Umbria 7.19. If we consider instead the child coverage ratio of these two regions the situation changes because Umbria has a level of the 37% while Molise 19.8%. Only a combination of the two variables is able to correctly describe the real availability of the early child care system, that is the reason why we used both the regressors in Table 14; the use of the Total Accessibility variable permits us to observe how the two variables interact in defining the Attendance to the early child care service.

Table 15 replicates Model I and Model II using Total Accessibility and its logarithmic value instead of Child coverage and Spatial Accessibility. The regressors used to control for the characteristics of the family maintain the sign and the magnitude of the Table 14; Age of the mother has a positive effect on the probability for the child to attend the early child care system, as having a mother with a tertiary education, while having another child in the family older than 2 years old of age keeps maintaining a negative effect on Attendance.
The Total Accessibility variable results to be statistically significant with a positive coefficient, this representation of the early child care system may then considered to be appropriated in analyzing the determinants that lead a family to enroll the child to the service. The peculiarity of this variable consists in the fact that positive variations in the value can be driven by an increase in the number of spots available in the facilities already present in a region, by an increase in the number of facilities or by a combination of these effects; using this variable we account simultaneously for the interactions between the spatial dimension and the characteristics of the quantity supplied. In Model III to an increase of a standard deviation in the level of Total Accessibility corresponds an increase of 5.2 percentage points in the likelihood for a family to enroll the child in the system, while in Model IV for a 1% increase of Total Accessibility, the probability to enroll

Table 15: Attendance OLS, Total Accessibility. HC heteroskedasticity-consistent standard errors have been used in the regressions.

	Dependent variable:	
	Model III	**Model IV**
	Attendance	
Age	0.009***	0.009***
	(0.004)	(0.003)
Other_child	−0.069*	−0.080**
	(0.039)	(0.038)
Education	0.134***	0.127**
	(0.051)	(0.049)
Couple	−0.027	−0.007
	(0.067)	(0.067)
Economic condition	−0.013	−0.015
	(0.040)	(0.039)
NE	0.003	−0.126**
	(0.047)	(0.056)
NW	−0.023	−0.162**
	(0.068)	(0.076)
C	−0.033	−0.163**
	(0.063)	(0.069)
Total Accessibility	0.003**	
	(0.001)	
log(Total Accessibility)		0.104***
		(0.022)
Constant	−0.107	−0.167
	(0.128)	(0.129)
Observations	439	439
R^2	0.061	0.094
Adjusted R^2	0.041	0.075
Residual Std. Error (df = 429)	0.401	0.393
F Statistic (df = 9; 429)	3.097***	4.967***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Own elaboration.

the child to the early child care system increases of 0.49%.
As for the Spatial Accessibility variable, also the Total Accessibility consists in four values per region, according to population dimension of the municipalities in which the families live. With more detailed data we would be able, hypothetically, to study the total accessibility value of any particular facility knowing the distribution of the population and the number of spots available.

4.3 Early child care system and maternal employment

The role of the early child care system in the literature refers mainly to the effect that it has on the employment decision of the mother; as demonstrated in the literature, more efficient is the formal system, higher is the probability for the mother to keep working during the first years of life of the child. Table 16 describes the ratio of the working mothers in the sample according to the macro area; in the Mezzogiorno regions (South and Island macroarea) we have the lowest percentage of working mothers, stressing the territorial differences that exist between different areas of Italy.

Table 16: Working mothers per macro area in the sample, 2013.

Macro area	Working mothers
North West	52.1%
North East	63.0%
Center	55.7%
South	40.0%
Islands	38.3%
Italy	51.03%

Multiscopo survey, ISTAT.

Table 17 presents the regression of Work, the employment situation of the mother of the child, on a set of regressors. In particular we used as regressors the Attendance variable to analyze if there exists a direct effect of having the child enrolled in the early child care system on the

Table 17: Work OLS. HC heteroskedasticity-consistent standard errors have been used in the regressions.

	Dependent variable:
	Work
Age	0.008^{*}
	(0.004)
Other_child	-0.077^{*}
	(0.045)
Education	0.216^{***}
	(0.052)
Couple	-0.138^{*}
	(0.076)
Economic condition	0.199^{***}
	(0.046)
NE	0.188^{***}
	(0.054)
NW	0.093
	(0.061)
C	0.133^{*}
	(0.070)
Attendance	0.167^{***}
	(0.052)
Constant	0.110
	(0.155)
Observations	439
R^2	0.179
Adjusted R^2	0.162
Residual Std. Error	0.458 (df = 429)
F Statistic	10.41^{***} (df = 9; 429)
Note:	*p<0.1; **p<0.05; ***p<0.01

Own elaboration.

employment situation of the mothers.
The age of the mother has a marginal positive impact on the dependent variable, having a tertiary education instead increases considerably the probability to work of 21.6 percentage points. Couple, the dummy variable equal 1 when the parents of the child are both present in the family, has a negative effect on the mothers' working condition decreasing the likelihood to work of 13.8 percentage points; this coefficient may represent the so called "breadwinner" pattern for which it is the father to be the member of the family designated to earn money for the whole family, in particular when there is a lack of job opportunities in the area in which the family lives. Also having a child older than 2 years old of age has a negative impact, while living in the North-East of Italy has a strong positive effect as living in the Center. The variable of interest, Attendance, is statistically significant and positive indicating the fundamental role that the early child care system plays in the maternal labour force participation. Attendance is the variable we assume to directly affect the maternal employment decision and its positive coefficient confirms that having a child enrolled in the early child care system increases the probability to work of 16.7 percentage points. We assume that the familiar effect of the early child care system is driven mainly by the Attendance variable being representative of each specific familiar context and not a regional variable representing an aggregate measure as it can be the child coverage ratio. The use of Attendance appears then to be indicated to describe this relation, but it may suffer of an endogeneity issue.

To consider Attendance as an exogenous variable, we should demonstrate that the enrolment in an early child care facility does not depend on the employment situation of the mother. In the literature we explicitly described the existence and the functioning of the eligibility criteria used by the public sector to ration the excess of demand; one of the criteria relies on the possibility to favour children from families

in which the mothers are working. Moreover, the choice to enroll the child to the early child care system may depend on the employment situation of the mothers; mothers that are working, as mothers with high job career expectations, may be more likely to try to enroll the child in an early child care facility than the unemployed ones. Simultaneity is then one of the clue that suggested us the presence of endogeneity for the variable Attendance with respect to Work. The regression in Table 17, due to a case of suspected endogeneity, should then not be used to describe the relation unless we do not find instruments for the variable to test whether it may be considered exogenous.

The instrument we propose in this research to specify a 2SLS model is the variable Total Accessibility, indicating the spatial proximity to the early child care system. For this variable to be a legitimate instrument it must affect Attendance, as shown in Table 15, and have no direct effect on Work. As validation for the use of the spatial variable as instrument for Attendance, we consider that the household decision on where to reside depends on many factors as the prices of accommodations, the proximity to the workplaces of the parents, in particular of the men due to the low female employment rate in Italy, the quality of life, the level of air pollution, and the level of accessibility to other services of an area, as the proximity to doctors, to parks, to the city center and so on; in this decision, a fundamental role it is also played by the proximity to relatives (who can be either care providers or care receivers) and friends. Moreover, the threshold of the 33% of child coverage ratio set by the European Council, as mentioned in Section 2, reinforces the assumption that when there is a substantial lack of the early child care provision, the distribution of the facilities depends on the population distribution and not on the employment rate of a particular area (Presidency European Council, 2002). We then consider the distribution of the early child care structures to be driven by the distribution of the population. Observing the regional

accessibility values, divided by the municipalities dimensions, we note in fact as the highest values are concentrated in the most populated cities to allow a greater number of children to access the service.
Even if applied in a completely different context, the approach used in the identification of the instrumental variable resembles the one in Card (1993) and Kling (2001), where the proximity to the college has been used as instrument for the college attendance in an analysis on earnings returns.
All these factors play against the argument that the accessibility measure used as instrument in our exercise can be strategically modified by households depending on women propensity to work, we then consider Total Accessibility to be a legitimate instrument for Attendance.

We proceed in the econometric analysis using a 2SLS procedure to study the effect of Attendance on Work through our instruments. In Table 18 we adopted as instruments for Attendance the Total Accessibility variable and its squared term. Thanks to this specification we are able to check, in the first stage, whether the effect of the early child care system on Attendance is decreasing as the service in a certain area becomes more and more present, meaning if it is stronger when positive variations happen in the case of low initial level of Total Accessibility. As expected, the coefficients of the Total Accessibility variable results positive and significant, 0.021, while the squared term negative and significant, -0.0002; an increase in the provision of the service has a stronger impact in the areas in which there is a substantial lack of the early child care system. These results suggest that the Total Accessibility variable may be indicated to be the instruments for Attendance.
In the second stage, the final effect of Attendance on Work results to be positive and significant, the coefficient indicates that having a child enrolled in the early child care system increases the probability for the

Table 18: 2SLS with Total Accessibility and its square as instruments for Attendance. HC heteroskedasticity-consistent standard errors have been used in the regressions.

	Dependent variable:	
	Attendance	Work
	OLS	*instrumental variable*
Age	0.009***	0.003
	(0.004)	(0.005)
Other_child	−0.074*	−0.049
	(0.038)	(0.051)
Education	0.131***	0.146**
	(0.050)	(0.065)
Couple	−0.014	−0.122
	(0.069)	(0.085)
Economic condition	−0.007	0.204***
	(0.039)	(0.050)
NE	−0.103*	0.174***
	(0.053)	(0.060)
NW	−0.103	0.060
	(0.070)	(0.069)
C	−0.133**	0.116
	(0.065)	(0.074)
Total Accessibility	0.021***	
	(0.005)	
Total Accessibility^2	−0.0002***	
	(0.0001)	
Attendance		0.658**
		(0.259)
Constant	−0.158	0.162
	(0.132)	(0.168)
Observations	439	439
R^2	0.093	0.026
Adjusted R^2	0.072	0.005
Residual Std. Error	0.394 (df = 428)	0.499 (df = 429)
F Statistic	4.378*** (df = 10; 428)	
Diagnostic	Statistic	p-value
Weak Instrument	10.826	2.59e-05***
Wu-Hausman	4.083	0.0439**
Sargan	0.052	0.8192
Note:		*p<0.1; **p<0.05; ***p<0.01

Own elaboration.

mothers to work of 65.8 percentage points. The difference between the coefficient for Attendance in OLS and 2SLS procedure explains that the OLS strongly under-states the effect on the probability for the mothers to work when the child results to be enrolled in an early child care facility. The 2SLS procedure, anyway, it is less precise than OLS, looking at the standard error of Attendance we note that the value is high, equal to 0.259; nevertheless the variable results to be statistically significant. Differently from the OLS regression, Table 17, the second stage produces different estimates for the coefficients, in particular considering the significance of the variables. The age of the mother, Couple and Other child are no more significant; while having a tertiary education, a good economic condition and living in the North-East of the country maintain the positive effect on the probability for a mother to work.
The 2SLS model is preferred to the OLS only when the instruments are not weak and the variable indicated as endogenous is effectively endogenous. The diagnostic tests are described at the end of Table 18. The Weak Instrument test has a statistic greater than the threshold level of 10 and also the p-value indicates that we must reject the null hypothesis of weak instruments. Wu-Hausman test for the endogeneity of Attendance rejects the null hypothesis of Attendance to be exogenous, confirming in our case the existence of a reverse causality issue and/or the presence of simultaneity. Sargan test for over identifying instruments accepts the null hypothesis of valid instruments.
The diagnostic tests confirm that the OLS model is biased and that the 2SLS procedure adopted produces consistent estimates.

We replicate in Table 19 the 2SLS procedure using the logarithm of Total Accessibility as unique instrument to study if the results obtained previously are robust to different specification of Total Accessibility. We focus directly on the second stage because the first stage replicates Table 15. The effect of Attendance on Work results to be

Table 19: 2SLS with the logarithm of Total Accessibility as instrument for Attendance. HC heteroskedasticity-consistent standard errors have been used in the regressions.

	Dependent variable:	
	Attendance	Work
	OLS	*instrumental variable*
Age	0.009***	0.004
	(0.003)	(0.005)
Other_child	−0.080**	−0.054
	(0.038)	(0.051)
Education	0.127**	0.160**
	(0.049)	(0.063)
Couple	−0.007	−0.125
	(0.067)	(0.080)
Economic condition	−0.015	0.203***
	(0.039)	(0.048)
NE	−0.126**	0.177***
	(0.056)	(0.058)
NW	−0.162**	0.066
	(0.076)	(0.068)
C	−0.163**	0.120*
	(0.069)	(0.072)
log(Total Accessibility)	0.104***	
	(0.022)	
Attendance		0.558**
		(0.248)
Constant	−0.167	0.151
	(0.129)	(0.164)
Observations	439	439
R^2	0.094	0.082
Adjusted R^2	0.075	0.063
Residual Std. Error (df = 429)	0.393	0.485
F Statistic	4.967*** (df = 9; 429)	
Diagnostic	Statistic	p-value
Weak Instrument	22.489	2.88e-06***
Wu-Hausman	2.678	0.102
Sargan	Na	Na
Note:		*p<0.1; **p<0.05; ***p<0.01

Own elaboration.

positive and statistically significant, having a child enrolled in the early child care system increases the probability for a mother to work of 55.8 percentage points. The coefficients of the other variables are similar to those obtained in Table 18, having a mother with a tertiary education and living in a family with a good economic condition increases the probability for a mother to work.
The results of the diagnostic tests of Table 19 differ from those presented in Table 18. The Weak Instrument test confirms that the logarithm of Total Accessibility is not a weak instrument, the statistic is higher than 10 with a value of 22.489; the Wu-Hausman test, anyway, with a p-value of 0.1 suggests that the variable Attendance should be considered exogenous or only marginally endogenous.

The information provided by the two models on the endogeneity issue of Attendance with respect to Work, suggest to treat this variable as endogenous, in particular referring to the results obtained in Table 18. Even if considering as instrument the logarithm of the Total Accessibility, rose the uncertainty on the effective endogeneity of Attendance, the p-value confirmed that we are facing a limit case for which it is difficult to clearly state that the variable Attendance is exogenous. The result obtained using two instruments, proven endogeneity of Attendance with a p-value for the Wu-Hausman test of 0.04, makes us lean towards the definition of Attendance as an endogenous variable. For this reason an OLS model with Attendance as regressor produces biased estimates; the 2SLS procedure instead, with Total Accessibility and its squared term as instruments, is able to correctly estimate the effect of having the child enrolled in the early child care system on the probability for a mother to work. The effect estimated in our model, Table 18, states that the probability to work increases of 65.8 percentage points.
The coefficient of Attendance defines directly the effect on the probability for a mother to work, differently from the child coverage ratio

that is a regional indicator able to describe the efficiency of the system on an aggregated level. Child coverage ratio has demonstrated anyway to be fundamental in the definition of the Total Accessibility variable and our research confirms the important role it play for investigating this topic.

5 Conclusion

In this essay we develop a measure of spatial accessibility of the formal early child care system in Italy; this measure has then been applied to analyse its correlation with the utilization of the early child care system and as an instrumental variable in a model analysing how the use of this service affects the maternal employment. Using a GIS software (QGIS) to map the distribution of the early child care facilities, both public and private, and the distribution of the Italian population; we built a spatial accessibility variable, thanks to a kernel density estimation procedure, able to measure the accessibility to the early child care system considering the distances between the families and the structures. This procedure permits to calculate the spatial accessibility value for every inhabited squared kilometer of Italy and to identify a variable describing the accessibility to the service at a regional level. The spatial variable allows us to measure which are the areas in a region that are more accessible and in which there is a lack of early child care facilities. To increase the precision of the variable, we divided the municipalities of each region in four groups according to the number of inhabitants to study the intra regional differences that exist between the spatial provision of the early child care system in small villages and in the highly populated cities. The four bands are: municipalities with less than 2,000 inhabitants, with a population between 2,001 and 10,000 inhabitants, with a population between 10,001 and 50,000 inhabitants and municipalities with more than 50,000 inhabitants. The lack of precise data forced us to consider approximations in the location of these observations introducing a measurement error in the analysis; the available data, due to privacy reasons, did not permit to know the exact location of the structures but only the municipality in which were located. To measure the magnitude of the error introduced in the model, we implemented a robustness check using the lists of early child care facilities of four

regions from the regional databases, with different years, for which we were able to know the precise location of the facilities. We compared the results of the spatial accessibility variable obtained locating firstly the facilities of these regions by the municipalities in which were situated, then using the addresses of the same facilities; the results show that the measurement error introduced not knowing the exact location of the structures was marginal, justifying the approximation of the model.

To analyse the effects of our accessibility measure on the decision of Italian households with children between 0-2 years old of age to use child care services, we combined our accessibility measure produced by QGIS with the data collected by the multipurpose survey "Multiscopo - Aspetti della Vita Quotidiana" run by ISTAT in 2013 and based on a sample of households representative of the Italian population. For privacy reasons, we do not know the exact place of residence of the households in the sample but we know their region of residence and the size of their municipality. The bands used by ISTAT to characterise the municipalities by the population dimension corresponds to those for which we already calculated the spatial accessibility variable; we then match each household in the survey with the corresponding accessibility variable considering the region in which it resides and the size of the municipality.

We estimated OLS linear regressions to assess how the probability of using the early child care system correlates with our spatial accessibility measure, controlling for individual and household characteristics, including the macro-area of residence and the coverage ratio. Our results show that 1% percent variation in our accessibility measure is associated with a statistically significant increase by 0.44% in the average probability of using child care facilities. This significance of the coefficient on our spatial accessibility index shows that it conveys additional predictive power to the model even if we are already

controlling for the aggregate regional child coverage ratio. For this reason we decide to calculate a new variable, called Total Accessibility, as the product between the spatial accessibility variable derived by the GIS model and the child coverage ratio; thanks to this variable we are able to gather the information on the early child care system in a single variable.

We then measure the effect of having a child enrolled in the early child care system on the employment probability of mothers with children between 0 and 2 years old of age. The OLS estimation shows that mothers with children enrolled in the system have an increased probability of being employed of 16.7 percentage points. Critics to this approach regard the endogeneity issue that concerns the use of Attendance as regressor for the working condition of the mother; the eligibility criteria, that may favour the enrolment in the system of children of working mothers, and the reverse causality, for which is more probable that are the working mothers those who look for the early child care system, suggest that the estimates produced with the OLS model are biased. We then implement a 2SLS procedure using a derivation of the spatial accessibility variable, the Total Accessibility variable (the product between the Spatial Accessibility variable and the regional child coverage ratio), as instrument for the probable endogenous variable. This approach assumes that our total accessibility measure is an exogenous instrument in the maternal employment equation. First, as stated in Barcelona targets, the early child care system should guarantee an adequate child coverage rate, at least 33%, for this service independently of the employment rate in a certain area (Presidency European Council, 2002). Second, matching the records of the ISTAT survey "Multiscopo - Aspetti della Vita Quotidiana" requires to aggregate accessibility indexes by region and group of municipalities based on their size. Although the area of residence is a choice of households, it is unlikely that they can have some direct con-

trol on the aggregate accessibility indexes defined in this way. Third, although mothers who want to work might have a clear incentive to live close to an early child care facility, it is also true that the place of residence depends on many other factors, including the prices of the accommodations and the economic resources available to the households, the proximity to the relatives of the parents (who can be either care-providers or care receivers), the proximity to doctors, and, given the sizeable gender gap in Italy for labour market participation, the workplace of men. All these factors play against the argument that the accessibility measure used as instrument in our exercise can be strategically modified by households depending on women propensity to work. Finally, although applied in a completely different context, our instrument resembles the approach followed by Card (1993) and Kling (2001), who use proximity to college as an instrument of college attendance in wage equations.

We implement two different 2SLS models to control for the endogeneity issue, in the first we use Total Accessibility and its squared term as instruments, in the second the logarithm of Total Accessibility. The first model in particular demonstrates the endogeneity of the variable describing the attendance to the early child care system; the Weak Instrument test rejects the null hypothesis of weak instruments, the Wu-Hausman test with a p-value equal to 0.04 confirms the endogeneity issue of Attendance with respect to mothers' working condition. The 2SLS specification, using Total Accessibility and the square term as instruments, describes that having a child enrolled in an early child care facility increases the probability for a mother to work of the 65.8 percentage points.

We demonstrate the importance of considering the spatial issue when investigating the impact of the early child care system on familiar choices, distances do influence the familiar choices in particular the mothers' decision to work.

Future research would try to acquire more refined data to locate the observations more precisely in the map to then calculate the spatial accessibility to the early child care system. It should also be analyzed if the results are similar using different methods in the calculation of the spatial relation, as the use of the road network.

6 Bibliography

References

Anderson, P. M. and Levine, P. B. (1999). Child care and mothers' employment decisions. Technical report, National bureau of economic research.

Baker, M., Gruber, J., and Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of political Economy*, 116(4):709–745.

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, 333(6045):975–978.

Becker, G. S. (1965). A theory of the allocation of time. *The economic journal*, pages 493–517.

Billari, F. and Kohler, H.-P. (2004). Patterns of low and lowest-low fertility in europe. *Population studies*, 58(2):161–176.

Blau, D. and Currie, J. (2006). Pre-school, day care, and after-school care: who's minding the kids? *Handbook of the Economics of Education*, 2:1163–1278.

Bratti, M., Bono, E. D., and Vuri, D. (2005). New mothers' labour force participation in italy: The role of job characteristics. *Labour*, 19:79–121.

Brilli, Y., Del Boca, D., and Pronzato, C. D. (2016). Does child care availability play a role in maternal employment and children's development? evidence from italy. *Review of Economics of the Household*, 14(1):27–51.

Burchfield, M., Overman, H. G., Puga, D., and Turner, M. A. (2006). Causes of sprawl: A portrait from space. *The Quarterly Journal of Economics*, 121(2):587–633.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. Technical report, National Bureau of Economic Research.

Chiuri, M. (2000). Quality and demand of child care and female labour supply in italy. *Labour*, 14(1):97–118.

Del Boca, D. (2002). The effect of child care and part time opportunities on participation and fertility decisions in italy. *Journal of population economics*, 15(3):549–573.

Del Boca, D., Locatelli, M., and Vuri, D. (2005). Child-care choices by working mothers: The case of italy. *Review of Economics of the Household*, 3(4):453–477.

Del Boca, D., Pasqua, S., and Pronzato, C. (2008). Motherhood and market work decisions in institutional context: a european perspective. *Oxford Economic Papers*, 61(suppl_1):i147–i171.

Del Boca, D., Piazzalunga, D., and Pronzato, C. (2018). The role of grandparenting in early childcare and child outcomes. *Review of Economics of the Household*, 16(2):477–512.

Del Boca, D. and Vuri, D. (2007). The mismatch between employment and child care in italy: the impact of rationing. *Journal of Population Economics*, 20(4):805–832.

Doiron, D. and Kalb, G. (2005). Demands for child care and household labour supply in australia. *Economic Record*, 81(254):215–236.

Foley, R. (2002). Assessing the applicability of gis in a health and social care setting: planning services for informal carers in east sussex, england. *Social Science & Medicine*, 55(1):79–96.

Gallagher, A. (2013). The politics of childcare provisioning: A geographical perspective. *Geography Compass*, 7(2):161–171.

Geertman, S. C. and Ritsema Van Eck, J. R. (1995). Gis and models of accessibility potential: an application in planning. *International journal of geographical information systems*, 9(1):67–80.

Havnes, T. and Mogstad, M. (2011). Money for nothing? universal child care and maternal employment. *Journal of Public Economics*, 95(11-12):1455–1465.

Hofferth, S. L. (1999). Child care, maternal employment, and public policy. *The Annals of the American Academy of Political and Social Science*, 563(1):20–38.

Hyndman, J. C. and Holman, C. J. (2001). Accessibility and spatial distribution of general practice services in an australian city by levels of social disadvantage. *Social Science & Medicine*, 53(12):1599–1609.

ISTAT (2016). Asilo nido e altri servizi socio-educativi per la prima infanzia: il censimento delle unità di offerta e la spesa dei comuni. anno scolastico 2013-2014. Technical report, ISTAT.

Kling, J. R. (2001). Interpreting instrumental variables estimates of the returns to schooling. *Journal of Business & Economic Statistics*, 19(3):358–364.

Langford, M., Higgs, G., and Dallimore, D. J. (2018). Investigating spatial variations in access to childcare provision using network-based geographic information system models. *Social Policy & Administration*.

Love, D. and Lindquist, P. (1995). The geographical accessibility of hospitals to the aged: a geographic information systems analysis within illinois. *Health services research*, 29(6):629.

McLafferty, S. L. (2003). Gis and health care. *Annual review of public health*, 24(1):25–42.

Morrissey, T. W. (2017). Child care and parent labor force participation: a review of the research literature. *Review of Economics of the Household*, 15(1):1–24.

O'Dwyer, L. A. and Burton, D. L. (1998). Potential meets reality: Gis and public health research in australia. *Australian and New Zealand Journal of Public Health*, 22(7):819–823.

OECD (2018). Programme for international student assessment (pisa) result from pisa 2018, country note, italy. Technical report.

Overman, H. G. (2010). "gis a job": What use geographical information systems in spatial economics? *Journal of Regional Science*, 50(1):165–180.

Phillips, R. L., Kinman, E. L., Schnitzer, P. G., Lindbloom, E. J., and Ewigman, B. (2000). Using geographic information systems to understand health care access. *Archives of Family Medicine*, 9(10):971–978.

Presidency, E. C. (2002). Barcelona european council. Technical report, European Council.

Ruhm, C. J. (2011). Policies to assist parents with young children. *The future of children/Center for the Future of Children, the David and Lucile Packard Foundation*, 21(2):37.

Silverman, B. W. (2018). *Density estimation for statistics and data analysis*. Routledge.

Van Ham, M. and Mulder, C. H. (2005). Geographical access to childcare and mothers'labour-force participation. *Tijdschrift voor economische en sociale geografie*, 96(1):63–74.

Sitography

EUROSTAT (2018). Formal child care by duration and age group. https://ec.europa.eu/eurostat/databrowser/view/tps00185/default/table?lang=en.

Google, D. T. (2019). Google earth pro. https://www.google.com/earth/versions/#download-pro.

ISTAT (2017). Annuario statistico italiano. https://www.istat.it/it/archivio/134686.

ISTAT (2019a). Datawarehouse istat. http://dati.istat.it.

ISTAT (2019b). Territory and cartography. https://www.istat.it/en/territory-and-cartography.

OECD (2018). Female employment rate. https://data.oecd.org/emp/employment-rate.htm.

QGIS, D. T. (2019). Qgis geographic information system. open source geospatial foundation project. http://qgis.osgeo.org.

7 Appendix

A GIS

Aim of GIS is to capture, store, manipulate, manage and analyse data from a geographic perspective; these features give to the researcher the opportunity to analyse different cases of study from a different point of view, investigating the spatial relations that may exist between the data that are difficult to understand with other statistical software. The use of GIS has been crucial for the development of many scientific and not scientific fields as Natural Science, Engineering, Architecture and Geography; and all of those subjects in which it is necessary to graphically plot images and data. To give some practical examples, this system is able to define the home range of an animal species in a bounded territory (allowing for the identification of the conditions that favour the presence of a certain animal); it can be useful in the analysis of climate change, tracking for example the rise of the sea level in a certain area for a certain period of time or the diffusion of a volcano's smoke after an eruption. The system may be used also in a social framework, it can be exploited for controlling the diffusion of specific diseases in a population considering variables as the daily movements of the people to contain its spread. In Economics it can be used to account for spatial competition of firms, when it depends on distance, or to improve models of Health Economics when talking about locations of hospitals and methods to reduce the costs of the health system in a country. While it is abundantly used in all these subjects, the GIS tool is still not completely exploited in many other cases; mostly because it is unknown.

While it appears clear the importance that such a system has for a geographer for example, it is not so immediate to understand how it can be used by an economist. To describe its use and some of the basic knowledge required for a first comprehension of this system, we are

going to refer to same authors that explained in their papers the most crucial aspects to be considered when talking about GIS and social science (Overman, 2010; Geertman and Van Eck, 1995; McLafferty, 2003; Burchfield et al., 2006; Gallagher, 2013).

With the acronym "GIS" we consider the geographic system as a whole, but there exist many platforms that use this system and implement on it different characteristics. In the GIS universe we can find open source software with limited features (as QGIS, the one used in our research) and paid software with more options (e.g. ArcGis). The most common types of formats for data that GIS elaborates are two: raster and vector. The raster format manages spatial data assigning a certain value to each cell on a regular grid (may it be square, or other regular spatial units); to visualize this format it is easy to think at the pixels of a screen, each pixel has a value that determines its colour. In the model adopted in the research the heatmaps were considered raster. The second format is the vector, vectors assign value to irregular polygons as line, points, surfaces and so on; let's say we want to list the characteristics and the exact positions of the trees of a public park, with GIS we are able to represent the extension of the park (by its geographic position in latitude and longitude) and for each tree, a point with a certain number of attributes as the species, its age and all those variable we are interested in. The same logic has been applied in our model when we considered the centroids representative of the distribution of the Italian population. GIS works differently according to the type of format in which the data are stored, but for both raster and vector data is able to graphically display the information they bring and to compute operations between different formats. One of the most simple operation is for example the count of points inside a polygon, while the most complex operations may regard analysis of the elevation of a mountain chain, water based analysis of streams or rivers and so on. In Regional Economic, thanks to the possibility to

implement and map images from satellites, GIS can be used in defining the land use in a region and analyse how it changes during the decades. On this topic, the Corine Land Cover project (CLC), with European Council as partner, aims to give to any citizen in the world a free and full access to an updated database on the land use of the European countries, with data easily readable by GIS systems. Data like these may be used in spatial and urban planning, forest and water management, in researches on the effects of the climate change and so on.

One of the most attractive aspect of the GIS structure is the possibility to combine data from different sources, manipulate them together and create new variables and datasets to be used in further investigations. While concordance between different sources of spatial units is common also for the statistical software, GIS is useful when such concordance is not readily available or is difficult to implement with standard econometric procedures. One of the most common example on this topic is the hedonic analysis of house prices, GIS can be used to identify the properties and their spatial characteristics (position, access to services and so on) that influence the final price from aerial photographs, combining variables from surveys and official dataset with images (Overman, 2010).

One of the potential use of GIS by an economist relies on the possibility to measure length of objects (line, polygons and so on) and distance between observations (e.g. average distance from firms to the railways). Through the geographic localization of the observations it is also possible to better characterize the variables thanks to the relations that these observations have with the surrounding environment, being it natural (elevation, rivers and so on) or artificial (road network, airports and so on). GIS allows the researchers to move from fixed boundaries, as NUTS in the European context or other kind of boundaries, to new patterns unbounded from the political limitations

and more connected to the economic sphere as it can be the analysis of commuting patterns or air pollution.
With the development of more and more efficient GIS platforms, able to compute new and more precise operations, it would be increased the possibility to integrate standard spatial econometric analysis with new sources of data. The demand of new data from the users of GIS will lead to the creation of new variables that could be useful also in economic studies.

To conclude, even if GIS software are not specifically designed to compute economic analysis, being aware of the existence of software able to compute conventional and not conventional spatial analysis may represent a possible answer to researchers willing to investigate such issue and to combine economic information with data from other scientific fields which use may be limited by the standard economic software.

B Kernel density estimation

In this Appendix we are going to briefly identify the characteristics of the kernel density estimation (Silverman, 2018). The density estimation, in the more general statistic framework, is the construction of the density function starting from the observed data. It is necessary when we observe data sampled from an unknown probability density function. While there exist many methods available for the calculation of the density estimation, using histograms for example, the kernel method appears to be one of the most efficient. To examine the characteristics of the kernel estimator we first need to define the naive estimator of which the kernel is a generalization.
Define the probability density function for a random variable X as:

$$f(x) = \lim_{h \to 0} \frac{1}{2h} P(x - h < X < x + h)$$

where h is the bin width of the interval.
Estimating the $P(x-h < X < x+h)$ as the proportion of the sample in the interval (n sample size), we derive the estimator $\hat{f}$ of the density given

$$\hat{f}(x) = \frac{1}{2hn}[\text{no. of} X_1, ..., X_n \text{falling in}(x - h, x + h)]$$

we call this the naive estimator.
Defining a weight function w equal to $\frac{1}{2}$ when $|x| < 1$ we rewrite the naive estimator as

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} w(\frac{x - X_i}{h})$$

The naive kernel estimator anyway is not satisfactory because $\hat{f}$ is not a continuous function, has jumps on the bin edges and the bin width itself is defined by the value h, which also controls how the data will be smoothed to produce the final estimate.

To overcome these issues, it is common to replace the weight function w by a kernel function K. The kernel function defines the shape of the density estimation, there exist many possible functions each with specific characteristics. Generally this function is a symmetric probability density function, which satisfies

$$\int_{-\infty}^{\infty} K(x)d_x = 1$$

and by analogy with the naive estimator, we define the kernel estimator as

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x - X_i}{h})$$

where h is named window width or smoothing parameter: when is small the spurious structure of the distribution becomes visible, when the value is large the estimated density appears smoother.
Moving to the univariate to the multivariate case is then straightforward to define the kernel density estimator with kernel K and window width h, in a d-dimensional space as

$$\hat{f}(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\frac{1}{h}(x - X_i)$$

For the purpose of our research we do not further describe all the other technicalities linked to the kernel density estimation (Silverman, 2018).