Università degli Studi di Milano
Università degli Studi di Milano - Bicocca
Università degli Studi di Brescia
Università del Piemonte Orientale

Ph.D. Programme
Economic Sociology and Labour Studies - 31st cohort

Doctoral Thesis

Neighbourhood and school segregation in urban Italy

Ph.D. Candidate:
Matteo Piolatto

Supervisor:
Prof. Gabriele Ballarino
Prof. Emmanuele Pavolini
Ph.D. Programme Director:
Prof. Gabriele Ballarino

SPS/09, SPS/07, IUS/07, SECS-P/07, SECS-P/10, SECS-S/04, M-PSI/06
Academic Year 2018/2019

The PhD programme in Economic Sociology and Labour Studies (ESLS) stems from the collaboration of four Universities, namely Università degli Studi di Milano, Università degli Studi di Milano - Bicocca, Università degli Studi di Brescia and Università degli Studi del Piemonte Orientale "Amedeo Avogadro". The University of Milan serves as the administrative headquarters and provides the facilities for most teaching activities.

Acknowledgements

This dissertation is the result of personal efforts, collective discussions, midnight suggestions and tea. I'd like to thanks Gabriele Ballarino and Emmanuele Pavolini for their teaching, suggestions, motivation and supportive supervisory. Beside my advisors, I would like to thanks Gianluca Argentin for its useful suggestions and Hans Schadee for the discussions and hypothesis betting on school and residential segregation. My sincere thanks also go to Marco Oberti, Carlo Barone, Tommaso Vitale and Bruno Cousin for the hints and discussions provided during my six months of visiting at the *Observatoire Sociologique du Changement - Sciences Po*. The final thanks are for my colleagues, especially for the time spent together in the difficult sea of firs year phd courses; my French *colocs*, Milena and Vivien for their supportive help during my French stay the theorethical discussions we shared and LaTeX help; my parents and sister, although they still ask what's my job about; my wife for her support.

Contents

List of Figures

List of Tables

List of Abbreviations

ACT	**A**merican **C**ollege **T**esting
CAD	**C**umulative **A**dvantage **D**isadvantage
EMI	**E**ffectively **M**aintained **I**nequality
EU	**E**uropean **U**nion
IALLSS	**I**nternational **A**dult **L**iteracy and **L**ife **S**kills **S**urvey
IALS	**I**nternational **A**dult **L**iteracy **S**tudy
INVALSI	**I**stituto **N**azionale per la **VAL**utazione del **S**istema educativo di **I**struzione e di formazione
ISTAT	**I**stituto Nazionale di **STAT**istica
MTO	**M**oving **T**o **O**pportunity for Fair Housing program
OECD	**O**rganization for **E**conomic **C**ooperation and **D**evelopment
OLS	**O**rdinary **L**east **S**quares
PIAAC	**P**rogramme for the **I**nternational **A**ssessment of **A**dult **C**ompetencies
PIRLS	**P**rogress in **I**nternational **R**eading **L**iteracy **S**tudy
PISA	**P**rogramme for **I**nternational **S**tudent **A**ssessment
PSID	**P**anel **S**tudy on **I**ncome **D**ynamics
SAT	**S**cholastic **A**ssessment **T**est
TIMMS	**T**rends in **I**nternational **M**athematics and **S**cience **S**tudy
US	**U**nited **S**tates

Chapter 1

Residential and school segregation

This first chapter is devoted to review national and international research on residential and school segregation in western societies. In Merton's words, the aim is to establish the phenomenon and provide a presentation of its main causes and effects. Section 1.1 investigates and discusses past research in order to provide a clear and comprehensive definition of segregation. Section 1.2 presents the main literature on residential segregation and its trends providing evidence for the United States, Europe and Italy. Causes and effects of residential segregation are discussed respectively in Section 1.3 and Section 1.4.

The second part is devoted to present research on school segregation. In Section 1.5 a literature review on levels of school segregation across countries (United States, Europe and Italy) is presented, while Section 1.6 presents a discussion on the main causes of school segregation. Section 1.7 presents a review of school segregation effects focusing on the consequences of school composition on individual educational performances, moreover peer effects are presented. Finally, conclusions of this chapter are presented in Section 1.8.

1.1 What is segregation?

The term segregation, namely the act of segregating, derives from Latin and is the noun of action from past participle of the verb *segregare*, a composite word of *grex gregis* "flock" and the prefix *se-*, to separate from the herd, isolate. Thus, segregation refers both to a separated environment and to the action of isolating. Within the fields of urban sociology and sociology of education the term usually refers to a situation in which groups experience separated environments in the urban space and schools, or policies that produce discrimination. However, it is rare to find completely segregated situations and the term usually refers to environments that are not mixed or heterogeneous in some degree.

Research on the different forms of social segregation and its effects has flourished since the second half of the 20th century across the academic communities of western countries and especially between 1920s and 1990s in the US. In the first decades of the new millennium the interest in such a topic spread also in Europe, as a result of several determinants: the need to understand the effects of migration flows in the '80s

and '90s; new migration flows from Africa and Middle East; the need to understand the link between segregation (especially its residential manifestation) and social stratification. Early concern with segregation derived from interest in ecological studies, an approach that stressed the idea of segregation as a natural phenomenon. In fact, while ecology is the branch of biology that considers the distribution of plants and animals in nature and their interdependence, human ecology is the study of the spatial and temporal relations of individuals as affected by the selective, distributive and accommodating forces of the environment (McKenzie 1924). Competition and selection are the main driving forces of spatial relationships in the ecological approach, as both change over space the physical basis of social relations is altered producing social and political attrition. A more cultural-based approach developed in criticism of the natural one, highlighting the influence of other factors on the spatial distribution of groups' population, especially personal values. Scholars from the cultural-based line focused especially on the role of prejudice, stigma and discrimination as the main determinants of segregation. Starting from this approach, economists implemented also the role of preferences and asymmetry of information, factors that have been studied extensively in the residential segregation literature.

Separation between individuals has been investigated in several domains of human life, most notably in patterns of residential mobility and school enrolment, and is still a trending research topic due to its implications for everyday life and its relevance in the policy design framework. This huge body of research addresses segregation as the uneven or non-random distribution of individuals in a given environment according to some sociological characteristic, usually income, social status, sex and ethnicity.

Following Reardon and O'Sullivan (2004), segregation can be considered as the extent to which individuals of different groups experience different social environments. The space in which segregation manifests itself is defined in broad terms as "social environment", thus making this definition relevant to the several fields of the social sciences' inquiry. The literature refers to social environment as a category that includes several components: the set of relevant social and cultural institutions, along with patterns and processes that influence the life of an individual or community. More specifically, elements of the social environment can be social and economic processes, power relations and beliefs about place and community (Casper 2001) and their variation over time and space.

This definition lacks two important elements. The first is the reckoning of the role of physical space (urban, schools, officies, etc.) as the place where social mechanisms translate into the differential distribution of groups, or differential concentration. Moreover, the physical side of segregation is the one that allow us to investigate the social mechanisms that determine it. This is important also in relation to the question of inequalities in opportunities and resources as dismissing the physical side of segregation makes difficult to focus on the actual inequalities and policies to address these. The second is the ability to clearly address the several structural mechanisms

underlying segregation. Schelling (1971) suggested the role of voluntary and compulsory behaviours, practices of organizations and cultural codes (e.g. Indian caste society, where a consensual basis for segregation was derived from shared religious values) and the interplay of individual choices. Moreover, segregation in some domain may result as the corollary of segregation in other domains.

Oberti and Préteceille (2016) build on this literature and group causes of segregation[1] according to three dimensions: economic factors, institutional factors and individual preferences. The broad range of economic determinants varies across the several fields in which segregation is investigated: in studies on residential segregation they relate mainly to the supply and demand of houses in the urban space, as well as their construction over time, while in occupational segregation studies they refers to the structure of the economic system. The second factor relates to a broad category of institutional features: policies, that may have explicit segregating aims, as the apartheid regime in South Africa, or unintended segregating consequences, but even the codification and institutionalization of cultural and religious codes. The classical example is the Indian caste society. Finally, individual factors that can be divided in two groups: positive and negative preferences. The first group includes behaviors that lead to aggregation of similar individuals, like homophily (McPherson, Smith-Lovin, and Cook 2001). The second group includes behaviors which effect is to generate physical distance between groups. Researchers have explained these behaviours as the consequence of a certain "taste for discrimination" in individual preferences (Becker 1957).

The debate around the interplay of these three main categories of causes of segregation developed since the late '80s thanks to the seminal contribution of Massey and Denton (1988). Their work addressed the complexity related to the mechanisms responsible for residential segregation by classifying five different aspects of the phenomenon: *evenness*, the differential distribution of two or more social groups in the social environment; *exposure*, the probability of contact or interaction between group members within a urban environment; *concentration*, the relative amount of physical space occupied by a minority group in the city; *centralization*, the degree to which a group is spatially located near the center of an urban area; *clustering*, the extent to which areal units inhabited by minority members adjoin one another in space. This line of reasoning highlights the multidimensional character of the concept. Moving from this discussion, segregation can be conceptualized as follows.

Definition:. *Segregation is a more or less institutionalized form of social distance between groups that appears in several ways and through different mechanisms, and results in some degree of separation in the physical environment.*

1. Causes of segregation are discussed in detail in Section 1.3

1.2 Residential segregation across countries

The initial steps toward the study of segregation were taken in the 1920s by the so called Chicago School, a group of researchers based at the Department of Sociology at Chicago University with a strong interest in urban sociology. The aim of the faculty members was to develop a more scientific approach of sociology, in antagonism to the more "philosophizing" one of the American sociology of that day that was yielding little consistency in the formation of social policy (Lutters and Ackerman 1996). This attempt was characterized by the adoption of basic concepts from the natural sciences, especially ecology. The ecological approach was centered on the role of competition and selection as the main driving forces that shape the formation of human natural areas. The activity of the Chicago School resulted in a huge body of research developed in the first half of the 20th century, especially in the period between the arrival (1914) and the retirement (1934) of Robert Park, one of the most prominent scholars of that group. His work, along with that of Burgess, were fundamental in understanding the phenomenon of residential segregation in American cities in the 1920s and 1930s. They focused especially on the city as a social laboratory in which to investigate social relations through rigorous qualitative methods. In fact, the primary assumption for the Chicago School was that qualitative methodologies were best suited to investigate the complexity of urban social interactions (Lutters and Ackerman 1996). Of course, while these methods allowed for in-depth analysis of the ongoing social processes, they did not permit to understand the entire process at the urban scale (Oberti and Préteceille 2016). As quantitative methodologies became more popular to investigate social phenomenon on large scale, the qualitative approach became one of the School's greatest liabilities. Starting from the '50s, the ethnographic approach in urban studies was progressively ruled-out, as more and more powerful instruments allowed scholars to exploit the rich set of information available in national census data and to rely on larger samples of metropolitan areas.

1.2.1 United States

Segregation by socio-economic status and ethnicity is a key issue in the American society since the beginning of the 20^{th} century, when the industrial expansion of northern cities attracted masses of black Americans from the south. The creation of segregated social environments in the U.S. was triggered by the movement of masses of southern African Americans to industrial cities in the mid-west and northeast after the first world war, and the contemporaneous adoption by the local and federal governments of discriminatory public housing (King 1995) and urban renewal policies (Weiss 1987). In fact, prior to the internal post-wars migration flows, and also during the 19^{th} century, the racial concentration of ethnic minorities was fairly low across American cities (Weaver 1948). At the local level two actions, both subsequently declared to be unconstitutional by the Supreme Court, facilitated the concentration of

minorities (especially blacks) in distinct parts of the city: zoning and racially restrictive covenants. The origins of zoning trace back to Los Angeles in 1908, Baltimore in 1910, Louisville, New Orleans and other southern cities (Weaver 1948) in the '20s and '30s. Zoning was aimed to regulate and control the type and intensity of land use along with the height, area location and use of buildings. However it was not only an instrument to shape the built environment and to stabilize land values, it soon became a way to achieve specific social objectives. Zoning was used as an effective tool to reshape the urban social landscape by protecting property values and excluding the undesirables (Silver 1997), especially African Americans and migrants. In 1917 the US Supreme Court ruling made racial zoning ordinances unenforceable (*Buchanan v. Warley*), thus a new instrument emerged: racially restrictive covenants, contracts between landowners or real estate agencies aiming at excluding African American from entering the property. The main idea behind this contract was that racial separation of residences was necessary to maintain property values, real estate profits and neighborhood stability (Gotham 2000). Explicit racially restricted covenants were used between 1900 and 1965 (Gotham 2000) especially in developing neighbourhoods, in order to prevent any block by block expansion of the black neighbourhood and to encircle existing black ghettos (Kushner 1979). As a result, separate house markets for whites and blacks were created. Later on, other factors contributed to the creation of urban ghettos: manufacturing jobs were lured away from the inner city centres, while industrial activities moved from the city toward the suburbs (Seitles 1998). Politics played a key role in shaping the spatial positioning of people: i.e., the federal government granted no-down-payment mortgages to veterans and granted them facilitated access to new houses. Since the policies of the armed forced resulted in greater numbers of whites being eligible for veterans benefits (Kushner 1979), blacks remained "trapped" in the city centres. Housing resettlement policies, implemented in order to mitigate the perverse effects of veterans benefits, had the same effect: facilitating whites' reallocation outside poor and racially or socially segregated neighbourhoods.

A relevant amount of research has extensively investigated the evolution of urban space's composition all over the last century, highlighting a general persistence of high levels of separation between ethnic groups. More precisely, the American society seems to be characterized by a sort of ranking based on social distance between ethnic groups: the bottom of the rank belongs to black Americans, the less desired group as neighbours, while higher stages of the social desirability ladder are occupied by Hispanics (with the exception of those with black antecedents) and Asians, the group more assimilated to white Americans.

High degrees of segregation has been well documented between occupational groups. One of the most relevant study for the discipline was the investigation of the residential distribution of occupational groups over tracts and zones sectors of Chicago in 1950 by Duncan and Duncan (1955). Subsequent studies extended this line of inquiry in time and space: Wilkins (1956) added eight middle-size cities in

1950 to the analytical sample;[2] Uyeki (1964) replicated the Duncans' study on Cleveland considering the 1940-1960 period, while Simkus (1978) provided a replication of the previous three studies in the total of the ten urbanized areas previously considered for the 1960-1970 period. These very first studies relied on the computation of distance matrices[3] for segregation indexes between each pair of occupational categories. If it is assumed that the ordering of major occupational groups corresponds with increasing social distance (Duncan and Duncan 1955) such that, e.g., professionals are more social distant from labourers than from sales workers, a general result in these studies should be that there is a correspondence between social and spatial distance across occupational groups. An expectation consistent with the human ecology framework. I.e., the relative dissimilarities of the residential distributions of members of occupational categories found by the Duncans resemble the expected differences in the general socioeconomic status of those occupations. Such a distance was confirmed by the replication of Simkus (1978), that highlighted also the presence of a mild trend of increasing residential segregation by occupational categories during the 1950s but no substantial change in the following decade. The small changes in urban segregation by occupational categories registered may reflect the composition of several potentially nested mechanisms, such as discrimination in the labour or house market, status groups differences in life style, place of work or affinities. Simkus estimated that up to one quarter of the gross occupational residential segregation could be attributed to racial segregation in both the occupational and residential distributions.

Coming to residential segregation by ethnicity, several studies proved the validity of the social distance assumption between ethnic groups. Guest and Weed (1976) studied the changes in patterns of ethnic residential segregation in the metropolitan areas of Cleveland (1930-1970), Boston and Seattle (1960-1970) and identified three major groups according to the degree of residential segregation experienced: old immigrant groups (Canadians and Northern Europeans), new immigrant groups (Italians and Eastern Europeans) and non-Europeans (Blacks, Mexicans and other non-whites). Two patterns were detected in the overall period: on one side, the segregation of the black population remained high over time while other non-white groups showed a general decrease; on the other, the segregation of new immigrant groups from Europe decreased while that of older groups increased. Segregation patterns for the next decade remained substantially unchanged. Massey and Denton (1987) found no substantial evidence of a significant change in levels of segregation across 60 Metropolitan Statistical Areas, confirming the validity of Guest and Weed's (1976) results but unveiling huge differences between ethnicities in the non-Europeans group. In fact, levels of segregation and probabilities to get in contact

2. The cities considered were: Hartford, Syracuse, Chicago-Gary, Cleveland, Columbus, Indianapolis, Fort Worth, Atlanta, Memphis and Richmond.

3. A distance matrix is a square matrix containing the pairwise distances between the elements of a set. In this case between census tracts.

with the dominant group (Anglos) were not the same for all the ethnic groups considered, *inter alia* Hispanics and Asians were those experiencing the higher degree of spatial assimilation, while African Americans remained entrapped in homogeneous neighbourhoods. The authors identified the main cause of these differentials in the non-participation to the sub-urbanization process, that due to the effects of previous local and federal policies was generally precluded to African Americans. The situation experienced by black Americans was defined as hyper-segregation (Massey and Denton 1989, a concept related to the different dimensions of segregation. In order to be hyper-segregated, a urban area must meet four of the five following criteria: minority members are unevenly distributed, namely the percentage of minority members in a given residential area doesn't equal the citywide minority percentage; minority members are isolated and have small potential contact with majority members; minority neighbourhoods are close one another in the urban space and clustered in certain parts of the town; minority members are settled close to the central areas of the town; minority members are concentrated within small and geographically compact areas. An extensive review of studies on residential segregation by ethnicity (Charles 2003) has shown that urban areas were far from being able to equalize or reduce the levels of segregation experienced by black Americans. In fact, in the last decades of the 20th century (1980-2000) the number of metropolitan areas matching the hyper-segregation criteria nearly doubled, passing from 16 to 29 (Iceland 2004), while in the same period the levels of residential segregation, both between ethnic groups and metropolitan areas, persisted or increased (Logan, Stults, and Farley 2004). Finally, in the first decade of the 21st century the degree of segregation of American metropolitan areas changed very little, although the number of hyper-segregated areas decreased to 21 (Massey and Tannen 2015).

Research on trends of segregation has shown the persistence of general high levels of segregation in the American society. However, composite effects are at place as trends depend by ethnic groups, economic status and structural factors of the residence area. In fact, the segregation experienced by African Americans slightly declined for those living in mid-sized metropolitan areas but remained the same in highly segregated metropolises (Farley and Frey 1994). On the other hand, segregation for Hispanic and Asian Americans increased at a small or moderate rate. This could be linked to the rapid increase of these groups' population, that may have resulted in a substantive change in the ratio with white Americans. However, attitudes and behaviours of the dominant group matter too and the conservative attitudes of whites and their fears in becoming a minority in a neighbourhood may limit the desegregation that can occur (Farley and Frey 1994).

1.2.2 Europe

European cities have been at the centre of massive outward and inward migration flows since the second half of the 19th century. The constant flux of people towards north and south America (especially United States, Canada, Argentina and Brazil)

between 1846 and 1940 has been the most relevant flow, involving about 60 million people from Europe, one-half of which from the British Isles. The migration process involved also North Africa, a minor destination with about 3 million French and Italians moving between the two continents (McKeow 2004). Beside the huge transoceanic migration flow, several resettlement movements took place within the same sending and receiving areas. Industrial areas in northern Europe were the main destination of internal migrations: British industrial cities received Irish workers while continental urban areas were the destination of eastern and southern Europeans. Differently from transatlantic migration flows, internal migrations resisted also after 1945. In the aftermath of World War II a general and persistent pattern of redistribution of the population within Western European countries occurred. Urbanization, or the rural depopulation and out-migration from old industrial regions towards the metropolitan areas with high concentration of modern consumer goods industries (Fielding 1989), was the dominant trend in the 1950s, powered especially by the development of industrial areas close to the city. In the 1970s and 1980s the urbanization process broke down and most of the Western Europe countries experienced counter-urbanization, a process characterized by a negative relation between the size of a human settlement and its capacity to attract new migrants. Finally, since the mid 1980s European cities became the destination of external migration flows. It's the case of Germany for Turkish and Northern Africa migrants, France for migrants from Central Africa and Italy for those coming from the Balkans.

Motivated by the interest in the effects of this huge movement of people, a well established stream of literature has been dedicated to the residential patterns of migrants in European cities, both in Northern and Southern Europe. Malheiros (2002) has summed up the different distributional patterns in a three-group scheme. The first aggregate includes the industrial metropolitan areas in Northern Europe characterized by a high percentage of non-EU foreigners that are spatially concentrated in the city centre, while natives are over-represented in the peripheral area. Migrants in these cities are experiencing the same pattern that characterize metropolitan areas in the US. It is the case of Brussels (Dujardin, Seldon, and Thomas 2008), Goteborg (Bråmå 2008), Birmingham and Rotterdam (Arbaci 2007) where the presence of non-EU immigrants is higher in the city centre due to a more favourable housing market.

The second group collects consolidating migration destinations, especially from the Western Mediterranean area, along with few Northern European cities. Urban areas in this group are characterized by a smaller proportion of non-EU foreigners and by a general concentration of migrants in the outskirts. Low or moderate levels of residential segregation, along with patterns of scattered peripheral settlement, have been documented in Milan, Madrid and Lisbon (Arbaci and Malheiros 2010). Surprisingly, Paris is not included in the sample of Malheiros's (2002) investigation, despite several contributions have shown that ethnic minorities are concentrated especially in the peri-urban and metropolitan areas. Préteceille (2009) has shown that the overall segregation of ethnic minoritis in Paris has been at moderate levels, but

it has increased at a low but constant pace between 1982 and 1999. This result has been confirmed both by a replication (Verdugo 2011) and an investigation of trends in segregation by ethnicity and social status in French urban areas (Pan Ké Shon 2010).

The third group is characterized by small metropolitan areas with high shares of migrants from other European countries. With respect to non-EU migrants, the cities in this cluster (Dublin, Geneve, Luxemburg) close to the first group than the second. With respect to these groups, some exceptions are present. Milan is characterized by the same spatial disposition of ethnic minorities than cities in the second group, however levels of segregation are higher than the average of the group (Malheiros 2002). Naples has extremely low levels of ethnic segregation, but concentration of minorities in the port zone, as several Italian port cities (Barbagli and Pisati 2012). Athens is an interesting case although not included in the sample of Malheiros' study. In fact, it presents features of the first and second group. It is not a Northern European city but, according to Maloutas (2007), ethnic minorities prevalently reside both in the old city centre and in scattered peripheral areas where there is a more affordable house-market. Finally, to the author's knowledge no studies on the spatial disposition of ethnic minorities in Eastern European cities have been carried out.

Despite the increasing level of inequality in Europe, which is likely to be expressed also spatially, a comparable interest has not been shown by scholars with respect to residential segregation by social status. However, some remarkable exceptions do exist. Using Swedish Census data Vogel (1992) has shown a trend of increasing residential segregation by socio-economic characteristics in several city centres, areas interested by numerous public intervents of urban renewal. In its comparison of segregation indexes by cities in previous studies on the topic, Musterd (2005) concluded that:

> Segregation levels are low, which implies the existence of many socially mixed neighbourhoods. In general, these findings indicate that in Europe the poor are not detached from the middle classes.

Although the urban space might be more socially mixed in Europe than in the US, the issues about the coexistence of people with different social status, and in particular of poor and middle classes, may be qualitatively different. First, the degree of coexistence varies according to the urban area size, as small cities might experience lower levels of segregation. Second, it is a methodological and measurement issue related to the statistical approach used (e.g. the type of segregation index used), the operationalization of the variables of interest and the dimension of the spatial units observed. Finally, the concept of *middle classes* is a broad one and represents a category characterized by qualitative differences between the occupations belonging to this group. These differences are pivotal in explaining the existence of socially mixed neighborhoods. Maloutas (2007) has shown that during the 1990s the middle classes in Athens have participated to the changes in the composition of the urban space in different ways. Intermediate and lower technical occupations have followed

the upper classes in their movement toward more homogeneous neighborhoods in the eastern part of the city, while the self employed and white collars remained in areas with a high presence of lower occupations. Something similar has happened for the metropolitan area of Paris, where the different occupations belonging to the middle class group share spaces with both upper and lower classes (Préteceille 1995, 2003).

1.2.3 Italy

The Italian case is largely under-represented in the international literature about urban segregation and the social division of metropolitan spaces. The available research covers the last two decades with a focus on the residential patterns of ethnic minorities rather than social segregation, and by paying little attention to the comparative perspective. However, some exceptions do exist. Petsimeris (1998) has compared the changes in population composition of the three cities of the *Italian Industrial Triangle*[4] in the decade between the two census waves of 1981 and 1991. Despite the difference in size and economic base, the three cities have all shown a general reduction in total population, especially among the working class (deproletarization) and the increase in relative concentration of the wealthy (professionals and business owners). The decrease in the numbers of the working class have involved especially the central and peri-central areas of these cities, leading to an increase in segregation indices for the working class during the considered period. In the same period the immigrant population of these cities increased substantially. Industrial cities of the North-West started to attract immigrants from North Africa (especially Marocco and Tunisia) and the Balkans (Albania) that initially located in the central areas of hosting towns.

The findings by Petsimeris (1998) have been substantially confirmed by subsequent investigation: in their analysis on the spatial patterns of individuals by ethnicity and social status in the eleven bigger Italian cities,[5] Barbagli and Pisati (2012) highlight a general trend of depopulation, gentrification and increase in the immigrant population. The diminishing figures of population is found especially in the city centres, as since the third decade of the 20th century these areas gradually became places of commercial activities rather than residential ones. The transformations of the city centres pushed upper classes "to migrate" from the central areas to other urban areas. As a consequence, their under-representation in peripheral zones decreased while the over-representation of the working class in the industrial areas of the town declined, probably as they redistributed within the city perimeter or most likely moved to adjacent municipalities. Milan is a typical example of these urban trends: the overall population growth rate has been negative in the period 1971-2011 with an increase in population of adjacent municipalities, in the meantime the share of business owners, professionals and independent workers increased while that of dependent workers decreased dramatically (Petsimeris and Rimoldi 2015).

4. Milan, Turin and Genoa.
5. Turin, Milan, Genoa, Venice, Bologna, Florence, Rome, Naples, Bari, Palermo, Catania.

With respect to both the occupational and spatial dimensions, a more fine grain analysis by Cousin and Préteceille (2008) has shown that professionals, entrepreneurs and in more general terms occupations belonging to upper classes are over-represented in the municipality of Milan with respect to the whole metropolitan area, on the other side the working class is under-represented. Moreover, census data for 1991 have shown that upper classes continue to reside in the city centre and in the western and eastern peri-central areas of Milan, while middle classes prevalently occupy the surrounding ring inside the municipal borders and working classes in the metropolitan areas.

Finally, Italian cities have been interested by the expansion of immigrant population, as a result of both migration flows and family reunions. Several studies have investigated the spatial distribution of migrants, finding some differentiation between coastal and inland cities. Immigrants in coastal cities are usually concentrated in the areas close to the port, historically characterized by a high concentration of working class indivduals and where rents are more accessible. That's the case of Genoa (Petsimeris 1998; Scarpa 2015), Naples (Mazza, Gabrielli, and Strozza 2017; Lelo, Monni, and Tomassi 2018) and Palermo (Barbagli and Pisati 2012), where migrants are greatly over-represented in the port zones with no distinction between ethnicities. The presence of migrants in inland cities is usually concentrated in peri-urban areas, however the 1991 census data have shown a consistent presence of migrants in city centres in Milan and Turin (Petsimeris 1998) with a subsequent dispersion towards peri-urban areas (Barbagli and Pisati 2012). These data highlights two facts. First, the initial settlement of migrants in these cities took place in city centres, probably due to more favorable housing conditions or public housing provisions. Second, across time the residential distribution of migrants has moved in the same direction of gentrification, that is from the center towards peri-urban areas.

1.3 Causes and mechanisms of residential segregation

The broad range of macro and micro level factors that have been highlighted as determinants of segregation can be summed up in three groups, according to as many dimensions (Oberti and Préteceille 2016): economic factors, institutional factors, individual preferences.

To the first group belong economic status, job location, discrimination in the labour, financial and housing market. In modern capitalist societies the more prestigious occupations come with the higher wages, which allow to have (by own means) and gain (by loans) a higher purchasing power on the housing market. Economic status generate a differential access to the housing market among occupations, such differences are reflected in the material and symbolic sorting of the urban space according to the income and the social status of residents (Oberti and Préteceille 2016). Job location is another economic factor that can help in explaining observed levels of segregation: workers should tend to cluster around their places of employment

in order to reduce time and costs associated to commuting (Galster 1988). However, the relevance of these mechanisms is questionable as they are not sufficient to fully explain the observed levels of residential segregation, not even a substantial part of it.

In a critical review of the causes of segregation in the American society, Galster (1988) has shown that there is no support to claim that the contribution of the economic factors to the overall levels of segregation is more than 10%, while individual preferences along with public and private discrimination may account for about 50% of it (Clark 1991). Given the multi-dimensionality of segregation, other causes should be considered in the economic domain such as private discrimination in the housing market. This can be conceptualized as the broad variety of discriminatory acts by landlords or real estate agents aimed at excluding minority home-seekers from non-minority neighbourhoods in which they would like and are financially able to move. This line of reasoning rely on two strong assumptions: (i) an amount of private discrimination exists; (ii) the level of private discrimination is associated with a fraction of observed levels of segregation. Scholars have tested the validity of these hypothesis by relying on experimental design, especially audits. Audit studies are a type of field experiment used to test for discrimination in which simulated market transactions are reproduced under controlled conditions. Two individuals are matched on all relevant characteristics except the one presumed to lead to discrimination. Subsequently, each person applies for the same job, housing, etc. The differential treatment they receive provides a measure of discrimination (Fix and Struyk 1993). Yinger (1998) has reviewed several audit studies from the 1980s and the 1990s in different consumer markets in the US, showing evidence for the presence of discrimination. Similar results were obtained by Galster (1990) in a review of fifty local audit studies run in the US in the 1980s. The U.S. Department of Housing and Urban Development's 1977 Housing Market Practices Survey estimated the effect of discriminatory practices at about 15% in the sales market and 27% in the rental market, a far more influent effect then that of economic status. The last economic factor of interest is information. Information in the housing market may be stratified according to some individual characteristic, e.g. members of minority groups may have not access to complete information on costs and quality of the housing market. As a result, they may underestimate houses' value thus choosing sub-optimal solutions. However there is mixed evidence in support of this factor. Farley, Danziger, and Holzer (2000) found little differences among ethnic minorities when coming to the estimation of the average cost of homes in several areas of Detroit, while Thompson (2000) found significant racial differences in the evaluation of the housing costs in Atlanta. Finally, the stock of housing supply and the evenness of its distribution in the urban space might determine segregation in some degree. For example, the concentration of low quality and cheap buildings in specific urban areas might allow the clustering of low-income households in that area.

The second group of causes of segregation gathers institutional factors that play

a role in creating a segregated urban environment. Public housing policies probably represent the most relevant institutional factor in this domain, as they involve directly all the elements that influence the distribution of people in the city. These policies might achieve segregation in different ways: (i) public housing buildings may be concentrated in a given part of the urban space, or might be concentrated in more disadvantaged areas with poor local transportation connections etc; (ii) housing policies may be connected to urban planning regulation and its effects, as in the US and France (Oberti and Préteceille 2016); (iii) residential segregation may results from policies in other public domains, or from the cumulative effect of policies in different domains. The most evident case is represented by policies of school choice. If the enrolment to schools depends on the residential position, households will try to move to the neighbourhoods that guarantee access to better schools (as in the US or the UK).

Finally, individual preferences of both dominant and minority groups constitute the third main cause of segregation. The complex role of this factor has been firstly highlighted by Schelling (1971), whose argument relied on the idea that the magnitude of differentials in preferences may not be that relevant, however even small differentials can lead to polarized results. The Schelling's theoretical model for neighbourhood population is based on the following assumptions: (i) there is a shared definition of neighbourhood and its borders; (ii) everybody is concerned about the composition of the neighbourhood. In this model households are the agents of interest, each of them has a preferred alternative in the composition of the neighbourhood and they will stay in the place until the composition changes such that a given threshold is reached. Once the threshold is reached, agents will choose another location as the neighbourhood they are living in does not match their preferences anymore. The model can be thought of also in terms of incentives: an household will reside in a given neighbourhood until the positive incentives (e.g. the closeness to members of certain groups) will outnumber the negative ones (e.g. the distance from members of certain groups) (Oberti and Préteceille 2016). While Schelling recognizes the complex process that lead to the formation of individual preferences, and that *preferences per se* hardly exist, he does not address the preferences' formation mechanism. Subsequent research has focused on unveiling the black box of preferences motivation formation. Scholars in the US have investigated the causes of persisting racial residential segregation in metropolitan by addressing preferences. The results of these multi-methods works (Clark 1991; Emerson, Chai, and Yancey 2001; Krysan and Farley 2002) show the relevance of positive and negative attitudes of ethnic groups toward each other in determining their preferences for neighbourhoods.

1.4 Effects and mechanisms of residential segregation

A burgeoning literature has addressed the relationship between various neighbourhood conditions and individual outcomes in several domains. Small neighbourhood

effects have been highlighted on the intellectual development of very young children, with a positive effect of the presence of affluent neighbours (Duncan, Brooks-Gunn, and Klebanov 1994). Effects have been found also on later life stages, especially adulthood. It seems that neighbourhood composition affects health conditions (Pickett and Pearl 2001; Reardon and Firebaugh 2002), the access to welfare provisions (Osterman 1991) and to the labour market (Massey, Gross, and Eggers 1991; Pinkster 2014), although there is scant evidence. The existing literature points to small effects of neighbourhood composition, however the identification of these effects is characterized by several methodological issues (proper definition of neighbourhood, limits in segregation measures used, etc.) such that it is not yet possible to argue in favour of their presence or absence. Although the effects of neighbourhood environment are found to be significant in many studies, they are consistently much smaller than the effects of family characteristics (Ellen and Turner 1997).

Although the relation between neighbourhood composition and individual outcomes has been addressed, less efforts have been put in uncovering the black box of neighbourhood mechanisms. In general, authors have shown little consensus around the causal pathways underlying the neighbourhood effects (Leventhal and Brooks-Gunn 2000). Some scholars distinguish between two main groups of mechanisms: those related to the neighbourhood composition and those related with the spatial distribution of resources. The former stem from the demographic composition of neighbourhoods; e.g. poverty rates, average educational attainment levels, and the proportion of single-parent families. In this group, socialization is the key factor. Mayer and Jencks (1989) identify four main dynamics. The first is the so called *contagion model*, in which socialization with bad or good attitudes and behaviours will push children toward being more sympathetic with attitudes and behaviours. According to this model, children reared in low income neighbourhoods will behave worse than those raised in high income contexts (Mayer and Jencks 1989). In the second dynamic, advantaged neighbours are a disadvantage. According to this view, if living with privileged neighbours creates resentment then low income individuals may feel the need to build a "deviant" subculture. Third, neighbourhoods, through institutions and different resources, matters more than neighbours. This mechanism is very close to the spatial distribution of resources. The idea underlying this dynamic is that it is fair to expect the local tax base to influence the quality of resources and institutions present in the area. Income segregation therefore may create disparities in these public resources and institutions among communities. However, authors that provide this explanation usually recur to black box explanations that do not address how the general spatial distribution of resources may influence directly the individual. A more fine grained review of the several possible mechanisms of neighborhood effects by Galster (2012) identifies 15 processes. The author sums them up in four categories. Social-interactive mechanisms, a set of factors that refers to social processes endogenous to neighbourhoods. This category includes the social contagion model previously discussed, along with social networks, relative deprivation and parental

mediation. Environmental mechanisms refer to attributes of the local space (both natural and human-made) that may affect mental or physical health of residents but not their behaviour. These mechanisms involve exposure to pollution, environments in deteriorated conditions and violence. Related to this category are geographical mechanisms. This set of factors refers to aspects of spaces that may affect residents' life courses yet do not arise within the neighbourhood but because of the neighbourhood's location (Galster 2012). Finally, institutional mechanisms involve actions by agents not residing in the neighbourhood but with a certain level of resources in that area, e.g. local institutional resources.

1.5 School segregation across countries

As in the case of residential segregation, the first steps towards the study of school segregation were taken by American scholars. This issue has been at the centre of the political debate and on top of the academic agenda of sociologists and economists since the famous *Brown vs Board of Education of Topeka* case (1954), in which the Court declared state laws establishing separate public schools for black and white students to be unconstitutional. Research soon focused on the effects of school segregation on individual outcomes, behaviours and attitudes. Thus, the study of school segregation is usually linked to its effects on school performance, attitudes towards school and study, probabilities to fulfill the academic path. This section contains a review of the literature on school segregation, while studies that address its causes and effects will be discussed in the next section.

1.5.1 United States

The landmark *Equality of Educational Opportunities* report (Coleman et al. 1966) included the first systematic study on school segregation in American schools. The authors described schools, especially public ones, as a heavily segregated environment. In fact, almost 80 percent of all white pupils in 1^{st} and 12^{th} grade attended schools that were from 90 to 100 percent white, while the same figure for black students was 65 percent (Coleman et al. 1966). At the geographical level, the presence of segregated schools was found to be higher in metropolitan areas and southern regions. The authors identified school as a segregated environment also for teachers, with a predominant presence of white teachers in majority white schools (and vice-versa). The elimination of the last vestiges of *de jure* segregation, and the implementation of desegregation policies, led to a reduction in school segregation levels in the 1960s and 1970s. Desegregation policies were aimed at eliminating dual school systems, in which one set of attendance zones was used to assign white children to one set of schools, and a second set of zones was used to assign black students to a different set of schools (Coleman, Kelly, and Moore 1975). The reduction of school segregation involved especially schools in the South, where the percentage of black students in

majority-white schools increased from 19 percent to 44 percent (Clotfelter 1998). Although the racial and ethnic diversity of the US in the '80s and '90s increased, and the population substantially concentrated in metropolitan areas (Littman 1998), the average segregation measures for school population remained substantially stable (Reardon, Yun, and McNulty Eitle 2000). This stability masks two distinct trends: groups other than white have become less segregated from each other, while segregation between white students and black, Hispanic and Asian students has been on the rise (Reardon, Yun, and McNulty Eitle 2000). In fact, according to the authors, about 80 percent of school segregation in American metropolitan areas is due to white segregation. Of this amount, one-quarter is due to within-district segregation while the remaining part is caused by between-district segregation resulting by residential patterns. While there is clear evidence that school segregation between blacks and whites declined between 1968 and mid-1970s (Reardon and Owens 2014), trends since the late 1980s are less clear. The difficulty in identifying the direction of these trends relies on the measures of segregation used. Scholars that support the *resegregation thesis* for black students (Orfield 2001; Frankenberg and Lee 2002) recur to exposure and isolation indices, while those supporting the *stability thesis* (Logan, Stults, and Farley 2004) use indices that measure the unevenness of distribution of groups between schools and districts.

1.5.2 Europe

There is rising interest in school segregation among researchers in European universities, although still less than the attention provided by American scholars. A literature is developing around some national cases, especially France and England, countries with a long tradition in inward migration flows.

A recent report by the French minister of education has investigated the levels of segregation by socio-economic characteristics in the *collèges*[6] and *lycées*[7]. Givord et al. (2016) have found high levels of social segregation in French lower secondary schools. By ranking the *collèges* by their composition, the top ten percent enrolls less than 15 percent of all students with low socio-economic origin while the bottom ten percent enrolls more than 60 percent. However, the main trend in segregation can be better understood by looking at the division of students between public and private schools (Oberti 2007b). Since 1989, the share of high socio-economic status students in private school is grown up to 37 percent in 2015, while that for low socio-economic status students deflated to 20 percent in the same year. The mechanism of segregation is in place also for the upper secondary school, as Fack and Genet (2016) have shown using the case of Paris.

Levels of segregation have been studied also for the English case, especially among different ethnic groups. Johnston, Wilson, and Burgess (2004) have investigated the

6. Lower secondary school.
7. Upper secondary school.

extent of ethnic school segregation in English secondary schools in 2001. Their analysis highlights that secondary school is substantially segregated on ethnic criteria, with white residents attending predominantly white schools. Half of all non-white secondary students are found to attend a school with more than 75 percent of white enrolment, and most segregation is registered within the state sector (Jenkins, Micklewright, and Schnepf 2008). A subsequent work by Burgess, Wilson, and Lupton (2005) confirmed Johnston *et al.*'s results showing higher levels of school segregation for South Asian pupils than for black ones. Finally, there is only one work that compares levels of segregation in secondary school at the international level (Jenkins, Micklewright, and Schnepf 2008). In this paper the authors compare 27 OECD countries on socio-economic segregation and tests the reliability of two different means to measure segregation. The results highlight some countries of continental Europe (Austria, Belgium, Germany and Hungary) as the most segregated secondary school systems, while Nordic countries have the less segregated systems (Norway, Sweden, Denmark and Finland). England and US occupy middle-ranking positions.

1.5.3 Italy

To the author's knowledge there is scant research on levels of school segregation in Italy. There are three works available that investigate to some extent school segregation by ethnicity in lower secondary schools. Due to data limitations these contributions are not able to provide a fine distinction between ethnicities other than "natives vs foreign born". The first work (Torri and Vitale 2009) is an investigation of the mechanisms (school segregation included) of reproduction of poverty inside one of the most deprived neighbourhood in Milan, Quarto Oggiaro. The authors do not provide any figure for indices of segregation in Quarto Oggiaro as they focus on highlighting the process that makes the neighbourhood a segregated one. The second work is a recent investigation by Pacchi and Ranci (2017). In their inquiry on school segregation in primary and lower secondary in the Milanese municipality, the authors concentrate their efforts in clarifying the complex structure of school segregation in Milan. They find a situation of mild segregation on ethnic and socio-economic bases. Moreover, they highlight the role of school choice in redefining school composition across the city. Despite their analysis of flows of pupils between school districts, the authors do not consider the whole metropolitan area of Milan. This is a major limitation as it doesn't allow to investigate flows of pupils between the core and the adjacent municipalities. The third work (Barberis and Violante 2013) addresses this issue and includes a comparison between four metropolitan areas: Milan, Bologna, Rome and Naples. The authors provide segregation indices for the selected cases showing mild levels of school segregation on ethnic bases. The more segregated metropolitan area is Naples, followed by Milan, Rome and Bologna.

1.6 Causes and mechanisms of school segregation

As for urban segregation, causes of school segregation can be grouped in the following three categories: economic factors, institutional factors, parental preferences.

The first group includes factors that are related to the economy, such as the residential position of households and demographic trends, that shape the composition of neighbourhoods and districts. In the last 30 years the spatial distribution of social classes and ethnic groups have changed according to the main local trends. Population on both sides of the Atlantic Ocean has become increasingly racially and ethnically diverse with a strong concentration in metropolitan areas and large cities. Neighbourhoods in old European cities are experiencing processes of gentrification characterized by an increasing presence of upper classes in previously majority-working class areas and several economic crisis have reshaped the distribution of economic resources among families. Schools are experiencing the consequences of these demographic and economic changes in their reference population.

However, levels of residential segregation are not enough *per se* in explaining school segregation but must be combined with institutional factors. The most relevant element of this group is the architecture of rules that define how to participate to the school system. There are two different systems depending on whether enrolment is on district base (United States, France) or family choice (Italy, Netherlands). Given the interplay between these two factors, the numbered contributions that have investigated this issue in the US share the general result that school composition reflects district composition (Farley and Tauber 1968; Farley and Tauber 1974; Reardon and Yun 2005; Frankenberg 2013). The situation is different in Europe. An analysis of school segregation in the metropolitan area of Copenhagen by Rangvid (2007) have highlighted that low residential segregation does not necessarily translate into low or moderate school segregation. Similar results are available for Amsterdam (Gramberg 1998) and Italy (Barberis and Violante 2013). In systems in which enrolment is on district base, students attending a specific school are drawn from the area close to that school, thus reproducing neighbourhood composition. Dissimilarities may be observed as a consequence of other institutional factors such as desegregation policies (Welch 1987) in the US and derogation systems (Oberti 2007a; Oberti, Préteceille, and Rivière 2012) in France. Moreover, schools may be able to exercise some degree of influence in the spatial positioning of individuals (Frankenberg 2013). This may happens through two mechanisms: schools can act as signal to home seekers influencing their preferences in the house market (Pearche 1980); a perpetuation effect may be in place such that graduates from integrated schools are more likely to live in more diverse neighbourhoods rather than their peers graduated from schools where the student population has higher degrees of homogeneity with respect to their socioeconomic characteristics (Phillips et al. 2009).

Finally, preferences of parents and children. School characteristics such as organization, reputation, extra activities may shape these preferences. A study on school

choice in a major Dutch city (Ledoux, Koopman, and Schaap 1999) has shown that school composition is not a factor in choosing a school but a factor in *not* choosing it. Karsten et al. (2003) have surveyed Dutch parents with respect to their ethnic preferences in choosing a primary school, finding that ethnic composition is not a relevant motive for school choice of natives, that rank reputation, organization and proximity to home as more important, but is relevant to foreign-born parents. In a similar study on Essen primary school system (Germany), Kristen (2008) has investigated whether variation in school selection varies because of differences in distribution of relevant resources or different perceptions in attractiveness or individual preferences for segregation. As a main result, the author find that family resources account for substantive differences in school choices between ethnic groups.

1.7 The effects of residential and school segregation on educational outcomes

As for literature on residential segregation, research on effects of school segregation on academic performance provides a number of studies on the US and little research on European countries. This growing share of contributions rely on the assumption that individual outcomes, such as school achievement, are not only the result of personal characteristics but also that of peers' achievement and behavioural patterns.

A first strand of literature investigates the effects of living in a poor neighbourhood on educational achievement, finding mixed evidence. Teenagers who live in high SES neighbourhoods attain more schooling than those from similar families who live in lower SES neighbourhoods (Datcher 1982). This study shows that quality of the neighbourhood is pivotal in explaining part (about one quarter) of the differences in schooling and earnings between whites and blacks in the US. However, the classical black-box problem arise as the author does not address the mechanisms underlying the relationship found. The literature review by Mayer and Jencks (1989) suggests that growing up in a high SES neighbourhood may lead high school graduates to get more education, even though attending a high SES high school does not. Moreover, given the methodological difficulties in clearly identifying effects of neighbourhood composition, it is possible that the actual estimates are over estimating the effect of neighbourhood composition on school achievement (Mayer and Jencks 1989), a conclusion shared also by Solon, Page, and Duncan (2000). Using data from the PSID they found that the proportion of the variance in educational attainment that can be ascribed to neighbourhood effects, after controlling for individual and family characteristics, is about 10 percent, far less than Datcher's results. Given the starting conditions, even complete equalization of neighbourhood backgrounds would leave inequality in educational attainment rather unchanged.

In order to gain better evidence on neighbourhood effects on educational outcomes, some researchers recurred to experimental approaches. The most relevant case is Ludwig, Duncan, and Ladd's (2001) inquiry on the effects on educational

outcomes of young children and adolescents of the Moving to Opportunity for Fair Housing program in Baltimore. The MTO was a residential-mobility program operating since 1994 in Los Angeles, New York, Baltimore, Boston and Chicago. The program was sponsored by the United States Department of Housing and Urban Development and organized as a randomized social experiment. Low-income families living in public housing in high-poverty zones who volunteered for the program were randomly assigned into one of three treatment groups: the *experimental group* received rental subsidies for private-market housing to be used in census tracts with very low-poverty rates for the first year. After a year, they could use their vouchers anywhere. The *comparison group* was offered rental subsidies but with no restriction to zones in which they can be used. The *control group* received no rental subsidies. The analysis of Ludwig, Duncan, and Ladd (2001) show an increase of educational outcomes (grades and test scores) for young children (up to 12 years-old) of families that moved to low-poverty neighbourhoods while there are no significant differences between the treated and the control group of young adolescents (older than 12 years-old). Although efforts put in disentangling neighbourhood effects, the relation with individual educational outcomes remains unclear. A later analysis of long-term effects of the MTO program (Ludwig et al. 2013) sheds different light on previous results, showing that there were no consistent effects on adult economic self-sufficiency or children's educational achievement. However, positive effects have been found on several mental and physical health indicators.

A second strand of literature investigates the effects of school composition on educational performance. This is usually measured through standardized test, a tool for assessing educational competences and abilities. Before entering the review of the relevant research on the topic, I present a brief introduction to standardized tests.

Standardized tests can be used in summative or formative evaluation frameworks. The two processes differ in how they seek and use information (Scriven 1967). Summative evaluation has a focus on outcomes and aims to compare individual results with a pre-defined benchmark, formative evaluation aims at obtaining information in order to improve the process investigated. Standardized tests can be used in these two ways, they can evaluate the effects of certain educational practices, policies or processes (summative) or can be used to highlight students' learning difficulties and inform the design of specific actions to tackle the problem (formative). An example of the latter case is represented by Nye and Konstantopoulos (2000) and Argentin, Romano, and Martini (2012) studies. The former analyses the effects of class size on academic achievement in Tennessee, the latter used a controlled experiment on a sample on thirty Italian primary schools to investigate the effects of playing chess as an extra-curricular program on learning mathematics.

At the international level, the most famous example of such a tool is the PISA program sponsored and organized by the OECD since 2000. PISA aims at investigating the ability of secondary school's pupils in reading, math and science. The history of

this program is rooted in a tradition of international school studies undertaken since the late 1950s. Among the most relevant precursors of PISA were the TIMSS program, a comparative assessment of students' achievement in mathematics and science for students in fourth and eighth grade, and the PIRLS assessment in reading involving only fourth grade students. Both these tests are policy-oriented instruments run by the International Association for the Evaluation of Educational Achievement. In recent years the OECD has developed the PIAAC program, a survey meant to measure adults' proficiency in key information-processing skills-literacy, numeracy and problem solving in technology-rich environments. However, PIAAC is the third survey of its kind. In fact, since the early 1990s the need for assessing literary skills in developed countries has been addressed by two large international surveys. The first was the IALS, the first program to provide and international comparative assessment of adult skills implemented in 1994, 1996, and 1998. The second was the IALLSS carried out in 2003, and between 2006 and 2008.

At national level, both private (as in US) and governmental (as in the UK and Italy) agencies have developed assessment tests for monitoring the educational system or selecting pupils in accessing the highest levels of education. In the US, one of the most relevant example is represented by the SAT score. This test is administered since 1926 by the College Board, a non-profit organization. The SAT test was originally not aligned to any particular curriculum, school of any order could adopt it for any purpose but in 1942 the SAT test was used to assess readiness of students for college (Baird 2012). The other relevant example is represented by the ACT test, developed in 1959 by Professor Lindquist at University of Iowa as a competitor of SAT. The ACT test is used in the process of selection of students at high school and university and includes five different sections: English, math, reading, science reasoning and writing. In the UK, since 1991, the Department of Education introduced the National Curriculum Assessment (commonly referred to as Statutory Assessment Test) for the purpose of testing seven-, eleven- and fourteen-year-old pupils' educational performance in English and mathematics. In recent years the test has gone through several criticisms by teaching unions, also a recent parliamentary report (Parliament 2008) has addressed the main shortcomings of the SAT. The main concern was that *teaching to the test* practices may emerge and thus narrowing the curriculum and focusing disproportionate resources on borderline pupils. Something similar to the British case has been done by the Italian Ministry of Education through INVALSI, a governmental agency with the purpose to evaluate and foster evaluative practices in the educational system. Since 2004, each year INVALSI runs a standardized test on Italian and mathematics involving pupils of grade two, five, eight and ten.

1.7.1 United States

A great contribution to highlight the relationship between class or school composition and results of standardized tests in the US has been given by economists Hanushek 1972; Hanushek, Kain, and Rivkin 2009. In these works they investigated the effects

of racial composition on test scores using Texas' schools. After controlling for family background, the results show that black enrolment shares adversely affects achievement of both black and white students and that the penalty is twice as large for blacks as for whites. Similar results have been provided by Bankston III and Caldas (1996) for Louisiana's schools, while Vigdor and Nechyba (2007) provided evidence of the effects of peer characteristics on student achievement in North Carolina elementary schools. As a general result, minority concentration exercises a negative impact independently of behavioural patterns and socioeconomic composition of schools, although this effect is non linear. Hoxby (2000) has investigated the effect of both ethnic and sexual composition of classrooms in the US, finding evidence of a positive effect of the share of females in class and confirming previous evidence of the negative effect on test scores of high proportion of black students.

The effect of racial composition of schools have been investigated also by looking at the effects of desegregation policies. A review of the literature on this theme by Crain and Mahard (1978) counted over one hundred studies on the effect of desegregation practices on students' achievement measured through test performance in a time span of ten years. Despite the number in studies aimed at clarifying this particular relationship, the authors suggest the presence of a positive but small effect of desegregation policies on average gain in performance for blacks in the US. Mixed evidence on both achievement and psychological differences among blacks and whites is also the result of Schofield (1995) in his literature review on desegregation effects in primary and secondary schools and of Simmons et al. (1978). The key element in evaluating effects of desegregation is that these practices can't be considered as laboratory-controlled experiments identical across different settings. The uniqueness of each policy plan and the context in which it is applied is determinant in shaping the results.

While evidence for effects of school racial composition is mixed, that for socio-economic composition is more clear (Rumberger and Palardy 2005a; Palardy 2013). Its effects may differ for attainment outcomes as compared with achievement outcomes due to different mechanisms at work (Rumberger and Palardy 2005b). The average socio-economic background of peers is likely to substantially influence school achievement (Kahlenberg 2001; Hanushek et al. 2003). The composition effect may not be the same across ethnic or socio-economic groups. In their review, Mayer and Jencks (1989) suggest that the average school composition may affect more black students than whites and higher-status students than lower-status ones. With respect to attainment, studies have focused on the effects of school composition on college enrolment (Perna and Titus 2005) and dropout (McNeal 1997; Rumberger and Palardy 2005b). While a positive relation has been highlighted for college enrolment (Engberg and Wolniak 2010), no effects have been found for dropout rates (McNeal 1997).

1.7.2 Europe

Coming to Europe, literature provides few cases. Ammermueller and Pischke (2009) found socio-economic composition of classrooms to be associated with PIRLS scores in grade four of six countries (Germany, France, Iceland, Sweden, Norway, the Netherlands). Using PISA data Brunello and Rocco (2013) investigate this association for ethnic composition in secondary school in cross country comparison. They identify a negative correlation between higher shares of immigrant pupils and achievement tests of natives. However, the effect is small and varies between socio economic status: natives with a disadvantage socio-economic background suffer more than their advantaged-background class mates, a result in line with Mayer and Jencks's (1989) prediction.

Using Swedish data, Szulnik and Jonsson (2007) found a small and negative effect of schools' ethnic composition on grades. Mixed evidence is available for France. Felouzis (2003)) estimates the gap in students' performance due to concentration of students with immigrant background in about 0.4 points on a scale ranging from 0 to 20. Using French secondary school data Cebolla-Boado (2007) finds no effect of the share of foreigners on various educational outcomes in lower secondary schools. However, estimates vary by the estimation technique used and caution is suggested in interpreting these results. Similar evidence is available for primary school in Spain (Cebolla-Boado and Medina 2011). Although all these studies show that once social individual characteristics are controlled for the average effect of ethnic classroom composition disappears, it is possible that the effect persists given a certain threshold in ethnic composition. In Cebolla-Boado and Medina's (2011) study the effect is significant only if immigrants represent at least one fifth of the student body. Coming to Italy, Contini (2013) has investigated the effect of the concentration of students with immigrant origin on student learning in primary and lower secondary schools using 2010 INVALSI data. Consistently with previous literature, the author identifies a weak negative effect of the concentration of ethnic minorities on educational performance of pupils. Moreover, when considering classrooms, the concentration of first generation immigrants has a larger impact on students of immigrant origin than on natives. This is especially true for the Italian test. Among natives, those who are affected the most are those from the lowest socio-economic background. Cardone, Falzetti, and Severoni (2015) run a similar analysis on pupils enrolled in grade 5 (last year of primary education) in 2012. They show that students enrolled in schools with segregated classes[8] perform worse than peers in not (or less) segregated classes. The gap is pretty the same across the distribution of ESCS, thus high ESCS stuents are affected as much as low ESCS peers. Interestingly, this result is not in line with previous research and may be a specificity of Italian primary school, or being the consequence of some degree of model misspecification. For example, the model presented in this

8. The authors identify segregated schools as those in which the between-classes variability of the index that measures socio-economic and cultural background (ESCS) is higher than the national mean.

study accounts only on structural characteristics of schools and does not control for individual characteristics of pupils.

1.7.3 The mechanism: peer effects

Researchers have highlighted two competing mechanisms to explain variability in the effect of school composition on the educational performance of pupils. The first mechanism is related to endogenous characteristic of schools. Schools in which pupils with immigrant and disadvantaged social background are more represented differ significantly to other schools in their structural characteristics (Cebolla-Boado and Medina 2011). In general terms, schools in poor neighbourhoods may be as deprived as the surrounding environment, thus negatively influencing results of pupils. However, there is little and mixed evidence about the strengh of this argument. Similarly to school effects, some authors have suggested the importance of class-level effects. In this case, teachers might adapt their requirements and demands to the average level of pupils (Duru Bellat and Mingat 1997).

The second mechanism is *peer to peer* interaction, or the social interaction between individuals in the same environment. Let "A" and "B" be two pupils with a given set of educational resources. If having student "B" as a classmate/schoolmate affects performance or educational behaviour of student "A", this can be considered a peer effect. Thus, this definition of peer effect includes any externality that spills over from peers or peers family background. Mansi (1993) distinguishes between endogenous peer effects (those that derive from peers' outcomes) and exogenous peer effects (those that derive from peers' background). Researchers have traditionally encountered difficulties in the proper identification, or estimation, of peer effect due to the difficulty to disentangle it from family and school characteristics (Hanushek et al. 2003). Some scholars have suggested to use panel data (Hanushek et al. 2003) while others support recurring to instrumental variable approaches (Card and Rothstein 2007). A third group of researchers have opened a third way: developing specific peer effect models in the attempt to capture the different mechanisms through which interaction between individuals can shape educational outcomes. Different models have been developed in order to capture specific transmission channels of peer effects. Hoxby and Weingarth (2005) provide a categorization of these mechanisms related to econometric models. Three groups can be identified. The first includes the most common model and explanation for peer effects: linear-in-means models. It suggests that only average characteristics of peers matter. To this group belong also models of class homogeneity and heterogeneity. This last specification, also known as "rainbow model", suggests that diversity of ability is good for all students. While theoretically plausible, this specification is not supported by empirical evidence. The second category groups two models, one the opposite of the other. On one side, a mechanism in which a disruptive student harms everyone. This is the so called "bad apple" mechanism, in which a student provide negative externalities to the whole classroom. Other students may be affected differently according to the

ways the disruption manifests. Moreover, students with different backgrounds may suffer peer effect in different magnitude. On the other side, one excellent student may provide positive externalities to the whole classroom. Empirical evidence and theoretical conceptualization lack on this specification (Lazear 2001). Finally, the third category includes mechanisms in which students perform better if surrounded by others like themselves. This mechanism is the usual justification for class tracking (Sacerdote 2011). A counter-intuitive consequence of this explanation is that the effect is non-linear across student ability: the less able pupils gain more by the presence of similar students than by the presence of high ability peers.

1.8 The nature of effects of segregation

A vast literature has shown that effects of ethnic and socioeconomic segregation may appear in different domains of human life. Through several mechanisms, segregation can affect mental health, educational outcomes, behaviours and attitudes, access to the financial market etc. Given the variety of domains in which these effects appear, we can think of them in terms of spillover effects. A *spillover* can be defined as a situation that starts in one place but then begins to happen or have an effect somewhere else (*Collins COBUILD - English Dictionary* 2003). Spillover effects of segregation have been extensively studied in the literature, but little attention and efforts have been put in studying the structure of these effects over time.

The sociological literature usually refers to the concept of cumulative advantage and disadvantage (CAD) in order to address the accumulation of any sort of advantage/disadvantage over time to individuals or groups. This concept was originally developed by Merton (1973) with reference to scientific reputation and productivity. Merton defined cumulative advantage as "the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable reputate and the withholding of such recognition from scientists who have not yet made their mark" (Merton 1973) and named this mechanism *Matthew Effect*, from the parable of the talents in the Gospel of Matthew. In other words, cumulative advantage is the mechanism through which an initial comparative advantage/disadvantage of trained capacity, structural location or available resource makes for successive increments of advantage/disadvantage (Merton 1988). These definitions contain three fundamental elements: a CAD process is capable of magnifying those that are very small differences at the begin of the process; the advantage must accumulate over time; it defines inequality in the population as an increasing function of CAD.

DiPrete and Eirich (2006) consider Merton's definition as the strict conceptualization of CAD and identify a second strain of theorizing rooted in Blau and Duncan's (1967) *The American Occupational Structure*. Blau and Duncan model refers to persisting direct and indirect effects of a status variable on individual outcomes, and attributes inter-group inequality to differences in returns to socioeconomic resources

(Rugh, Albright, and Massey 2015). A status variable produces a cumulative disadvantage when it has both direct and indirect effects on outcomes in life-course perspective (i.e. they consider being black as a cumulative disadvantage because race had both direct and indirect effects on individual outcomes). Differently from the strict "Mertonian" conceptualization of cumulative advantage that requires growing differences between individuals, or growing mean differences between groups, over time, the Blau and Duncan model requires only persistent effects over time. This sufficient condition leads to the corollary that cumulative disadvantage does not necessarily lead to growing inequality in the population.

As suggested by DiPrete and Eirich (2006), the Blau and Duncan model can be generalized by widening the concept of status variable. Status variables attached to individuals (race, social position, educational attainment, etc.) can be thought of in experimental terms as the exposure to their effects over time, that is the exposure to a treatment. Thus, we can identify two groups of status variables: those attached to the individual; those that are not possessed and to which the individual is exposed over some duration. In other terms, the second group represent exposition to other individuals and their characteristics. Despite DiPrete and Eirich (2006) do not discuss the idea of exposure underlying their generalization of Blau and Duncan's model, a specification of this concept is needed. The idea is that the concentration of a given status variable in a social environment actively influences an individual and its outcomes and life chances. We can think of exposure as a situation similar to exposure to a virus or a chemical agent. The concept can be enhanced by framing it in different terms. Let consider each individual as a vehicle of a status variable. In social environments characterized by an high concentration of people with a given status variable the probabilities to enter in contact with individuals carrying that variable, or the probabilities "to be infected", are high. In this conceptualization, being exposed to a status variable means to socialize in that social environment. Growing in a rich family means to socialize with high status individuals, growing in a black neighbourhood means to socialize predominantly with black individuals, etc.

In Blau and Duncan terms, school and neighbourhood segregation can be defined as cumulative advantage processes when growth rates of an outcome variable vary by some status, and the status-unequal growth rates persist over time or across multiple stages of the life course (DiPrete and Eirich 2006).

1.9 Conclusions

This first chapter reviews literature on residential and school segregation, its causes and main effects in western societies. It starts from a discussion on the concept of segregation, highlighting the contribution of leading scholars in producing a feasible and sound definition of segregation that can be applied to several phenomenon without any restriction. The first section is concluded by a more comprehensive definition of segregation that builds on the contributions of Kuper (1968), Schelling (1971),

Massey and Denton (1988) and Reardon and O'Sullivan (2004). Three are the key points in this definition. First, segregation is the manifestation of social distance between groups. Second, segregation appears in several ways and in different degrees. It follows that it is not necessary to have a perfectly physically separated space to have a segregated environment. Third, segregation applies to any social environment, that is to any environment in which there are social interactions.

The chapter continues in three parts. The first part contains the review of literature on residential segregation by ethnicity and social status. The contributions presented highlight a huge body of literature developed by American scholars starting from the Chicago School tradition, while there is little research on Europe and Italy. As a general result, residential segregation by ethnicity is high in American cities, where ethnic minorities are confined in the city centre while white Americans are clustered in the suburbs. Arbaci (2007) has shown that levels the spatial distribution of ethnic minorities varies across Europe. Northern cities are more similar to American ones, with high levels of segregation and spatial location of minorities in city centre, southern cities are characterized by lower levels of ethnic segregation and spatial disposition in peri-urban areas. The Italian case has been little investigated on the issue of residential segregation. Some evidence is provided especially for large metropolitan aggregates (Rome, Milan, Naples) but little is known for middle-size cities. However results shown little average segregation levels with few clustering processes in urban peripheries. Residential segregation by socio-economic characteristics is little studied, however the few contributions available highlight a general low level of spatial clustering due to this attribute.

The second part investigates the actual levels of school segregation and its causes. As for residential segregation, the US are characterized by high levels of ethnic and socio-economic segregation in schools. Scholars share the idea that this is due to how district composition reflects in schools composition. With reference to school segregation, the American case is a perfect example of how institutional factors can produce segregated environments. Conversely, in Europe there is not a strong correspondence between districts and school composition as enrolment on district base is less present (France, England). In general terms, schools are more segregated than related districts. This is mainly due to familiar strategies in school choice. Contributions on Italy (Torri and Vitale 2009; Pacchi and Ranci 2017), Germany (Kristen 2008) and France (Oberti 2007b) have shown substantial flows of pupils between school, as a result of the strategic behaviour of families that want to secure better education to their offspring.

The third part contains the reviewed of literature on effects of school and class composition on educational performances measured through standardized tests. The body of research on this theme has highlighted a small negative average effect of peers with immigrant background on educational performance, but a consistent effect of socio-economic composition. This effect varies across different social groups:

disadvantaged-background pupils suffer more than their advantage class mates. Moreover, as found by Cebolla-Boado and Medina (2011), effects may vary upon certain thresholds in average composition of schools or classes. The mechanism that can explain these composites effects is peer to peer interaction (or socialization). Let "A" and "B" be two pupils with a given set of educational resources. If having student "B" as a classmate/schoolmate affects performance or educational behaviour of student "A", this can be considered a peer effect. Evidence on this mechanism is mixed and requires more methodological investigation. However, the current state of the art shows that segregated environments may lead to several consequences (even small) in different domains thus influencing later life chances.

Chapter 2

Explaining school segregation: a discussion

Social closure dynamics and their determinants have not been addressed sufficiently in literature about school segregation. Available contributions usually see school segregation as a natural phenomenon and address "white flight" as the main "natural" cause. However, they do not consider enough the role of the schooling system, how it is perceived by ethnic and social groups, its structure and organization. According to Collins (2000), schooling in modern capitalist societies is characterized by a bureaucratic and hierarchical structure, where educational institutions grant formal degrees as certification of possession of some knowledge. The possession of the highest degrees allows economic returns in the labour market (*via* premium wages or access to professions), but represents also a *status*, the key to access the highest positions in the social structure. Given the advantages deriving from the possession of educational titles, along with the mass structure of schooling systems of modern capitalist societies and the scarcity in the supply of high-level degrees, families and pupils compete to gain access to educational paths and institutions that grant this asset. Social closure dynamics between groups may emerge as a main consequence of competition, thus determining segregation of groups between educational paths, schools, classes.

This chapter relies on the idea that competition in the schooling system is the main driver of segregation, and social closure dynamics are the mechanism through which this happens. The chapter discusses theories that may be helpful in understanding *why* competition arises. The discussion starts by considering the role of institutional actors. Section 2.1 presents a theoretical discussion on the institutional factors that allow for the appearance of competition in the schooling system, in particular those related with the degree of differentiation and autonomy of schools. Section 2.2 elaborates on the link between social stratification and social closure dynamics. The section presents four different theories that might be helpful in explaining why schools and educational tracks are segregated. The first is the Neo-Ricardian structuralist theory by Sørensen (2002). This line of reasoning is rooted in the idea that rents are the cause for antagonistic interests and conflicts in the society. These conflicts happen between owners and non owners of rents leading to strategies of social closure and

usurpation. The issue is whether education can be considered a rent-producing asset or not. The second is the Relative Risk Aversion model of educational transitions proposed by Breen and Goldthorpe (1997). The third is the Effectively Maintained Inequality theory by Lucas (2001), while the fourth is the credential theory by Collins (1979, 2000). Finally, Section 2.3 presents the research questions of this work.

2.1 Institutions, competition and segregation

Scant research has investigated the profound causes of school segregation, nor particular attention has been paid to the role of institutional actors although a vast literature has been devoted to address the phenomenon of school segregation and its effects. Despite its relevance, the way in which institutions regulate social interactions of actors in the schooling system, and its possible link with school segregation, is still a not addressed issue. Fiel (2015) suggests that institutional factors related to the degree of decentralization of the educational system are the most important in determining the occurrence of segregation. With *degree of decentralization* the author refers to a continuum between central organization, in which all the aspects of the system are defined by a central authority, and local autonomy, in which the organization is demanded to a peripheral authority.

However, it is difficult to define a rank of relevance between factors that participate in the complex process that leads to school segregation, and other factors may be considered as relevant as decentralization. It is the case of *differentiation*, a concept that addresses the variety of a given provision. School differentiation expresses the degree of variety of the educational provision: on one side of the continuum there is homogeneity, that's the case when each school of a given grade have the same educational plan, schedule, subjects, activities; on the other side there is heterogeneity. Differentiation can be achieved through policies that promote the local autonomy of schools and their active action in responding to the educational needs of the area in which the school is located. Differentiation and decentralization are surely linked, Collins (1979) suggests that decentralization causes differentiation, as the transfer of power from the center to a peripheral unit favors competition. However, also policies from a central authority can fuel competition between schools (i.e. tracking or curriculum differentiation) and families. In theoretical terms, the homogeneity-heterogeneity continuum can be applied to each dimension of the educational systems. Starting from whatever position in the continuum, a small change in the degree of differentiation is the necessary and sufficient condition for the introduction of (additional) qualitative differences.

When policies that enhance differentiation are introduced, the educational supply looses its characteristic of standardization and become a scarce resource, where the scarcity resides in the fact that for each combination of school characteristics there is a fixed supply, but potentially unlimited demand (the upper limit is represented by the existing population). Access to scarce resources is usually determined by a

market-like mechanism, in which competition within the groups of buyers and sellers determines a price for the good or resource. As a consequence of competition, quality standards of the supplied good or resources might change upwards. Since the last decades of the 20th century the educational systems of western societies have undergone a process whose result has been the introduction of market principles. The theoretical basis for this kind of policies are rooted in the seminal work of American economist **Friedmann1955**. Introduction of market principles in education, especially in the form of free-choice, was an instance for improving United States public school system and to achieve freedom of individuals (families). The introduction of market principles was presented as a tool to reduce economic segregation among students in public schools, by creating access for lower-income families to schools outside their neighbourhood. This prediction relies on the assumption that school choice is available to all families, and that the choices are supported with public transportation and full information about school options. This argument has been subsequently defined as the "liberation model" (Archbald 2004), as school choice would liberate children of low-income families from their reference school if they choose to take advantage of it. It is worth noticing the strict correlation of this proposal with the specificities of the American educational system: enrolment on district basis and school fundings based on district tax base.

The adaptation of the market model to the education sector is based on several basic principles characteristic of a *quasi-market* (Dumay and Dupriez 2014). First, free choice among a set of alternatives. Families should be able to choose the preferred school and be free to change school. Second, the system should enjoy a high degree of autonomy. This should allow schools to make choices specifically adapted to the respective situation and to differentiate their provisions from that of competing schools (Dumay and Dupriez 2014). Third, founding implement rewarding systems such that low performing schools with respect of a defined output get less money in order to prompt them to enhance their educational practices.

How these policies participate in creating conditions for more segregation? The key point is that family free choice and school autonomy create conditions for actors to compete on the educational provision. The starting point of this line of reasoning is school autonomy, that leads to differentiation of the educational supply. This goal can be achieved by different means, e.g. the implementation of rewards in funding systems, a mechanism leading to differential distribution of resources and consequent differential opportunities in expenditures for structures, wages, extra-activities, or through organizational autonomy of schools on different levels (schedules, classes, subjects, extra-activities, etc.). All of these elements provide the conditions for competition between families, as they transform good schools in a scarce resource. When free choice is implemented, families and pupils are able to freely move among schools and competition arises.

When the number of competitors increases in relation to the profit span (Weber

1968), agents try to secure profitable resources through individual and collective action aimed at reducing access and controlling competition. This process of monopolization of social and economic opportunities takes the name of *social closure*, the mechanism through which competition translates into segregation. Neo-Weberians such as Parkin (1979) have stressed the relevance of social closure as the main driver of collective action, along with the role played by institutional factors in shaping closure dynamics. Parkin (1979) identifies two main types of social closure: exclusion and usurpation. Strategies of exclusion are the predominant mode of closure and refers to the attempt by one group/individuals to secure for itself/themselves a privileged position at the expenses of some other group/individuals. This practice leads to the definition of a group of excluded, or negatively privileged. Strategies of usurpation are a response to exclusion strategies, the attempt by excluded to win a greater share of resources (Parkin 1979), but also to challenge institutions and the resulting inequality.

Murphy (1984) suggests that social closure dynamics are derived by rules of closure and proposes a framework for the analysis of these rules, whose absence has been identified as a main limitation in Parkin's work. Exclusion in a social environment may be principal, derivative or contingent. The principal form of exclusion refers to the set of exclusion rules that are backed by the legal apparatus (laws). From the principal form are derived two other type of rules. Derivative rules of exclusion derive directly from the principal form, but are not identical. Derivative forms can be formally written into laws. Finally, contingent rules of exclusion are not directly derived from the principal form and depends on the context determined by it. For example, a principal form in capitalist societies is private property, from which it is possible to identify derivative and contingent forms of exclusion that arose around credentials, ethnicity, race, gender, etc. When coming to the educational system, the principal form of exclusion can be identified in the role of selecting individuals for apical position in the society. Derivative forms of exclusion are norms that determine enrolment on quota basis, tracking by ability, etc. Contingent forms of exclusion (determined by the context) are informal practices aimed at granting access to a given institution. The assumption that derivative and contingent forms of exclusion in education rely on the principal form is not violated. Unlikely exclusion, usurpation can be articulated in two distinct forms. Inclusionary usurpation, the struggle by the excluded to be included in the current structure of positions, and revolutionary usurpation, aimed at changing the structure of positions. We have inclusionary usurpation in education when certain groups struggle to have access to some given level of education. History of the 20^{th} century is dissaminated of examples of inclusionary usurpation. On the other side, home-schooling might be conceived as a form of revolutionary usurpation. Parents bringing out their offspring from the classical education system is a way to change the structure of positions in education.

Although policies aimed at introducing quasi-market mechanisms are the main cause of segregation, they are not the only one. A special case is represented by

policies of district definition. These policies refer to the definition of school districts, their implementation and relevance in the enrolment mechanics of school systems. With respect to a given educational level (primary education or secondary education), districts could be designed to contain only a school or several of them, they can be relevant in the enrolment process or not. When districts are relevant in the enrolment process and include a single school, school's population reflects district composition and segregated schools will appear in segregated districts. When districts are relevant in the enrolment process but include several schools so that families can choose, or they are irrelevant in the process of enrolment, school's population does not necessarily reflect district composition. Here, school segregation is also the result of family choices.

2.2 Social closure and stratification

Since the 1950s the schooling systems of western societies have experienced a process of expansion, also known as "massification". This expansion has involved both the numbers of people staying in education and the structure of the educational systems themselves. Educational systems have changed substantially: tertiary education has been characterized by a relevant process of stratification, while the school system has experienced a gradual destratification. It is the case of Italy, the selected case of this work. In the same period, several research contributions have investigated class differentials in educational attainment finding mixed results. Persistence over time of inequalities of educational opportunities in Italy have been found in early studies (Cobalti 1990; Cobalti and Schizerotto 1993), but consensus on this view was challenged by subsequent contributions (Ballarino et al. 2009). It is worth noticing that when both social position and parental education were introduced in the analysis, only inequalities of educational opportunities related to family education was found to be persistent (Ballarino and Schadee 2008). In order to explain the persistence of these differentials, several models have been proposed. This section presents three theories of macro-level educational inequalities, linking them to the micro-level phenomenon of segregation.

2.2.1 Neo-Ricardian structuralist approach

The aim of applying a Neo-Ricardian structuralist approach to education is to identify whether the possession of educational titles might grant a rent in the social structure. The solution is found in the intergenerational transmission of education, a situation in which access to education is dependent on family background characteristics instead of pupils' ability.

The starting point for any structuralist approach is the concept of class, that in the original proposal by Marx was aimed at providing a structural theory of inequality and building a theory of history. The cornerstone in Marx's work for class definition is exploitation, the process through which one class obtains economic advantages at

the expenses of other classes (Marx 1996). In the capitalist mode of production, the Bourgeois state defines through the property rights the legal framework of legitimate property relations. This process is based on legal property rights, that correspond to the capitalist mode of production. Marx defines production as a process of appropriation of nature by an individual within and through a specific form of society, stressing that the two concepts (production and property) are inseparable. Given how it is defined, the process of exploitation generates antagonistic interests between classes. Through the process of class formation, the incumbents of classes realize they have these interests and organize themselves in collective actors that engage in conflicts. Depending on the results of this struggle the class structure and society may change.

Sørensen (2002) has proposed to refine the causes of antagonistic interests in order to provide a theory of exploitation more consistent with modern economic theory and the organization of market economies. His suggestion is to widen the definition of property rights at the base of the class concept, in order to comprehend all assets and resources that produce economic benefits. According to Barzel (1997) property rights can be seen as the ability to receive the return on an asset, directly or indirectly through exchange. That is, the ability to consume or have access to a good, or the ability to have control on its supply. These rights may be supported by the state, and they are then legal rights, but people also obtain advantages from rights that are not legally enforceable (Sørensen 2002). Clearly, these rights are not constant in time as both individual and collective action are able to change their distribution in the society. However, the rearrangement of property rights is costly given that some attributes of assets are difficult to measure and not fully known to actual or potential owners (Barzel 1997). This difficulty translates into capture by others who then obtain rights to the benefits from this attributes.

The sum of advantages produced by the ownership of both assets and resources and those produced by not legally enforceable rights constitutes the person's total wealth. Moreover, part of this total wealth may come from rents, generated upon resources or assets that are fixed in their supply. Economic theory defines rents as the difference between the actual value of an asset and the value that would have emerged under competition. This conceptualization follows from the classical definition by Ricardo (1951), in which rent is related to ownership of land and defined as "that portion of the produce of the earth which is paid for the use of the original and indestructible power of the soil". In other words, a rent is a surplus earned by a factor over the sum necessary to induce it to do its work (Wessel 1967). Rents are then obtained independently by the effort of the owner of rent-producing asset and exist as long as owners of assets are able to exercise some control on the supply.

This conceptualization makes, for Sørensen, the ownership of rents the mechanism of exploitation through which defining the new class structure of a society. To sum up, in Marxist theory the concept of exploitation refers to labour relations and is the process through which agents are not receiving according to their work or needs (Elster 1978). This process is possible because there is asymmetry in the ownership

of productive factors, thus the class structure is dual: those who have the means of production and those who have not. In Sørensen's proposal, exploitation is not confined to labour relations but extended to all factors capable of producing a rent. Like in Marx, the process of exploitation is rooted in the asymmetry of ownership of these assets, unlike Marx, Sørensen does not derive a class structure according to its specification of the exploitation process.[1]

Theoretically, non owners are better off in two situations. First, as a group they gain in a perfectly competitive market where there are no rents. In fact, in this fictional situation there is no exploitation processes between classes and total wealth is improved by the absence of additional prices paid to access productive goods or resources. The second situation involve the upper social mobility of non-owners that happens when they are able to gain control over a rent-producing asset and exploit, even partially, its benefits. This situation is characterized by both vertical and horizontal competition in the social structure. The former is competition between actual owners and non-owners, the latter is that within the group of non-owners. Both kinds of competition have a collective and individual dimension. They can be the sole action of a single individual, that adopt strategies to defend its position and gaining (further) control on scarce resources producing rents, or collective action. It is important to notice that it is not possible to separate the two aspects of competition, as attempts of gaining control over a rent-producing asset are inherently competing on the vertical and horizontal dimensions.

The conflict in interests between those who have access to a rent and those who have not generates rent-seeking and rent-protecting activities that take the form of "class action". In accordance to the neo-weberian approach (Parkin 1979; Murphy 1984), Sørensen has stressed the relevance of social closure as the main driver of collective action, along with the role played by institutional factors in shaping closure dynamics. The modes of action can be summed up in two categories: on the one hand, processes of closure, aimed at reducing social and economic opportunities to outsiders of a given social group; on the other hand, processes of usurpation represented by the broad ensemble of actions by the excluded in order to challenge institutions and the resulting inequality. Closure theory suggests that collective action at the macro-level is fundamental in controlling entry and exit to key positions in the division of labour (occupations) and then to social classes. On the other side, Grusky and Sørensen (1998) argue that regulation of entry and exit to occupations happens at the local level. What is relevant to set this dispute is to firstly identify at which level the rent is generated, and interests are formed Sørensen (1996, 2002).

The theoretical approach discussed above has never been applied to a rigorous discussion on whether education can be considered a rent-producing asset in modern

1. It is worth noticing that it is far beyond the aim of this contribution to solve the open problems in Sørensen's discussion and to provide a class structure that represents the duality between owners and non owners of rent-producing assets.

capitalist societies. Our aim is to move some steps in this direction, starting from a general situation with restrictive hypothesis that are subsequently relaxed.

Let's start by considering a fictional "simple" society where all actors have same tastes for work and leisure, there is no educational system and in which all economic transactions take place in a perfect market, that is a market characterized by a perfectly competitive equilibrium. Following Sørensen (1996), we introduce one more restrictive hypothesis: this fictional society exists only for the lifetime of its starting incumbents. Two corollaries are derived by definition. First, in this market there are no extra prices to access goods or resources. In other words there are no rents. As a consequence, the total wealth of individuals is the combination of revenues from labour and savings. Second, in this market there are no incentives to competition, that is the situation in which different actors seek to obtain a share of a limited good. Competition is defined as a zero-sum game: if "A" obtains the good (or advantage), "B" will be excluded from it. Even if we relax the hypothesis of same tastes for leisure and work among individuals there will be more differences in total wealth but no part of it will result from rents.

In our fictional society the absence of an educational system means that knowledge is produced in a shared manner among citizens, or not produced at all. Let's relax the hypothesis of no education and consider the introduction of a simple educational system, that guarantees access to all who apply for it and which provides, for simplicity, a single title after n periods of attendance. The presence of such a system means that now a portion of knowledge is "produced" in codified ways and certified by that institution. The possession of knowledge is valuable in the labour market, where titles are rewarded with premium wages. Until access to education is granted to all who apply for it, there are no conditions for education as a rent.

We can now relax the hypothesis of a single-institution educational system, and introduce a second institution or school. The two schools, let them be "A" and "B", formally provide the same title. If the two institutions differ for some characteristic and these differences are unknown and relevant in determining different returns in the labour market (e.g. in favour of "A") owners of that title will have a lifetime advantage at the expenses of non owners only because they enrolled in a school instead of another. We can conceptualize the mechanism through the structuralist model proposed by Sørensen (1979): employers hire degree-holders from the high level as possible, and workers accept the best jobs offers. As a consequence, the most desirable jobs go to the more educated, while the less attractive to the less educated (Triventi et al. 2015). This model holds also when qualitative differences among titles are present, as we can imagine that degree-holders are sorted according to amount and type of education. As far as enrolment is determined by free choice, the situation described above does not identify education as a rent-producing asset, as enrolment reflects individual investment strategies. If enrolment is determined by chance, it might be the case for education as a rent given that some level of inequality is produced due to premium wages granted on random allocation. What is less clear to

identify is whether a rent generated by random allocation can be the cornerstone of antagonistic interests.

Let's now relax the final assumption of our fictional society and allows it to go beyond the lifetime of its starting incumbents. Here, the problem of inter-generational transmission of education appears. With the concept of *inter-generational transmission of education* we refer to actions aimed by a generation at securing at least the same levels or type of attainment to its offspring. What makes education an asset capable to produce a rent in the labour market is the intrinsic nature of inter-generational transmission. The idea at the base of this action is the transfer of a good or resource, that means the transfer of property rights on that asset. Clearly, it is not possible to transfer the property of an educational title between actors, as the title certifies the possession of a certain amount of knowledge of an individual. What is relevant in this case is to find a proxy for the transfer of property rights. Such proxies might be identified in two moments. The first proxy is represented by attempts to maximize possibilities of enrolment to a given educational level or institution (Breen and Goldthorpe 1997; Lucas 2001). Here, the transfer takes the form of granting the same (relative) possibilities of access to a given level or school. The second regards resources mobilized by parents in the attempt to maximize the probabilities of attainment. Here, the transfer takes the form of granting the successfull conclusion to the educational carreer. Of the two moments, the first is the most relevant as it is a necessary condition for the second one. As a consequence, access to education is no more on ability or personal inclination basis. When the offspring of the more educated are able to access the higher stages of education because of parental action, intergenerational transmission of education makes education a rent-producing asset. Access to education is then characterized by a latent property relationship, on which antagonistic interests and social closure dynamics arise.

School segregation is then a consequence of inter-generational transmission dynamics. In fact, successful attempts by the more educated to enroll their offspring in certain schools might result in a general redistribution of pupils among schools. The selection of pupils into school is thus biased by social position, resulting in some degree of segregation.

2.2.2 Relative Risk Aversion

Breen and Goldthorpe (1997) developed a model of educational decision rooted in rational choice and relative risk aversion theory.

When deciding whether continuing in education or entering the labour market, families and their offspring make a rational choice aimed at maximizing their utility. When deciding if continuing in education or not, families evaluate the attached direct costs of education (tuition fees, books, etc.) along with discounted earnings, and the probabilities of success of the pupil. Moreover, families consider their beliefs about chances of access to possible destination classes with reference to different outcomes (entering the labour market, continuing in education with success, continuing

in education with failure). The assumption of the model is that parents want their offspring to acquire a class position at least as advantageous as that from which they come from. In other terms, families maximize their utility when they are able to minimize the risk of downward mobility. Several factors enter the equation generating class differentials. Upper class families can afford more easily the direct and indirect costs of education, since they possess more resources; given the position in the social structure, they have more incentives to stay longer in education; pupils ability may not been randomly distributed among social groups. Another key element is represented by information. The accuracy of the evaluation can be improved by a better knowledge of costs, benefits, possibilities to succeed of each educational level.

From this model follows that competition among families for avoiding social demotion leads to inequality of educational opportunities. In this scheme, strategies of usurpation take place when lower status pupils are able to go as far as high status pupils in the educational system, thus sharing the probabilities to enter in the service class. Strategies of exclusion are represented in the model by the usage of resources (money, information, etc.) from high status families to push their offspring towards the higher stages of education.

2.2.3 Effectively Maintained Inequality

The effectively maintained inequality theory (Lucas 2001) takes into account the structure of the educational system to investigate both processes of school continuation and track mobility. Similarly to the model proposed by Breen and Goldthorpe (1997), this theory posits that socioeconomically advantaged actors try to secure for themselves and their children some degree of advantage wherever possible. The model goes on in specifying two types of situations. Whenever quantitative differences are common, high status actors will look for quantitative advantages. This means that if an educational level is not universal, actors will seek for an advantage at the level (this is similar to Breen and Goldthorpe's model). Conversely, if qualitative differences are common, high status actors will look for whatever qualitative difference in order to gain advantages and secure better education.

Similarly to the model previously discussed, competition arises around the "conquest" of some degree of advantages. On the one hand, when quantitative differences are common, it will be possible to observe segregation by social status as a consequence of inequalities of educational achievement. On the other hand, social segregation by tracks. Moreover, the effectively maintained inequality theory takes into account hidden tracking and all the different kinds of qualitative differences. This implies that segregation, as the result of competition upon the securing of advantages, can result also at the micro-level (schools) when qualitative differences are allowed at this level. This requires policies of differentiation and autonomy. Here, strategies of closure will be all that strategies used by high status families to defend the most remunerative educational paths.

2.2.4 Education and status

The above presented theories are concerned with the labour market value of education, that is the associated premium wages, and the position in the social division of labour that it allows. What is missing is the other side of the coin: the relationship between education attainment and social status. The historical inquiry by Collins (2000) on patterns of education across societies has shown that (i) the evolution of schooling has been a process rooted in the evolution of human communities; (ii) education has been a tool for selection of élites. The first element highlights the importance of considering educational systems in relation with the societal organization, while the second is usually embedded in it. Collins (2000) identifies four ideal-typical educational organizations: (i) community schools; (ii) family or guild education; (iii) professional licensing; (iv) bureaucratic organization. The third and fourth organizational types are the ones relevant in nowadays school systems, usually characterized by a combination of the two. Professions are occupations accessible only through the possession of a specific degree (i.e. the lawyer) that certifies a formal training. Professions have a high social prestige derived by the social closure strategies (licensing) put in place to control access. Large scale, bureaucratic school organization is typically related to a division of labour among circumscribed positions coordinated by hierarchic channels of authority (Collins 2000). What is relevant in the discussion about education as a status symbol is the perceived value of a title in a given society at a given time. Collins (2000) shows several historical examples of inflation of educational credentials (Collins 1979), a situation that implies decreasing returns and prestige to educational titles in time.

2.3 Research questions

From the presented literature emerges the role of inter-generational transmission of education as a possible factor in creating school segregation. Four blocks of research questions derive from this idea, each aimed at contributing to existing literature by providing some evidence from the Italian case and by addressing limitations of current research.

- Which is the level of segregation of school districts and schools in Italy? Which are the determinants of residential and school segregation?

  The literature review on school segregation in Italy presented in Chapter 1 has shown the lack for a nationwide investigation of the phenomenon. Pacchi and Ranci (2017) have investigated the Milanese case in detail with reference both to school and district segregation, while Barberis and Violante (2013) have focused only on mean levels of school segregation in Milan, Bologna, Rome and Naples. The determinants of segregation are investigated through the ecological approach proposed by Farley and Frey (1994). The ecological tradition in studies of segregation recognizes the role of structural elements of the urban

space and history of a metropolitan area in influencing socio-economic and ethnic segregation.

- How is district composition reflected in school composition? What makes families more prone to enroll their offspring in a school different from the closest public school?

  To the author's knowledge, only the recent work by Pacchi and Ranci (2017) on Milan has addressed how district composition reflects in school composition and its reasons. These questions aim at providing more evidence on the issue by extending the case to other relevant Italian metropolis and cities. With reference to the second question, and according to previous work on this theme (Pacchi and Ranci 2017), we test the hypothesis of socio-economic and ethnic composition of schools as main determinants of pupils flights between schools. Moreover, we test the hypothesis of school quality as a relevant element in determining enrolment to a given school.

- Is there a significant link between the degree of segregation of neighbourhoods and schools/classes on the performance in achievement tests? Is it heterogeneous over pupils?

  Contini (2013) has shown that the concentration of first and second generation migrants in classes of lower secondary school has a small but significant negative effect on the individual INVALSI test score, in accordance to previous literature. The aim is to extend this contribution by investigating whether also urban segregation can have a detrimental effect on educational performances.

- Are those coming from segregated schools/classes more inclined to make any school transition? Is it heterogeneous over pupils?

  The idea is to test the following hypothesis: everything being equal, those coming from schools where the student population has higher degrees of homogeneity with respect to their socio-economic characteristics have the same probabilities to make any school transition with respect to their peers coming from schools where the student population is more diverse according to its socio-economic characteristics. To the author's knowledge there is no research on the Italian case that tests this hypothesis.

2.4 Conclusions

The role of institutional factors is pivotal in creating conditions for dynamics of segregation. These factors take the form of policies, in particular those aimed at creating differentiation among schools, in increasing institutional and organizational autonomy and in fostering family free choice. These kind of policies transform the educational provision in a scarce resource thus producing competition among actors (both schools and families) in order to access it. When the number of competitors increases

in relation to the profit span (Weber 1968), agents try to secure profitable resources through individual and collective action aimed at reducing access and controlling competition. This process of monopolization of social and economic opportunities takes the name of *social closure*, the mechanism through which competition translate into segregation.

The theories presented in this chapter have highlighted the profound reasons that make families prone to compete for education. In advanced societies education is linked to increased opportunities to get a better wage, better working conditions and occupation in the labour market. According to authors presented (Sørensen 2002; Breen and Goldthorpe 1997; Lucas 2001) families compete on education in order to secure some degree of advantage. But education is also a status symbol. This complementary explanation by Collins (2000) let us consider competition for the educational provision as driven by a mixture of both advantage securing practices and symbolic status attainment. Obviously, it is impossible to determine which explanation is driving the process, as the mixture may vary among individuals, social groups, societies.

Chapter 3

Data, methods and the institutional context

This chapter is devoted to the discussion of the general methodological issues of this work, while the strategy adopted to answer the research questions will be highlighted in the specific chapters (Chapter 4-6). Section 3.1 introduces the rationale for focusing on lower secondary school in Italy, highlighting this level as pivotal in educational careers of pupils. Section 3.2 presents the institutional context in which our environment of interest (school) is embedded. Section 3.3 presents the databases used to answer the research questions previously discussed (Chapter 2): the ISTAT population census (2011) and INVALSI survey on learning (2015). The section includes a focus on available variables and the strategy used to merge the two sources of information in a single database, as well as a section devoted to presents the issue of missing data in the INVALSI database. Section 3.4 discusses three methodological issues related to the measurement of segregation. The first is how to measure segregation. In their seminal work, Massey and Denton (1988) identified five dimensions of segregation: eveness, exposure, clustering, centralization, and concentration. Section **??** focuses on the first two aspects, including a review of main indices of segregation presented in chronological order, along with a discussion of their pros and cons. An alternative strategy proposed by Préteceille (2003) is also discussed. The second issue is about how occupations and nationalities, our two dimensions of interest, are measured in the two databases and make comparable. The third discussion concerns the proper level of analysis and the sample selection. Finally, Section 3.5 sums up chapter conclusions reviewing the methodological choices and discussing limitations that should be addressed in future work.

3.1 Why focusing on lower secondary school?

In the last century, the Italian educational system has experienced a general de-stratification of educational levels accompained by a strong expansion in school enrollment and completion rates. During Fascism, the educational system was characterized by a strong tracking regime where lower secondary school play a pivotal role. After five years of primary education, students could choose one of four different tracks: the

Ginnasio, a five-years track granting access to the more academic upper secondary path; the *Istituto tecnico inferiore*, a four-years track granting access to thecnical upper secondary education; the *Istituto magistrale inferiore*, granting access to a track devoted to the formation of teachers; the *Scuola di avviamento professionale*, a vocational track not permitting the continuation of studies. This structure characterized the educational system until 1962, when tracking was dismantled: the vocational track was abolished and the remaining tracks replaced by a comprehensive public-funded lower secondary school granting access to all upper secondary tracks. The de-stratification of Italian school has accompained the educational expansion process and somehow influenced it. The increase in enrolment rates has involved the country as a whole, although a North-South divide has been documented (Ballarino, Panichella, and Triventi 2014), especially for higher levels of the educational system. The educational expansion process has somehow lessened the inequality of educational opportunity by social class of origin, especially for lower grades. While primary and lower secondary school completion rates are similar across social strata, probabilities to enrol to upper secondary or university (after completion of the previous level) vary substantially between the two social groups at the ends of the social structure: the working class and the service class.

Despite its being universal and de-stratified, the lower secondary educational provision is not homogeneous across the country, this is due to qualitative differences at the school level. The variability involves the endogenous characteristics of schools such as teachers' quality, organization of curricular and extra-curricular activities, and facilities status. All these elements might shape family preferences about enrolment in a given school, making the composition of the student population non-random. Such a non-randomness in the student population, along with it's determinants and consequences, is the phenomenon of interest. Moreover, the Italian lower secondary school is the last educational level before upper secondary tracking. Thus its selection, and orientation mechanisms are pivotal in shaping educational transitions of pupils.

Finally, a practical reason: data on lower secondary school are more complete and robust than data on primary school. This is due to the fact that individual level data for lower secondary are collected by INVALSI at grade 8 through a test that is part of the final examination. Thus, it is completed by all students but those not admitted to the final exam. Since its introduction, the INVALSI test has gone through a wide dispute by both students (in upper secondary) and parents. The grade 5 test data we collected are severly affected by this dispute, making them not suitable for reliable statistical analysis.

3.2 The institutional context

In Chapter 2 I've discussed the importance of the institutional context in creating conditions for practices that may lead to appearance of segregated environments. The

main institutional cause is represented by policies that introduce differentiation in the supply side and competition on both supply and demand of educational provision. The institutional context for Italian lower secondary school has moved towards both kind of policies.

On one side, school autonomy was introduced with Law 15 March 1997, n. 59 in the form of autonomy in the articulation of the didactic activity. The aim was to carry out educational and training interventions aimed at the development of the person, the expansion of the offer and its adaptation to the various contexts and requests (families, local economy). Subsequent regulation (Presidential Decree No. 275/1999) specified that schools plan these interventions in the plan of educational offer (POF), a document that must be discussed every three years by each institution and presented to the users of the service, students and families, at the time of enrolment in school. Autonomy can involve the organization of teaching activities through the regulation of time devoted to each discipline by adapting it to the type of studies and the learning pace of the pupils. Schools can foster schedule flexibility, activate individualized educational paths, plan training courses in coordination with the requests of the territory, choose methodologies and teaching tools. Autonomy is declined also as integration and coordination between the different articulations of the school system through the creation of networks.

On the other side, enrolment procedures have been reviewed in an attempt to foster family free choice. Law 7 August 2012, n.135 has introduced an online enrolment procedure in which families express three preferences of schools to which they want to enroll their offspring. There are no criteria limiting the school choice except the maximum number of preferences expressed. Applications are accepted by the school up to the maximum of available places, however the school board defines criteria to solve priority in case of extra applications. This criteria usually accounts for home-school distance, school-work distance, occupational condition of parents, number of siblings enrolled in that school. The Ministerial Frequently Asked Questions (FAQ) on this topic clearly states that schools are not obliged to process enrolment requests following the receiving order, thus granting a great leverage to schools. Previously to 2012 enrolment to lower secondary school took place through applications through the attending primary school (2004-2011) or through direct application to the school headmaster (prior to 2004). Differently from systems adopted in other countries, school districts are irrelevant in enrolment procedures to Italian lower secondary school. Districts were created in 1974 (Law 31 May 1974, n.416) as an administrative unit defined by municipalities, each district had to have at least 10.000 inhabitants up to a maximum of 200.000, and it was governed by a board representative of school personnel, and local economic and political actors. In 2003 (Law 27 December 2002, n.289) the personnel and functions of school districts were transferred to schools, resulting in their *de facto* elimination.

To sum up, the evolution of the Italian institutional context took place in two different moments: first, since the '90s it was introduced the possibility for schools to

differentiate the educational provision through organizational autonomy (possibility to redefine schedules and activities); second, the abolition of school districts (2003) and the implementation of an enrolment system fostering family free choice (2012).

3.3 Data

In order to answer the research questions I refer to a database built from two different sources merged together by using spatial information, allowing to investigate the micro-spatial structuring of education and the effects of a homogeneous composition of the student population, according to social class and immigrant background, on standardized test scores.

3.3.1 Population census - ISTAT

The first source is the *2011 Population Census* provided by ISTAT, the latest census wave available. Census data are collected every ten years at the household level but provided in aggregate form. The elementary geographical units for both collection and output are the *sezioni di censimento*, the smaller territorial partition at which data are aggregated. The *sezioni* have been created primarly for enumeration purposes and are defined according to boundaries that coincide with physical elements of space (rivers, roads, railways, etc.) using centrelines as demarcation. As a result of this criterion, the polygons are characterized by highly irregularity. Given their nature the number of the *sezioni di censimento* is not stable over time. Moreover, their boundaries have changed in the last three waves of the Italian census and new polygons have been added due to new urbanized areas. This makes comparison across time a very difficult task. Moreover, the defining criteria do not include lower population treshold, thus resulting in a great territorial variety among urban areas. Finally, the *sezioni*'s dimension poses problems when coming to analysis of residential segregation, as they are too small to provide reliable measures.

Census data include information on residential population, its distribution among different classes of age, education, occupation and nationality, along with information on buildings' condition. The information have been collected through long forms (complete quesionnaires) and short forms (regarding only demographic and a sub-set of socioeconomic variables), depending on population levels of each urban aggregate. For municipalities with less than 20.000 inhabitants a traditional approach has been carried out: the long form has been submitted to the whole population, allowing for paper or on-line completion. For those municipalities that are capital provinces, or that have more than 20.000 inhabitants, a sampling strategy has been carried out: the short form has been submitted to about two third of the municipality's population, while additional information have been collected through the long form submitted only to sampled households (the remaining third).

3.3.2 2015 Survey on Learning - INVALSI

The second source of data is the *2015 Survey on Learning - INVALSI*. INVALSI is the governmental agency for the evaluation and monitoring of the educational system, each year it runs a survey aimed at investigating the skills and knowledge of pupils. The survey has been introduced with the law 25/10, 2007, n. 176, and administered for the first time in the 2007-2008 along with the final examination of grade 5. Table 3.1 sums up the waves of INVALSI test by grade in the last ten years. The survey is built of two different tests aimed at measuring literacy in mathematics and Italian,[1] a standardized score corrected for a cheating parameter is computed for each pupil.

Grade	School Year									
	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17
II		✓	✓	✓	✓	✓	✓	✓	✓	✓
V	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
VI			✓	✓	✓					
VIII			✓	✓	✓	✓	✓	✓	✓	✓
X				✓	✓	✓	✓	✓	✓	✓

TABLE 3.1: INVALSI test by grade and year.

The standardized score is produced according to the Rasch model (Rasch 1961), a probabilistic model developed with the aim to operationally define a way to estimate, by the results of a test, individual' skills and difficulty of items. The Rasch model belongs to the family of latent trait models within test theory. Latent trait models specify a relationship between observable performance and an unobservable dimension assumed to underlie performance (Gustafsson 1980). An important characteristic is that parameters describing the test items can be estimated in such a way that they are invariant among different groups of individuals, and estimates of ability can be produced in a way they are invariant from one sample of items to another. INVALSI refers to the most famous model of this family: the Rasch model for dichotomous data. Responses to the test are organized in matrices in which each line is an individual and each column an item of the test. Each cell can assume two values, [0] for wrong answers and [1] for correct answers. The idea behind this model is that answers are a function of two parameters: individual ability and the difficulty of the item. The greater the ability (e.g. preparation) to correctly answer to a given item, the greater the probability to provide a correct answer. This idea is synthesised in the following mathematical form:

$$Pr\{X_{ni} = 1\} = \frac{e^{\beta n - \delta i}}{1 + e^{\beta n - \delta i}}. \tag{3.1}$$

1. For grade 2 a third test on reading abilities is calculated.

The model has the form of a simple logistic funcion and express the probability for an individual *n* to correctly answer the item *i*. In equation 3.1 β_n is the ability of the n^{th} person and δ_i the difficulty associated to item *i*. Despite its simple structure and the power to solve complex mathematical problems, the Rasch model's limits are predominantly rooted in strong assumptions. First, the model consider all items to have the same discriminative power, this is the *homogeneous item discrimination assumption*. Second, the model assumes that there is only one parameter of ability, that is ability is explained by a single factor or trait. This is the *unidimensionality assumption*. From this assumption derives the third condition: the probability of an examinee response pattern is given by the product of all probabilities of item responses conditional to the unobservable trait. This is the *local statistical indipendence* assumption, and it is inseparable from the unidimensionality assumption. Theoretically speaking, the specific factors of an item must be uncorrelated with the ability measured, thus fulfilling the assumption of orthogonality.

The Rasch model appears to be violated when there are heterogeneous items that measure different abilities, when the items do not discriminate in the same way, when some items are systematically too easy or too difficult for certain subgroups of the sample (Gustafsson 1980). In practice, each situation that leads not to fulfill the orthogonality assumption represents a violation to the Rasch model. To control for these issues, INVALSI submits the items to pre-testing techniques that investigate whether the discriminatory power of items is homogeneous (or in a given range) and that test the unidimensionality assumption. Gnaldi and Bacci (2016) have found evidence for the reliability of multidimensional latent class models: with respect to the national test administered to lower secondary school students in 2009, the assumption of unidimensionality is strongly rejected for the Language test, while it is plausible for the Mathematics test. No other independent analysis are available for further waves of the national test, but despite the early reports the most recent technical analysis (INVALSI 2016) include a specific test for the unidimensionality assumption. The test accepts the assumption.

The INVALSI database has been enriched with information from the student questionnaire. This survey provides data on personal and family characteristics such as parental occupation, parental education, demographics, and immigrant background. Moreover, it provides a single continuous indicator called ESCS as proxy for the socio-economic and cultural capital of pupils' family. The ESCS is a composite index drawn from the PISA test methodology, it is a standardized index computed with principal components analysis techniques and built upon three different indicators: (i) HISEI, for socio-economic status; (ii) PARED, for the parental educational level; (iii) HOMEPOS, an indicator of home possessions. The HISEI is the highest value of the ISEI of either parents, the international socio-economic index of occupational status developed by Ganzeboom (2010), PARED accounts for parental education measured as number of years of schooling, finally HOMEPOS is a summary index of all household possesions (technological possesions, desks, quiet places to study etc.) included the

number of books. Given that the information for the computation of these indices is provided by the pupil through questionnaire completion, the risk of missing data is high. To solve this problem INVALSI recurs to two strategies depending on whether missing information regards only one indicator or more. In the first case INVALSI recurs to imputation of data using regression predictions, in the second case the modal value of the pupil' school is imputed.

3.3.3 The merging process

The merging process that led to the final database has been possible thanks to the collaboration of INVALSI, especially for the linking of data characterized by privacy information. The first step was to obtain two separate lists of addresses: one for lower secondary school buildings and one for residence addresses of pupils. Secondly, I've georeferenced these addresses by translating the individual record into a pair of ground coordinates (latitude and longitude). Consequently, the precision of the encoding has been hand checked with respect to several entries characterized by bad completion, special characters, and other typos. Particular attention has been devoted to the precision of school coordinates, whether possible missing addresses have been completed by identification of the school through its fiscal code, the *Codice Meccanografico*, verification and subsequent attribution of an address. Once the two lists have been geocoded, only coordinates and identification codes for schools and individuals have been retained. INVALSI have merged the pupils' residence coordinates with the pupils' record in the INVALSI test. Since the *Codice Meccanografico* of the attending school is attached to the pupils' record, the coordinates of each school have been added. The third step has been to add information from the population census. Each pair of coordinates (both for schools and pupils) has been assigned to a *sezione di censimento* through the STATA command `[geoinpoly]` developed by Picard (2015). This command uses an algorithm to determine if a point falls inside a polygon. It uses the geographic locations in memory in the form of polygons from a shapefile and checks whether they spatially overlay the points.

The final database is build by individuals nested in classes and schools, as this allows to investigate school segregation and its effects on school performances. Moreover, individuals and schools are nested in the geographic space and information on the reference census tract is attached to each record. This allows to investigate the micro-spatial dynamics of school enrolment and segregation. Although this structure is particularly useful in computing indices of segregation, isolation and concentration for schools (when algorithms that rely on individual or aggregated data are used) it is not for the analysis of segregation at the urban level. In order to solve this problem, a pair of auxiliary databases have been created: one for the analysis of segregation in the urban space, for which each record is a *sezione di censimento*; the other for the analysis of segregation in school districts, for which each *sezione* is linked to the related school. After computation of results and some reduction operations, both databases can easily be merged with the main one.

3.3.4 Areas of interest and sample selection

A long tradition in studies on residential segregation looks at functional areas as the relevant level of analysis. These areas are usually defined considering local labour and housing markets, thus identifying urban areas characterized by a strong degree of integretion. About Italy, sampling strategies in line with this tradition might look at Functional Urban Areas as defined by OECD (OECD 2013) and Local Labour Systems (*Sistemi Locali del Lavoro*) as defined by ISTAT (ISTAT 1991). The Functional Urban Areas defined by OECD address the problem of comparability of metropolitan areas at both national and international level by considering urban unit as functional economic environments. The methodology uses population density to identify urban cores, and travel-to-work flows to identify the sorrounding hinterland. A key feature of this definition is that hinterland and core's labour markets are highly integrated. More specifically, municipalities are identified as hinterland of a core whether at least 15% of resident population <is working in the core area.

The process results in a four-category scale: (i) Small urban areas, with a population below 200.000 people; (ii) Medium-sized urban areas, with a population between 200.000 and 500.000 people; (iii) Metropolitan areas, with a population between 500.000 and 1.5 million people; (iv) Large metropolitan areas, with a population greater than 1.5 million people. Similar to that of OECD, ISTAT's Local Labour Systems (LLSs) are conceptualized as aggregation of municipalities characterized by highly integrated labour markets. Despite the OECD methodology, integration of urban cores' labour markets is not defined upon a fixed share of the local workforce that dayly moves to the core. Conversely, ISTAT relies on self-containment functions of both labour supply and demand, jointly considered with the absolute figures of resident workforce.[2]. Since self-containment functions are characterized by strict conditions, ISTAT's LLSs are more conservative in defining functional areas than the OECD methodology.

Studies on school segregation are less related to a specific tradition. The majority of available research investigates school segregation at the upper secondary level looking at national and cross-country comparisons, research on school segregation in the US investigates trends at the metropolitan area level while the few studies focusing on primary and lower secondary schools in Europe usually analyze the phenomenon at the municipality level. This is the case also for the few Italian studies. However, the question about why school segregation in Italy should be investigated at the municipal level usually remains with no answer. The main causes of school segregation might push towards different methodological choices. On one side, institutional factors. As discussed in section 3.2, the Italian enrolment system allows pupils to freely move among schools. Thus between-municipality flows are allowed. As a consequence, framing the analysis at the broader available sensible geographic level could be a good solution (provinces, functional areas). On the other side, school

2. See https://www.istat.it/it/files//2014/12/nota-metodologica_SLL2011_rev20150205.pdf for detailed methodology

segregation is, among other factors, the result of school and residential preferences of families, along with school behaviours. About family preferences, research on school choice and school segregation on primary and lower secondary has shown that parents evaluate a limited set of short-range alternatives. Coming to school behaviours, school boards of Italian lower secondary have some degree of discriminatory power in accepting applications, and school proximity is usually a criterion of selection. The combination of these elements makes school segregation, as deriving from school choice and residential position, a local phenomenon. Here, the municipal level seems a good option for investigating school segregation dynamics in Italy. Moreover, between-municipality flows can be accounted for by considering the complete pupil population of a municipality for a given grade.

Since the aim of this work is to investigate school segregation and its determinants, I opted to focus on a selection of functional areas' urban cores in order to analyze the related lower secondary school systems. The Local Labour Systems with the highest shares of migrants and unemployment rates with respect to the same geographical sub-area (North-West, North-East, Centre, South, Islands) have been selected. In the first phase, all provincial capitals with more than 250.000 inhabitants have been selected, giving a list of the top twelve municipalities[3] according to population figures. In the second phase, provincial capitals with the highest share of foreign residents by macro area have been selected, this resulted in a list of 12 additional municipalities.[4] In the third step, provincial capitals have been selected among those with the highest levels of unemployment in the related macro area, resulting in additional 11 municipalites.[5] Given the typical north-south divide in the economic development of the country, Northern cities are characterized by higher shares of foreign residents but lower shares of unemployment, conversely for Southern cities. The process results in 35 selected municipalities that account for the 0,5% of the total number of municipalities (8000), 19.19% of the Italian population and about 23% of grade 8 students.

3.3.5 Missing data

Due to several reasons, both ISTAT and INVALSI data are affected by the presence of missing data, that might pose different threats to results. About census data, the problem involves the variable recording occupations. Table 3.2 sums up how many *sezioni* in the entire database are affected by this problem. About 67% of areal units has complete information or just one missing category, but 17% presents missing values in all occupational categories, and less than 19% in more than 5 categories. This situation is typical of areal units with a small number of residents, usually located in villages or small towns, in fact the 50% of these *sezioni* has less than ten individuals

3. Bari, Bologna, Catania, Firenze, Genova, Milano, Napoli, Palermo, Roma, Torino, Venezia, Verona.
4. Asti, Brescia, L'Aquila, Mantova, Modena, Olbia, Piacenza, Perugia, Prato, Ragusa, Reggio Emilia, Teramo.
5. Agrigento, Caltanissetta, Crotone, Frosinone, Imperia, Latina, Padova, Rovigo, Savona, Udine, Vibo Valentia.

Missing count	Freq.	Percent	Cum.
0	218,343	59.52	59.52
1	28,146	7.67	67.19
2	17,243	4.70	71.89
3	15,137	4.13	76.01
4	12,276	3.35	79.36
5	7,503	2.05	81.41
6	3,290	0.90	82.30
7	949	0.26	82.56
8	166	0.05	82.61
9	63,810	17.39	100.00
Total	366,863	100.00	100.00

TABLE 3.2: Number of *sezioni di censimento* with n missing values.

with a job position. The geographic distribution of *sezioni* with missing values is uneven, they are concentrated in four regions: Emilia-Romagna, Lombardia, Piemonte, Veneto. Each of these regions counts about 13% of the total figure of areal units with more than 5 and less than 9 occupational categories with missing values. However, these percentage means that 1-2% of all *sezioni* of the above mentioned regions has a considerable amount of missing values.

For each *sezione di censimento*, if the missing information refers to just one category, then we are able to fill information by computing the difference between variable $p61$ (15-years old individuals that are occupied) and the sum of all the occupation variables. If the missing information concerns more than one category, we are not able to correctly allocate the remaining individuals. Finally, if all the occupational categories are missing, the *sezione* is not used in the computation of the segregation measure.

The same problem affects different variables in the INVALSI database, but two are the most important: the geographic data, and family background information. Of the 113.700 individuals in the final database, residence address is missing for 263 students. However, addresses are not reported with a common output and mistakes, typos, and special characters are usual, resulting in 3.261 individuals with missing latitude and longitude due to impossibility of geocoding.[6] When coming to geographic data of schools, only one address is missing out of 1098 recorded institutions. However, information on attended school, registered via an anonymous id code, is missing for a minority of students. In the following table, the pattern of missing data for residence and school addresses are reported. The plus (+) indicates the presence of data, while the dot (.) represents missing information. The table shows that in the database there are 4.074 pupils with residence address but missing information on the attending school. To this group we add the 2.977 pupils for which we know

6. The goodness of geocoded addresses has been checked at the highest detailed level.

the attending school but no information on residence is available. Finally, the small group for which there is missing information on both variables.

Pattern	Frequency
++	106.365
+.	4.074
.+	2.977
..	284
Total	113.700

TABLE 3.3: Pattern of missing data for residence and school addresses.

Coming to family background information registered through the INVALSI' student questionnaire, the situation is characterized by a higher amount of missing data. Let's have a look to the most relevant variables, those characterizing family social and immigratory background. Information about the ESCS index (socio-economic and cultural capital of a household) is available only for 84.401 individuals (74% of sample), family socio-economic background as measured by social class is available for 97.625 individuals (85% of sample), while citizenship (Italian vs first and second generation migrants) is available for 97.600 pupils (85%) and geographic area of origin of parents for 91.528 individuals (80%). The following table reports the patterns of missing values for these four variable of interest in the following order: parental geographic area of origin; ESCS index; citizenship; social class. As in the previous table, the (+) indicates the presence of data, while (.) represents missing information.

Pattern	Frequency
++++	73.081
.+++	3.870
+.++	12.562
++.+	12
.+.+	7.438
+.+.	5.870
..++	652
+..+	1
+...	2
..+.	1.565
...+	9
....	8.638
Total	113.700

TABLE 3.4: Pattern of missing data for parental geographic area of origin; ESCS index; citizenship; social class.

3.4 Methodology

This section presents a discussion of four methodological issues relevant to this work. Subsection 3.4.1 discusses the evolution (in chronological order) of segregation measures, their rationale as well as their advantages and criticisms. Finally, it presents criteria for the evaluation and selection of proper segregation indices. Subsection 3.4.2 and 3.4.3 discuss how occupations and nationalities are classified in the two databases of interest and which solutions are available in order to make them comparable. Subsection **??** presents the sample used in this study and how it has been defined.

3.4.1 Measures of segregation

The definition of measures of segregation requires the solution of some methodological and conceptual issue, not always addressed in the literature. First, the definition of the reference unit on which segregation is computed. Given that the majority of indices presented has been developed in order to provide a measure of residential segregation, the reference unit has always been a territorial partition usually conceptualized as "neighbourhood" and operationally identified as (agglomeration of) census tract. Second, the definition of reference population and the characteristic of interest. Different segregation indices can be derived if the variable of interest is discrete (categorical or ordinal) or continuous and if segregation is measured between two or more population groups. Third, a conceptual definition of segregation is required. Massey and Denton (1988) define segregation as a multi-dimensional concept, in contrast to the mono-dimensional previous conceptualization. They identify five dimensions: *evenness of distribution*, *exposure*, *clustering*, *concentration*, and *centralization*.

In this subsection, the most relevant and useful indices of segregation for unordered categorical variable (i.e. race, class) are discussed and the notation suggested by Reardon and Firebaugh (2002) is used. Consider a region R populated by M subgroups indexed by m. The region of interest is divided in r sub-regions, p index points within each spatial partition. Finally, τ indicates population density and π population proportions. Thus we have

$$
\begin{aligned}
T &= \text{total population in area } R \\
t_{\mathrm{r}} &= \text{total population of subarea } r \\
t_{\mathrm{rm}} &= \text{absolute frequency of group } m \text{ in subarea } r \\
\tau_{\mathrm{p}} &= \text{population density at point } p \\
\tau_{\mathrm{pm}} &= \text{population density at point } p \text{ for group } m \\
\pi_{\mathrm{m}} &= \text{relative frequency of group } m \text{ on total population} \\
\pi_{\mathrm{rm}} &= \text{relative frequency of group } m \text{ in subarea } r
\end{aligned}
$$

The discussion around the development and testing of measures of segregation occupies the literature, with alternate intensity, since the late '40s. Three waves of

debate can be identified: the first wave goes from late 1940s to 1955, when the seminal work by Duncan and Duncan (1955) led to conceptualization of the dissimilarity index. This was followed by a *Pax Duncaniana*, as defined by Massey and Denton (1988), that lasted until 1980s. The second wave goes from late 1970s to around 1990, and it is characterized by several contributions that examine the inadequacies and shortcomings of the dissimilarity index and that propose indices suited for computing multi-group segregation. The apex of this wave of debate can be identified in the often cited works by James and Tauber (1985) and Massey and Denton (1988). In the first, four criteria for the evaluation of segregation measures were presented:

1. *Organizational equivalence*: segregation is unchanged if the organizational unit is splitted in sub-units with the same group proportions of the original.

2. *Size invariance*: segregation is unchanged if the number of persons in each group is multiplied by a constant factor.

3. *Transfers*: segregation is reduced if an individual of group m moves from one areal unit to another where the proportion of individuals of that group is lower than the starting unit. This is also the case if the individual of group m is exchanged with an individual of group n from another areal unit, and proportions of group m and n are higher in the starting unit than the target one.

4. *Composition invariance*: segregation is unchanged if the number of persons in a group increases by a constant factor while the number and distribution of persons of all other groups is unchanged.

In the second, the authors elaborated on the concept of segregation by articulating it on five dimensions and classifying existing indices. Finally, the third wave goes from the 1990s to the first decade of the 21st century, and it is characterized by contributions aimed at developing indices that account for the spatial dimension of segregation.

The first contribution on the theme can be traced back to Jahn, Schmid, and Schrag's (1947) proposal of four simple measures of segregation based on census tracts as reference unit of computation and rooted in the assumption that no segregation exists if place of residence by census tracts is uninfluenced by racial factors (Jahn, Schmid, and Schrag 1947). This article was followed by strong criticism (Cowgill and Cowgill 1951) and proposal of alternative measures (Shevky and Williams 1949; Bell 1954; Duncan and Duncan 1955). Since sociologists of that time were concerned with finding a compositionally invariant segregation index (Massey and Denton 1988) the Duncans' dissimilarity index D became popular instead of isolation measures (as those proposed by Bell). The Duncans derived this index from the Lorenz curve, which plots one against the other the cumulative distributions of minority and majority groups ordered from smallest to largest minority proportion. The dissimilarity index is the maximum vertical distance between the Lorenz curve and the line of evenness

($y = x$), and can be interpreted as the proportion of minority members who would have to change their tract of residence in order to have the same proportion of minority members in all tracts. The index is defined in the range [0,1], where [1] represents a situation of maximum segregation.

$$D = \frac{1}{2} \sum_{r \in R} \frac{t_r |\pi_{rm} - \pi_m|}{T \pi_m (1 - \pi_m)}. \tag{3.2}$$

The D index became popular due to its easiness of computation and interpretation, although several limitations were addressed (Cortese, Falk, and Cohen 1976; James and Tauber 1985). First, it relies on the idea that the opposite of segregation is evenness of distribution. However, this does not mean that the spatial location of individuals will be unrelated to minority's characteristics (e.g. ethnicity). The index would be better defined if based on the assumption, as Jahn, Schmid, and Schrag (1947) did, that the opposite of segregation is a random distribution. Second, it does not respect the size invariance criterion. Third, it only satisfies a weak form of the transfer principle (Reardon and Firebaugh 2002): the index is insensitive to the redistribution of minority members among units with minority proportions above or below the city's minority proportion, while the transfers of individuals from units of higher to lower proportions may result in no change in D, but never result in an increase. Finally, the index is not suited for the investigation of multi-group segregation. Several contributions provided generalizations of the index in order to make it more suited to this task (Morgan 1975; Sakoda 1981), however the shortcomings persist (Reardon and Firebaugh 2002).

Closely related to D, the Gini index of segregation represents the area between the Lorenz curve and the diagonal of evenness, expressed as a proportion of the total area. It is defined in the range [0,1]. Note that the Gini index of segregation is related, but distinct, from the Gini coefficient of inequality.

$$G = \sum_{r \in R} \sum_{s \in R} \frac{t_r t_s |\pi_{rm} - \pi_{sm}|}{2T^2 \pi_m (1 - \pi_m)}. \tag{3.3}$$

Differently from dissimilarity index, the Gini index is sensitive to the transfer principle and somehow more difficult to compute. Its multi-group version is a desirable alternative to the multi-group dissimilarity index.

In the early 1970s an index of multi-group segregation based on the concept of entropy was derived from information theory (Theil and Finezza 1971; Theil 1972) and applied to the investigation of school segregation: the information index (H). Information theory is a branch of mathematics proposed by Shannon (1948) and applied to issues of communication between machines, signal processing and data compression. *Entropy* is a key concept in information theory, it can be conceived as a measure of how much information we can get from the happening of an event. Moreover, it can be interpreted as a measure of diversity of a population. If an event is very probable,

e.g. drawing a red ball from a pool populated only by red balls, no new information will added by the drawn of a new ball and entropy will be zero (the population is homogeneous). Conversely, if the pool is populated by balls all with a different color, the entropy will be maximum as the diversity of the population is complete. By generalizing the Theil and Finezza's (1971) definition of entropy we have:

$$E_{\mathrm{r}} = \sum_{m=1}^{M} \pi_{\mathrm{m}} \ln \frac{1}{\pi_{\mathrm{m}}}. \tag{3.4}$$

E_{r} is the level of entropy (diversity) of each areal unit. The H index is the average deviation of each areal unit's entropy from the system's (schools, city, metropolitan area) entropy weighted for the population share of each unit. In other terms, it is a measure of average variation of diversity across subareas.

$$H = \sum_{r \in R} \frac{t_{\mathrm{r}}}{T} \frac{E - E_{\mathrm{r}}}{E}. \tag{3.5}$$

The information theory index is defined in the range [0,1], where [1] is complete segregation. Despite its ostensible computational difficulty, several contributions have highlighted the desirable properties of this index. Zoloth (1976) has shown that H depends on the entire distribution of the reference population, while D relies only on a portion of it. Moreover, H can be decomposed in a between-group and within-group component, as also highlighted by Reardon, Brennan, and Buka (2002). Unlike the dissimilarity index, Theil segregation index satisfies the transfer principle, but it fails on compositional invariance since its value is also determined by the relative number of minority members (James and Tauber 1985). Finally, it is naturally defined as a multi-group segregation measure and does not require any mathematical adaptation. In the sub-sample of multi-group segregation measures, H is the only one that satisfies the transfer principle and is thus preferable to other measures (Reardon and Firebaugh 2002).

The dissimilarity index, the Gini index of segregation and the information theory index all belongs to measures of evenness of distribution. In 1954 Bell proposed an index[7] suited for the measure of exposure, namely the average degree to which, in an areal unit, members of a group are exposed to members of other groups (Bell 1954). The Bell's exposure index can be interpreted also as the probability of interaction between groups, or degree of potential contact. The popularity of measures of exposure is due mainly to Lieberson (1981).

$${}_{m}P_{n}^{*} = \sum_{r \in R} \frac{t_{\mathrm{rm}}}{T_{\mathrm{m}}} \pi_{\mathrm{rn}}. \tag{3.6}$$

As all the other indices of segregation already discussed, the exposure index is defined in the range [0,1]. It is worth noticing that the exposure index is not asymmetric, thus exposure of group M to group N is not complementary to exposure of

7. Bell's contribution was rooted in the work of Shevky and Williams 1949.

group N to M.

The measures presented in previous paragraphs are all "aspatial", meaning that they do not adequately account for the spatial relationships among residential locations. Upon the limitations of aspatial measures, since the '90s developed the third wave of debate on measures of segregation. This new wave of debate is primarly concerned with two problems: the *checkerboard problem* and the *modifiable area unit problem*. Both stem from the fact that aspatial measures are based on fictional areal units (census tracts), that do not necessarily represent meaningful social or spatial environments and that do not account for the internal disposition of individuals. The debate is informed by the principle that using individual's exact location would be a better solution to overcome these problems.

The checkerboard problem concerns the fact that aspatial measures concentrate only on internal composition of neighbourhoods and are not interested in their spatial relation. For an aspatial index of segregation it is irrelevant whether, let's say, a black neighbourhood is surrounded only by exclusively black or exclusively white units. A typical example: the D index is usually identified as a measure of segregation suffering this problem. Let's consider a checkerboard divided in black and white squares. Each square represents a neighbourhood characterized by its own composition. Any spatial arrangement of them will result in the same calculation for D. In other words, the index is not able to taking into account whether the squares were concentrated in to a single portion of the checkerboard or evenly dispersed. This arguement can be generalized to the broad category of aspatial measures.

The modifiable area unit problem stems from the fact that census data are collected, aggregated and reported for spatial units whose boundaries have no correspondence with meaningful social or spatial divisions of space. The problem is strictly connected with the ecological fallacy concept, that is inferring that results based on aggregate zonal data can be applied to the individuals within the zone itself. In fact, spatial associations created by the aggregation of data are not equal to real associations between individual data prior to aggregation (Dark and Bram 2007). Moreover, MAUP involves the size (scale) and variation of areal units. Aspatial indices as a particular case of statistical analysis are sensitive to these changes: indices for the same urban aggregate may vary consistently if we are considering tracts or tracts aggregation, and if we are relying on tracts defined in a way or in another. In order to solve these issues, spatial indices implement a Kernel Density function that describes the spatial proximity between all pairs of points (latitude and longitude) in a given region R. This function can take different forms according to different definitions of the local environment, relevance and weight given to certain points (condominium vs independent house), impact of features (i.e. highways vs narrow roads). We have a special case when census data are available in the aggregate form instead of individual level, a special case is made for the spatial proximity function: the group composition of each point in each areal unit is conceived as identical to group proportion in the areal unit, independently from how population groups are

actually distributed inside the tract.

Reardon and O'Sullivan (2004) have tested the spatial versions of the dissimilarity index, the information theory index, and the exposure index with respect to the classical dimensions highlighted by James and Tauber (1985) and the modifiable area unit problem. The spatial version of both H and P are identified to have the desirable mathematical properties of reliable measures of segregation.

Finally, when coming to measuring segregation by continuous variables, variation is typically measured by some index of the dispersion of distribution like variance or standard deviation. Variance can be conceived as the most basic and simple measure of segregation for some random ordinal variable (X). It is defined as the square deviation from the mean of our random variable.

$$Var(X) = \sum_{i=1}^{N}(x_i - \bar{x})^2). \tag{3.7}$$

A desirable property of variance is that it is decomposable in its between and within group variation. In fact, total variation of data can be decomposed into two additive components: the variability due to differences among groups; the total variability within each group. The former tells about the distance of a group from the mean of the population, the latter informs about the variation of the variable within the group. Let's imagine a school whose population of students is grouped in k classes. The variation between groups (classes) of a given variable is defined by the following:

$$Betweengroupvariance = n_k \sum_{k=1}^{K}(\bar{x_k} - \bar{x})^2), \tag{3.8}$$

where n_k is the sample size for each k^{th} group, $\bar{x_k}$ is the mean of the X variable in each group and $\bar{x}$ is the population mean. Between groups variance is minimized when each group has the same mean. It is worth noticing that a first problem arise: same mean can be achieved by different distributions. Here enters the withing group variation, defined as follows:

$$Withingroupvariance = \sum_{j=1}^{N}\sum_{k=1}^{K}(x_i - \bar{x_k})^2). \tag{3.9}$$

Within group variation is minimized when the variable of interest assumes the same value for each individual in group k, thus when a group is pretty homogeneous. The combination of the two components of variance, along with information on mean and median, might help in describing the general picture of phenomenon of segregation when the variable of interest is a continuous one.

3.4.2 Classification of occupations

Since the aim of this work is to investigate segregation by socio-economic factors and ethnicity in both the urban space and schools, a key element is represented by having a good degree of comparability between measures. How can we say that schools are more or less segregated than the sorrunding environment if the indices we compare are measuring different things? The need for a unique classification of occupations is rooted in the need to compare multigroup indices of segregation, since these indices are not invariant to the number of groups used for the calculation. E.g. if the index of urban segregation is computed on four occupational groups, and the index of school segregation is computed on three of these four groups due to different operationalization of the occupational scale, the two indices will not be comparable.

The following table shows the classification of occupations used by the ISTAT 2011 Population Census and the INVALSI 2014 Student Questionnaire. The two scales used are different and require some work to make them talk together.

ISTAT 2011 Population Census	INVALSI 2014 Student Questionnaire
1. Unskilled workers	1. Unemployed
2. Semiskilled manual workers	2. Housewife
3. Skilled manual workers	3. Manager, university professor, military officer
4. Agricultural workers	4. Entrepreneur/landowner
5. Skilled sales and service workers	5. Professionals (freelance and employee)
6. Clerical workers	6. Self-employed worker
7. Intermediate occupations	7. Teacher, employee and graded soldier
8. Highly specialized technical, intellectual and scientific activities	8. Blue collar, cooperative partner
9. Employers and managers	9. Retired
10. Armed forces	10. N.A

Table 3.5: Classification of occupations in data.

In order to ensure a certain degree of comparability the two occupational scales must be forced inside a common structure. The identified solution, presented in Table 3.6 is to derive a scheme of social classes according to the ESEC (Harrison and Rose 2006) classification. However, because these data are not classified in the same way nor using any detailed ISCO88 reference, this matching work is tentative as well as a limitation of this study. The scheme derived is a three class structure that discriminate between upper class, middle class and lower class. While there are few problems in the definition of upper class occupations in both classifications, problems arise with middle and lower classes. The ISTAT classification does not include categories other than higher grade white collar workers, this is a problem as it does not allow to identify a key component of middle classes: independent workers (self-employed). Similarly, with respect to the lower class, the INVALSI classification includes only

lower grade white collar workers, with no attention to skilled and semi-skilled blue collars.In the INVALSI database, parental occupation is defined with the dominance criterion (Erikson 1984): the highest occupational position between father and mother is selected as the occupational position of the household and associated to the pupil.

Class	ESEC	ISTAT classification	INVALSI classification
Upper class	Higher salariat Lower salariat	9 - 8	3 - 4 - 5
Middle class	Higher grade white collar workers Petit bourgeoisie or independents Petit bourgeoisie or independents Higher grade blue collar workers	5 - 6 -7	7 6
Lower class	Lower grade white collar workers Skilled workers Semi- and non- skilled workers	3 1 - 2 - 4 4	8
Excluded	Excluded	Specific variable in census data	1 - 2 - 9

TABLE 3.6: Class scheme adopted.

The analysis reported in this work will recurr also to the continuous variable ESCS presented in previous sections. Figure 3.1 presents the distribution of the index of socio-economic and cultural status (ESCS) for the four classes defined in this section. The distribution of ESCS by defined class is quite clear, and suggests that at we can clearly associate low values of ESCS to lower class, and higher values to the upper class.

3.4.3 Classification of nationalities

The situation concerning the classification of nationalities is less complicated, although not optimal, than that for the classification of occupations. On one side, ISTAT provides counts data of each nationality for *sezioni di censimento* if the number of occurances of that nationality is equal or greater than three, otherwise the count is provided but not the nationality. However, the information is nested in other two criteria: geographic area of country of origin and continent of origin. On the other side, INVALSI collects information about pupils and parents' nationality by grouping birth countries in four alternatives: (i) Italy; (ii) European Union;[8] (iii) European

8. This group includes Austria, Belgium, Bulgaria, Cyprus, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Latvia, Luxemburg, Malta, The Nederlands, Poland, Portugal, Romania, Slovak Republic, Slovenia, Spain, Sweden, United Kingdom.

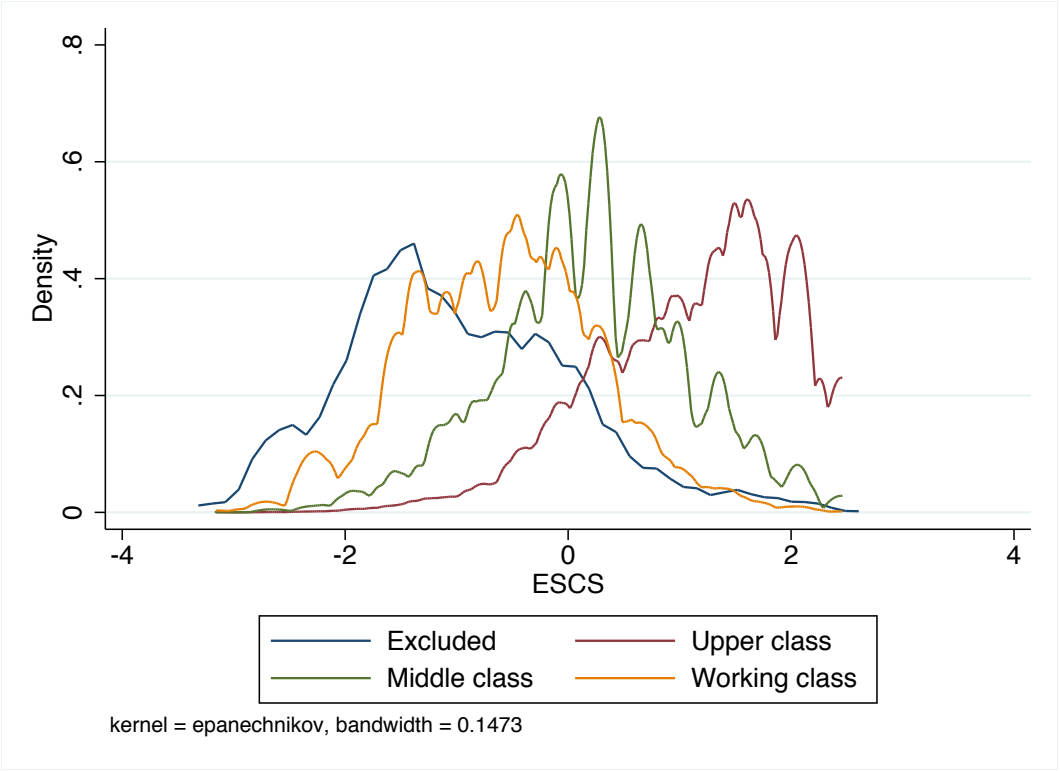

FIGURE 3.1: Kernel density estimation of ESCS index by social classes.

country outside the EU;[9] (iv) Other. Since it is not possible to further disaggregate the information obtained through this scale, the only option available is to fit the ISTAT geographic area of countries into the INVALSI one. This sub-optimal solution does not allow to distinguish among several groups of countries. The main limitation is that is not possible to uniquely identify individuals coming from different parts of Europe, e.g. those coming from Scandinavian countries or from Eastern Europe. An alternative is represented by building a variable that distinguish between Italian and foreign residents, without any specification of nationality.

3.4.4 School catchment areas and neighbourhoods

Districts were created in Italy in 1974 (Law 31 May 1974, n.416) as an administrative unit defined by municipalities. Each district had to have at least 10.000 inhabitants up to a maximum of 200.000, and it was governed by a board representative of school personnel and local economic and political actors. In 2003 (Law 27 December 2002, n.289) the personnel and functions of school districts were transferred to schools, resulting in their *de facto* elimination. So far, school districts are empty containers, but continue to exist as geographical entities to which schools sometimes refer to. Since districts were defined by municipalities, there is no database or public repository

9. This group includes Albania, Andorra, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Georgia, Iceland, Kosovo, Liechtenstein, Macedonia, Moldova, Monaco, Montenegro, Norway, Russia, Serbia, Switzerland, Turkey, Ukrain, Vatican City.

providing information on their borders. Although lists of schools and related districts can be found online, precise definition of borders is not available.

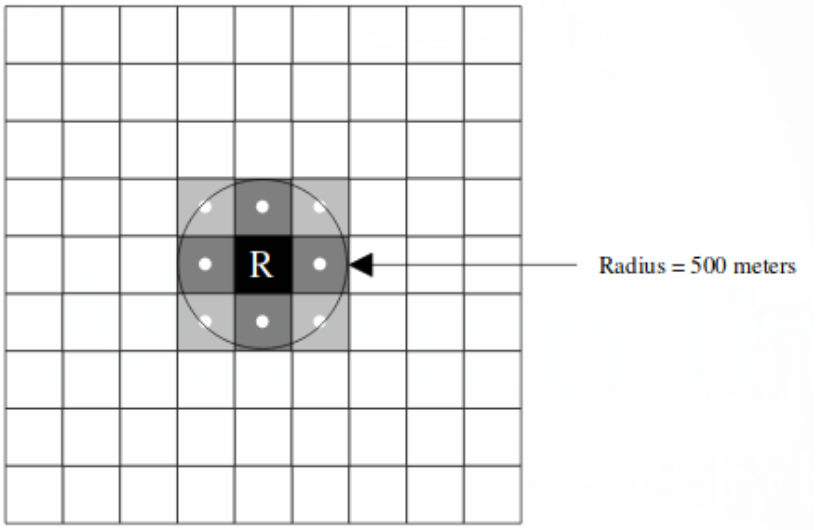

FIGURE 3.2: Neighbourhood definition (Barbagli and Pisati 2012).

To solve this issue, a dual strategy is used. First, the segregation profile of districts is computed using the *sezioni di censimento* as areal units. Once each school is inserted into the corresponding *sezione*, an egocentric local environment is drawn. Each census tract is a polygon characterized by a centroid, the geometric center of a plane figure computed as the arithmetic mean position of all points in the figure. The aggregation is defined as follows: each *sezione* hosting a school is considered as the center of a circle constructed using alternative radii. All the *sezioni* with a centroid that follows inside the circle, are considered in the aggregation and define our school district (Figure **??**). Neighbourhood measures are then calculated on the total aggregated *sezioni*. An alternative solution is presented in Figure 3.3. Instead of using the *sezioni di censimento*, our egocentric local environment relies on punctual data of pupils residing in the school district.

In the figure, school 5 is surrounded by concentric circles with different radius. As far as we do not know the exact extension of each district a possible solution is to build fictional districts characterized by different radius. Pupils located in the circle are retained as population of the school district irrespectively from census tracts borders. Differently from the solution highlighted, districts in the real world host more than a single school. Accordingly, our fictional character should be renamed as "school catchment areas". An alternative solution might look at Voronoi polygons, a way of partitioning a plane into regions based on distance to points (schools in our case) in a specific subset of the plane. This solution is not adopted in this work, but its features suggest controlling for school distance and catchment areas overlapping in the process of districts definition.

Both approaches presented in this section are used in order to get a better understanding of the phenomenon and the goodness of the measurement processes.

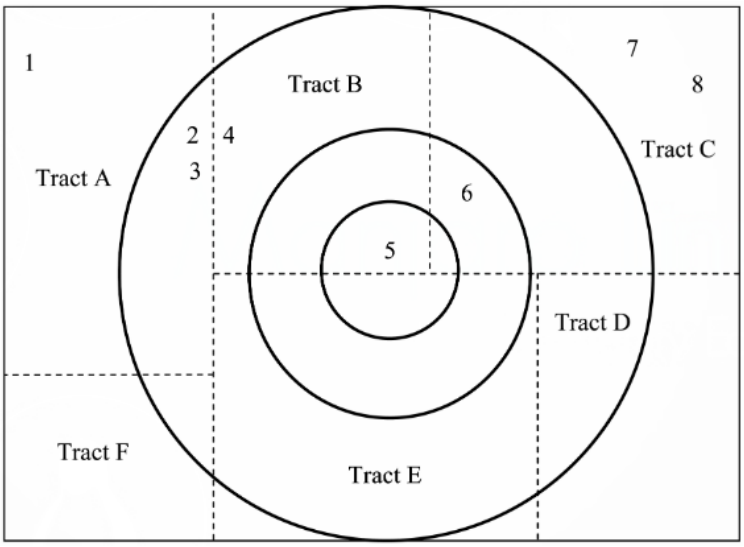

FIGURE 3.3: Districts as egocentric local environment (Lee et al. 2008).

3.5 Conclusions

In this chapter I have discussed the methodology to address the following research questions:

- Which is the level of segregation of school districts and schools in Italy? Which are the determinants of districts' segregation?
- How is district composition reflected in school composition? What makes families more prone to enroll their offspring in a school different from the closest public school?
- Is there a significant link between the degree of segregation of neighbourhoods and schools/classes on the performance in achievement tests? Is it heterogeneous over pupils?
- Are those coming from segregated schools/classes more inclined to make any school transition? Is it heterogeneous over pupils?

The database used to address these question is the result of a merging procedure between the ISTAT - Population Census (2011) and INVALSI - Survey on learning (2015). The population census contains information on the distribution of socio-economic and demographic characteristics of the population across *sezioni di censimento*, the smaller areal unit at which data are aggregated. The INVALSI database contains information on grade 8 pupils, a score of test ability in Italian and Mathematics along with demographic and socio-economic background information. Through the help of INVALSI, it has been possible to merge these two databases by a geo-referencing procedure. Residence addresses of pupils have been inserted in *sezioni di censimento* and connected to the urban space. Two are the main limitations of this database. The first is given by the methodological issues deriving from the *sezioni*'s

boundaries. On one hand, these boundaries are inconsistent across census waves and make time-variant analysis of urban segregation very difficult. On the other hand, the high variability in the size of *sezioni*'s population contributes to difficulties in the proper estimation of urban segregation. The second is due to the high amount of random missing values in variables describing family background, generating a snowball effect on the reliability of statistical models.

One of the aims of this work is to investigate whether the socio-economic and ethnic composition of school districts reflects in the school population, since information on these characteristics is recorded in both ISTAT and INVALSI databases in different ways, a strategy that allows for comparison is needed. The only available solution is the one presented in this chapter: to reduce both occupational classifications and nationalities records at the minimum common level. For the specific case, this means to recode both occupational classifications into a three class scheme (service class, middle class, working class) with the addition of the excluded-never worked category. Although reducing variability and complexity of the phenomenon, this strategy is valuable as far as comparisons between school and neighbourhood composition are at the centre of the research questions.

Instead of looking at the phenomenon of school segregation in the whole country, I've decided to focus on schools from a restricted sample of urban environments (Section 3.3). I've selected urban cores of economic functional areas as defined by ISTAT's Local Labour Systems among those with the highest share of immigrant population and unemployment rate by geografica area.[10] This resulted in a sample of 35 municipalities that account for 19.19% of the Italian population and about 23% of grade 8 students. The interest in the municipal scale is due to the nature of the school segregation phenomenon: research on school choice (Kristen 2008) has shown that parents evaluate primarly a limited set of short-range school alternatives, usually a single alternative to the closest school to home. Moreover, despite the Italian enrolment system allows pupils to (theorethically) move among schools, school boards have some degree of discriminatory power in accepting applications, and school proximity is usually a criterion of selection. Finally, focusing on urban cores of functional areas instead of all Italian municipalities allows to investigate sizable school systems. An alternative choice, not used in this work, should be that of considering the entirety of functional areas. This solution might provide more evidence on between-municipality flows in an area characterized by a highly connected labour market, as well as testing whether labour market flows are reproduced at the school level.

In order to investigate and compare the composition of the selected urban areas, school districts, and schools, I rely on measures of segregation presented in this chapter and on ethnic and socio-economic composition. The aim is to combine different measures in order to produce a segregation profile of the environments of interest. The Dissimilarity index D in both its two-groups and multi-groups version, the Information theory index H, the Exposure index P and its normalized version are used for

10. North-West, North-East, Centre, South, Islands.

this aim. Combining different measures can help in overcoming the main limitations associated to each single measure. On one side, the dissimilarity index; it is the most widely used measure of segregation for urban environments, however several criticisms have been raised. Reardon and Firebaugh (2002) shown that the index fails to satisfy several desirable mathematical properties, the most important of which is the transfer principle: segregation is reduced if an individual of group m moves from one areal unit to another where the proportion of individuals of that group is lower than the starting unit. On the other side, the Information theory. Zoloth (1976) suggests it as a measure of school segregation, while Reardon *et al.* suggests to recurr to this index due to its mathematical properties (Reardon and Firebaugh 2002). However, recent contributions (Abascal and Baldassarri 2015; Roberto and Hwang 2016) have criticized the use of the H index as it compares the diversity of a local unit to the overall diversity of a region rather than measuring differences in proportions of groups at the local and general level. The authors suggest that heterogeneity indices tend to flatten the hierarchical relationship between groups in the urban space. As a consequence we might expect lower levels of H, pushing our interpretation towards the presence of "less segregated" environments. Finally, school composition is analyzed with respect to previous indices and to the variability between and within classes of the index of socio-economic and cultural capital (ESCS) provided by INVALSI.

Chapter 4

Residential and school segregation in urban Italy

[1]

A substantial body of research investigates the link between school and residential segregation, finding different evidence depending on the selected country. For example, a major result in studies on the US is that school composition reflects district composition (Farley and Tauber 1968; Reardon and Yun 2005), while studies on European countries report mixed evidence. Such difference in the relationship between school and residential segregation has been explained by two main reasons: first, residential segregation seems to be more moderate in Europe than in American cities (Musterd 2005); second, residence and attendance of the school of the related catchment area are not so strictly linked in European cities as they are in the North-American case. These two facts might be the result of several elements. They might be related to the city extension, to the functional characterization of cities or to historical patterns in city development that include variations in the institutional context.

As highlighted by Jenkins, Micklewright, and Schnepf (2008), school segregation is an actual phenomenon across European countries even if not as large as in the US. However, this seems to be more related to parental choices than to urban segregation. Results supporting this idea can be found in investigation on Copenhagen's metropolitan area (Rangvid 2007), Sweden (Andersson, Östh, and Malmberg 2010), England (Burgess, Wilson, and Lupton 2005), and France (Oberti 2007b). These studies found that children from ethnic minorities and low socio-economic background are more segregated in schools than in their neighbourhood. Moreover, Burgess, Wilson, and Lupton (2005) highlight that the ratio between school and neighbourhood segregation increases in areas with higher population density.

Chapter 1 has shown that few contributions have been devoted to the study of levels of segregation by ethnic and socio-economic background in Italian school. Torri and Vitale (2009) investigated the mechanisms of reproduction of poverty inside one

1. A very preliminary version of the analysis presented in this chapter have been presented at the 2017 - ECSR Annual Meeting.

relatively deprived neighbourhood in Milan, Quarto Oggiaro, with a particular attention to dynamics of school segregation. Pacchi and Ranci (2017) have recently investigated school segregation in primary and lower secondary school in Milan, finding weak segregation on both ethnic and socio-economic basis. Parental choices are identified to be the main driver of pupils allocation among schools. Finally, Barberis and Violante (2013) investigate school segregation by comparing four metropolitan areas: Milan, Bologna, Rome and Naples.

4.1 Research questions and hypothesis

The aim of this chapter is to establish the phenomenon of segregation in Italian lower secondary school at different levels (classes, schools, districts), as well as to widen knowledge on school segregation in Italy and to address similarities with other national cases. The rationale for focusing on lower secondary school is manyfold. First, the Italian schooling system is not characterized by tracking in compulsory education, however hidden tracking may be in place and consequent predatory strategies by families in order to secure "better schools" to their offspring might emerge. Second, lower secondary schools are bigger and less spatially diffuse than primary schools. Italy has about 18,000 primary schools against about 8,000 of lower secondary. Thus, families have to make choices about enrolment and this makes it a phenomenon of interest. Third, data on lower secondary school are more complete and robust than data on primary school. This is due to the fact that the INVALSI test of grade 8 is part of the final examination and thus completed by all students but those not admitted to the final exam. Since its introduction, the INVALSI test has gone through a wide dispute by both students (in upper secondary) and parents. The grade 5 test data we collected are severly affected by this dispute, making them not suitable for reliable statistical analysis. Results are analysed with respect to available national and international research.

This chapter aimes at investigating the following research questions:

- Which is the level of residential segregation by socio-economic and ethnic characteristics in Italy?

  According to the existing literature presented, the hypothesis to be tested is that segregation in Italian cities is relatively low. Residential segregation by ethnicity is expected to be higher than that by socio-economic factors.

- Which are the structural determinants of segregation in the urban space?

  We aim at applying the ecological model by Farley and Frey (1994) to the Italian case and investigate whether it is possible to identify common causes in structural elements of urban areas for segregation.

- Which is the level of segregation of classes, schools, and school districts by socio-economic and ethnic characteristics in Italy?

We expect a medium level of segregation for schools, higher than segregation found for urban areas although the two aggregates are difficut to compare.

- How do levels of segregation vary across geographical levels?

  On average, residential segregation by socio-economic status is expected to be low but higher in the south. Barbagli and Pisati (2012) estimate that residential segregation by social status in the eight largest Italian cities is less than 10% according to the H index, but highlighted variation along the north-south dimension. We expect higher levels of segregation in cities of Southern Italy. In fact, immigration is a relatively new phenomenon in Italy, concentrated especially in the more developed part of the country (northern Italy). Moreover, we expect processes of assimilation of ethnic minorities to be different across the country, thus leading to different segregation levels.

- Are private schools more or less segregated than public schools?

  We expect private schools to be severly more segregated than public schools, especially by socio-economic background. We expect private education to be quite homogeneous its composition by socio-economic background. In fact, Italian private school is characterized by relatively high tuition fees that only families with relevant economic resources can afford.

- Are schools more or less segregated than the related district?

  If a process of school segregation is in place, we should expect that schools' population is more homogeneous than the related district on both the socio-economic and ethnic dimensions.

4.2 Data and methods

As discussed in Chapter 3, two databases are used for the investigation of residential and school segregation. The 2011 - ISTAT Population Census is the latest available wave of the Italian census, that recently turned to a permanent structure characterized by annual collection of data, and provides characteristics on social position and immigrant background of respondents. The census database presents 366,863 *sezioni*. From this database, the Local Labour Systems (LLSs) related to the 35 selected municipalities have been extracted. The whole LLSs account for 116,806 areal units, while the urban cores for 67,042. Both urban cores and entire functional areas are used in order to investigates residential composition. Count data on foreign residents by nationality in each areal unit have been provided by ISTAT. When the count is less than three units, the information of the country of origin has been deleted, but not that for the macro-area and the related continent. Thus, it has been possible to categorize individuals in the common variable presented in Chapter 3. The 2015 - INVALSI Survey on Learning of grade 8 is used to investigate school composition. It contains data at the individual level, where each observation is a pupil of grade 8. The database

for the selected municipalities is build of 113.700 individuals nested in 1.098 lower secondary schools.

Urban and school composition are investigated through several methods. The aim is to provide a segregation profile based on differend measures, in order to overcome limitations related to a single measure. This task is carried out by referring to indices of segregation analyzing eveness of distribution of groups in the selected environment and exposure between groups. The jointly investigation of these two aspects of segregation provides a detailed map of the phenomenon. Evenness of distribution is investigated through the Dissimilarity Index (D), and the Information Theory Index (H), while exposure through the Exposure Index (P^*). Both indices varies between [0] and [1], where the former denotes absence of segregation/exposure while the latter denotes environments that are perfectly homogeneous and groups perfectly exposed. The Dissimilarity index expresses the average share of group m individuals that must be relocated from areal units where the group is over-represented to units where it is under-represented. The Information theory index expresses the average diversity of areal units from the overall region. The exposure index expresses the extent to which members of different groups share the same residential area.

It is worth noticing the the smallest areal unit at which data are gathered and aggregated by ISTAT, the *sezioni*, vary by dimension and population across the country. In fact, they are defined *ad hoc* on each municipality borders. As a consequence, they are very irregular polygons. Moreover, *sezioni* are not defined according to any lower population threshold, and a large proportion of them has few inhabitants. A fact that threatens the computation of reliable measures of segregation. As a matter of fact, our estimates of residential segregation will always be affected by the two classical problems of aspatial measures presented in Chapter 3: the checkerboard problem and the modifiable areal unit problem. Another possible threat is posed by confidentiality concerns leading to not-provided information for the variables of interest if the *sezione* hosts less than three individuals in a given category. However, for larg urban aggregates as those selected this is likely to have a very small influence. Finally, school composition is investigated also through variance decomposition of the average school and class ESCS. This method, already used by Cardone, Falzetti, and Severoni (2015), is implemented in order to define a categorization of types of segregation experienced by schools.

Municipalities have been divided in five groups according to their population:

(i) villages, with a population lower than 5,000;

(ii) small towns, with a population between 5,000 and 20,000;

(iii) medium towns, with a population between 20,000 and 40,000;

(iv) big towns, with a population between 40,000 and 500,000;

(v) metropolis, with a population higher than 500,000.

	Village	Small city	Medium city	Big city	Metropolis
% Upper class	.081	.103	.123	.146	.181
% Middle class	.244	.282	.295	.319	.335
% Lower class	.292	.281	.248	.222	.194
% Excluded	.224	.243	.276	.277	.278
% EU	.017	.017	.015	.014	.017
% non-EU	.010	.016	.015	.014	.006
% Other	.020	.029	.025	.029	.053

TABLE 4.1: Average composition of Italian municipalities by dimension.

4.3 Empirical results

4.3.1 Residential segregation - Italy

As described in the previous section, Italian municipalities have been divided into five categories. Table 4.1 reports the composition of Italian municipalities by population size. The presence of upper class individuals seems to be positively correlated with city population (correlation of about .18), as higher shares are present in larger towns. On the other hand, the presence of the working class seems to be negatively correlated with city population (-.08). On average, ethnic minorities are evenly distributed across the five aggregates, with a slightly higher presence in metropolis. In general terms, the hierarchy of socio-economic groups in the urban space is related to two latent factors: the structure of the labour market in that city and the housing market. With the end of the Fordist era industry and services have relocated them-selves across the urban space, between city cores and periurban areas (Lever 2001), changing the urban geography of labour demand. The main effect of this relocation process have been a general polarization of the urban space characterized by the duality between high-waged workers in urban cores and low-waged workers in the periphery. The diachronic investiation on larger Italian cities by Barbagli and Pisati (2012) has provided some evidence on this issue, highlighting the expulsion of low-waged workers in the same direction as theorized by post-fordists.

Table 4.2 reports average Dissimilarity (D) and Information theory (H) indices for residential segregation by population size for socio-economic position and ethnic background. According to the Dissimilarity index, the upper class appears to be the most segregated socio-economic group in all the typologies, followed by the lower class, and the middle class. In fact, between 22% of the upper class in small cities and less than 40% in metropolis have to be relocated between areal units in order to obtain eveness of distribution among all the municipality' *sezioni*. The D index for the middle class hovers around 20% in all urban typologies, while that for the lower class is around 24%. Values for both these socio-economic groups have smaller variability than that for the upper class. Finally, the Excluded are the less segregated group, scoring low values of Dissimilarity and suggesting they are quite distributed in the

	Village	Small city	Medium city	Big city	Metropolis
Upper class (D)	.271	.219	.280	.342	.394
Middle class (D)	.240	.175	.192	.213	.186
Lower class (D)	.237	.183	.218	.263	.277
Excluded (D)	.112	.096	.115	.147	.163
Italians (D)	.285	.290	.348	.405	.436
EU (D)	.395	.412	.489	.591	.545
Non-EU (D)	.342	.463	.547	.645	.687
Other (D)	.396	.400	.463	.529	.510
Upper class (H)	.090	.062	.089	.127	.151
Middle class (H)	.101	.064	.064	.071	.044
Lower class (H)	.105	.067	.072	.089	.081
Excluded (H)	.025	.012	.018	.027	.032
Italians(H)	.084	.074	.102	.140	.175
EU (H)	.125	.112	.150	.211	.186
Non-EU (H)	.114	.141	.191	.255	.303
Other (H)	.124	.112	.146	.195	.215

TABLE 4.2: Mean values of segregation in Italy by city population categories.

urban landscape. Since the data are population counts, a possible explanation is that members of this group, that includes unemployed and never worked, are spouses of members of the other groups.

The segregation profile of socio-economic groups by city size is confirmed by the Information theory index. However interpretation is different: H measures the average diversity of each areal unit from the overall region, and low values indicate that areal units are as diverse as the overall region. Values of H are sensibly smaller than D, but the hierarchy of space is relatively respected. This is not true for the two smaller aggregates (villages and small cities), where the ranking of segregation change substantially. In this case, differences in the index value are really small and change in rank probably not reliable.

About segregation by ethnic groups, Table 4.2 reports high values for D and low or moderate values for H. The majority group (Italians) is the less segregated one, although characterized by high values of dissimilarity (especially in larger aggregates). High levels of segregation are registered for all other ethnic groups. On average, Europeans from non-EU member states is the most segregated group, followed by Europeans from EU-member states, those coming from continents other than Europe and Italians. This ranking is confirmed by the H index, although values are flattened and lowered with respect to D. On average, about 40% of Italians in large urban areas and 30% in the small one need to be relocated in order to obtain an even distribution. This figure is higher for minority groups: i.e. more than 60& of Europeans from non-EU member have to be relocated to obtain an even distribution in the related region.

In order to clearly grasp the phenomenon, additional investigation is needed,

	(1)	(2)	(3)	(4)
(1) Upper class		.686	.685	.622
(2) Middle class	.241		.499	.420
(3) Lower class	.221	.447		.400
(4) Excluded	.256	.473	.501	
(1) Italians		.017	.012	.022
(2) EU	.892		.207	.326
(3) Non-EU	.704	.257		.294
(4) Other	.882	.279	.182	

TABLE 4.3: Average exposure index P, Italy.

for instance by looking at exposure rates as measured by the Exposure index (P). Table 4.3 reports the two matrices of exposure. In the top part of the table, exposure between socio-economic groups, in the bottom part exposure between ethnic groups. The table suggests that the upper class is sharing residential areas with all other groups, the probability for an upper class gentleman to meet a member of other groups is higher than 60%. Despite being the less segregated group, the middle class is sharing more areal units with the lower class and the excluded rather than the upper class. In fact, the first column of the upper part highlights general low exposure to the high-waged workers. The situation is different for ethnic groups. The very low levels of exposure of Italians to other groups are driven by the low percentages of minority groups. Exposure between minority groups hovers in the range 20-30% and suggests they are not occupying the same urban areas.

Finally, territorial variation of segregation is described in the following figures, where the distribution of municipalities according to their Dissimilarity index for upper class *vs* all other groups and foreigners *vs* Italians are presented by geographic area and population size. Metropolises are not reported in this figure, as the group is made only of six towns[2] and they are discussed in the following paragraphs. Figure 4.1 highlights a pattern of levels of segregation by town dimension that is consistent in all the five geographic areas: larger towns are characterized by higher average levels of segregation. However, villages are the group with more internal variation, probably due to computational issues. There are not huge territorial differences in segregation levels by population size. Cities in the north are experiencing similar levels of segregation, while these are on average higher in the South and Islands, especially for medium and big urban aggregates.

The situation with reference to segregation by ethnic groups is shown in Figure 4.2, where the phenomenon is more evident. Variation in levels of segregation by nationalities groups across the country is higher than variation in segregation by socio-economic factors. Villages and small cities have similar levels of segregation irrespectively from their geographic area. When coming to medium and big cities, towns in the south and islands presents higher levels of segregation.

2. Torino, Milano, Genova, Roma, Napoli, Palermo.

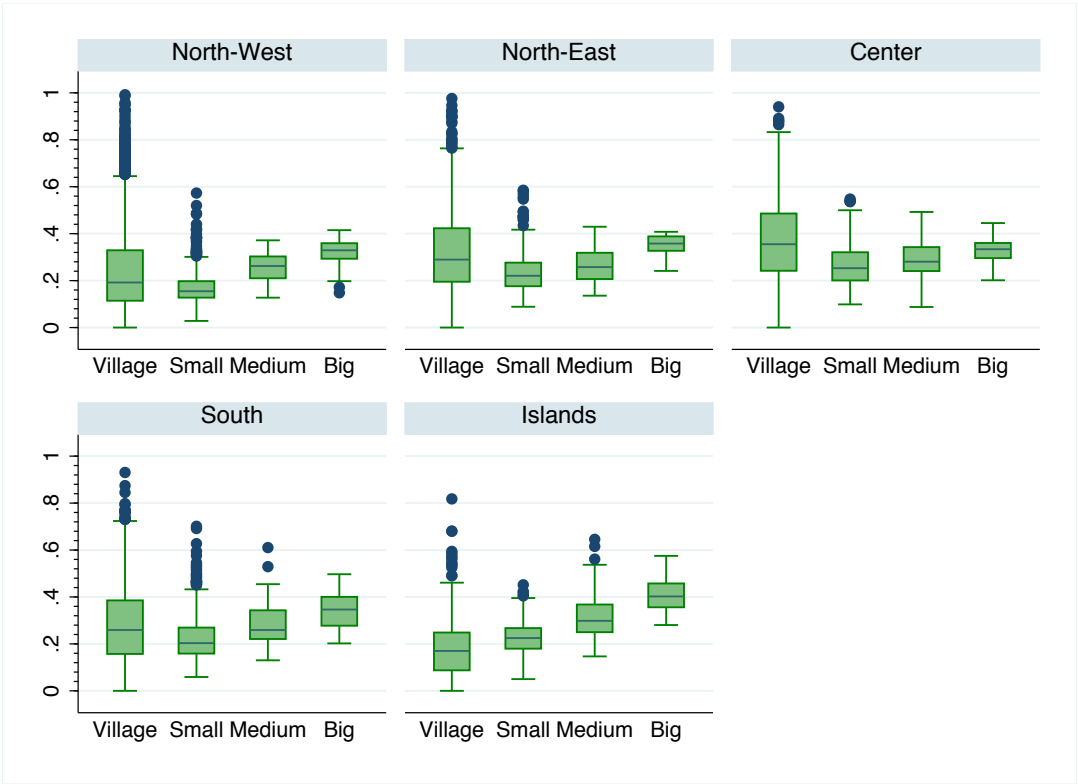

FIGURE 4.1: Boxplot of D_{sec} index by geographic area and town population.

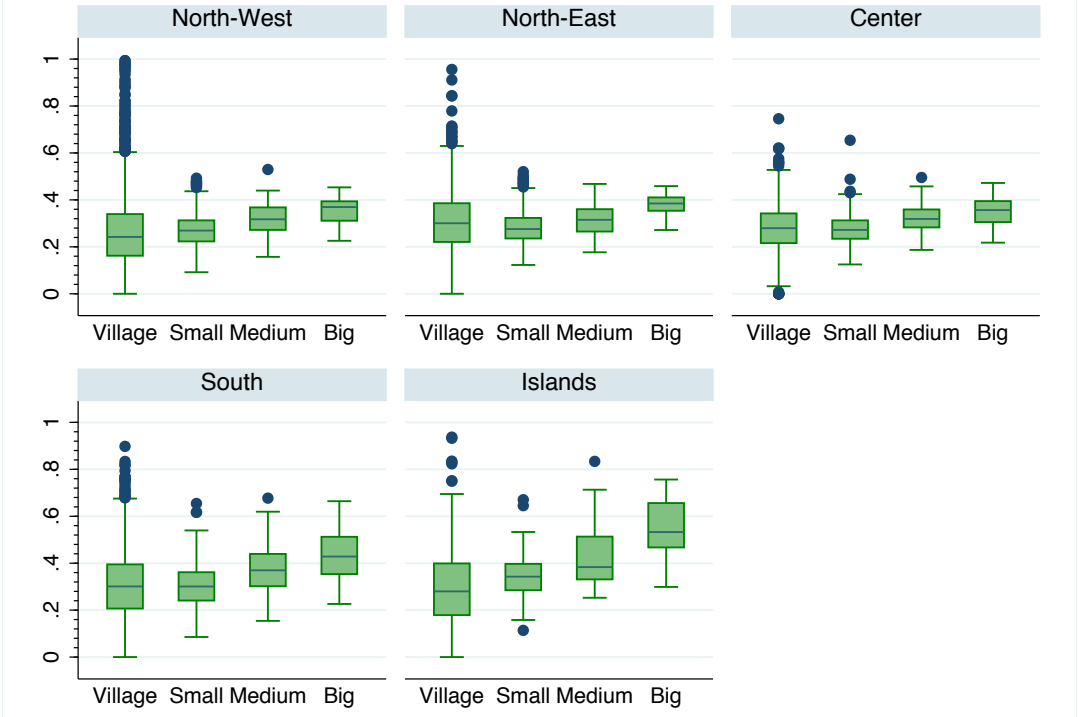

FIGURE 4.2: Boxplot of D_{nat} index by geographic area and town population.

4.3.2 Residential segregation - selected LLSs

Urban composition for the selected sample has been analysed by recurring to Local Labour Systems, functional areas characterized by a high degree of interconnection of local labour markets. Each LLS has been divided in the urban core, represented by the 35 selected municipalities, and a hinterland, represented by the other municipalities in the system. The 35 LLSs involve 1,145 municipalities, accounting for 21.7 millions inhabitants and divided by size as follows: 527 Villages; 445 Small towns; 101 Medium towns; 66 Big cities; 6 Metropolis. The sample of urban cores is made of one medium city, 32 big cities and 6 metropolis, for a total population of less than 12 millions inhabitants. Concerning the active population, the sample encompasses less than 40% of the total active and excluded in the labour market. The sample presents a slightly over-representation of upper and middle class, more concentrated in large urban aggregates characterized due to a major presence of high-skilled jobs, and under-representation of the working class. Also the foreign population is slightly over-represented in the sample (8.8% against 7.5% for the whole country).

The selected LLSs can be partitioned in a urban core, the larger urban aggregate in terms of population and labour market as defined by ISTAT, and the related hinterland. On average, urban cores are characterized by higher shares of upper (17.4% vs 9.8%) and middle (33.5% vs 29.5%) class, while the working class resides more in the hinterland (21% vs 26%) where the housing market is more affoardable. The excluded (unemployed and never-worked) are evenly distributed in the two aggregates (24%). Italy is a new comer as a destination country of migratory flows. This reflects in the low figures of the immigrant population in Italian cities. In 2011, individuals from EU member countries account only for 1.5% in the urban core and 2.2% in the hinterland, while the same figure for individuals from non-EU member states is about 2% in the core and 1.2% in the hinterland. Finally, individuals from all other countries account for 4.4% of the core pupalation but only for 2.2% of the hinterland.

The core-hinterland partition can be analyzed in a different way, by looking at urban aggregate classified by population size. In fact, the selected urban cores all fall in the Big city and Metropolis categories, while hinterland municipalities distribute in the other three. Once again, the higher presence of service and middle class in larger urban aggregates is certified, and linked to both labour and housing market characteristics. On the contrary, working class shares are higher in hinterland municipalities, especially the smallest, where the housing market is more accessible. Slightly higher shares of those coming from EU member states are reported in villages and small cities (less than 20,000 inhabitants), but these are really small figures. Interestingly, immigrants from continents other than Europe are more concentrated in metropolis and big cities rather than small ones.

Through segregation measures, Table 4.4 investigates eveness of distribution in the urban space by social class and migratory background. According to both Dissimilarity (D) and Information theory (H) indices, LLSs' cores are characterized by a

Groups	*D*		*H*	
	Core	Hinterland	Core	Hinterland
Upper class	.365	.247	.142	.076
Middle class	.213	.201	.069	.078
Lower class	.289	.206	.098	.078
Excluded	.164	.107	.032	.018
Italians	.410	.291	.154	.008
EU	.561	.405	.201	.119
Non-EU	.617	.433	.256	.136
Other	.503	.398	.202	.115

TABLE 4.4: Mean values of segregation for LLSs' cores and hinterland.

more disproportional distribution of social classes than the related hinterland.[3] The Dissimilarity index highlights a ranking in the relationship between socio-ecoxnomic groups. The service class is more "segregated", and about 36% of its incumbents should be redistributed in the city in order to obtain an even distribution. The working class is the second most segragated group, about 29%, followed by the middle class (21%) and the excluded (16%). This hierarchy is similar in the hinterland municipalities, although there is no difference in segregation measures between middle and lower class. In general terms, levels of segregation are substantial, but far from the high levels registerd for US metropolis. The same hierarchy is found by the H index, although interpretation is different: Theil index expresses the average lack of diversity. Thus, at a first glance at H index we might conclude that *sezioni* in the selected municipalities are charaterized by high diversity with respect to the overall city. An even sharp result for hinterland municipalities, where the heterogeneity index highlights high levels of diversity.

Results for segregation by ethnic group show general high levels of segregation in urban cores, while substantially less pronounced in the hinterland. Europeans from non-EU member states[4] represent the most segregated ethnic group in both spatial partition (about 61% in cores and 43% in the hinterland), followed by Europeans from EU-member states (56% in cores and 40% in the hinterland), individuals from continents other than Europe (less than 40% in both partitions) and Italians (less than 30% in both partitions). In terms of diversity (H), it seems that *sezioni* are characterized by higher levels of average variation of ethnic rather than socio-economic composition.

Table 4.5 and 4.6 present two matrices of exposure for core and hinterland municipalities. The top part of each table presents exposure indices between socio-economic groups, the bottom part for ethnic groups. Results of the two tables present a common structure: high rates of exposure of the upper class to other socio-economic groups (first row), compared to low levels of exposure of these groups to the upper class (first

3. Indices for each group are computed using the dichotomized version: group of interest *vs* all other groups together.

4. See Section 3.4.3 for detailed list.

	(1)	(2)	(3)	(4)
(1) Upper class		.563	.387	.459
(2) Middle class	.290		.331	.376
(3) Lower class	.320	.534		.480
(4) Excluded	.347	.534	.429	
(1) Italians		.015	.019	.041
(2) EU	.927		.278	.452
(3) Non-EU	.892	.239		.426
(4) Other	.850	.172	.184	

TABLE 4.5: Average exposure index P, sample (cores).

	(1)	(2)	(3)	(4)
(1) Upper class		.707	.666	.617
(2) Middle class	.233		.456	.397
(3) Lower class	.245	.496		.420
(4) Excluded	.278	.513	.498	
(1) Italians		.022	.013	.022
(2) EU	.934		.234	.351
(3) Non-EU	.818	.339		.355
(4) Other	.917	.334	.213	

TABLE 4.6: Average exposure index P, sample (hinterlands).

column); high exposure of ethnic minority groups to Italians and low exposure rates of Italians to ethnic minority groups. In general terms, hinterlands present higher levels of exposure between groups of both dimensions. Thus, tables are suggesting that on average core municipalities are characterized by a more rigid hierarchy of the urban space rather than the related hinterland. Here, different groups have higher probabilities to share a common residential area.

Results presented in this section suggest that the upper class is sharing some portion of the urban space with other socio-economic groups, although this might happen in different ways. The combination of medium levels of exposure and segregation might suggest that the areas where the upper class is living (usually centres and residential areas) are characterized by some degree of mix, or shared at least with one group. In general terms, the upper class of the selected sample is living in places where it is more exposed to the middle class (56%), while the probabilities to encounter a member from the lower class are less pronounced (38%). This last case might be that of areas subject to some process of "gentrification", where the housing market is more affoardable than city centres. Looking at the hinterland, the situation is more pronounced: there are less service class individuals, they are more evenly distributed than in urban cores and have high probabilities to encounter members of other socio-economic groups. Howerver, the opposite is not true due to the high disproportion in absolute figures. In general terms, the selected LLSs are characterized

by hierarchy in the occupation of the urban space. The upper class is living especially in cores, while the lower class in the related hinterlands. The middle class occupies an intermediate position, but this is probably due to the composition of this group and a more in-depth investigation should bring light on this issue as the upper tier of middle class occupations might live more in cores than in hinterlands. Finally, excluded are characterized by low levels of segregation and high levels of exposure, suggesting they are quite distributed in the urban landscape. Since the data are population counts, it might be that members of this group, that includes unemployed and never worked, are spouses of members of the other groups.

Results of this descriptive analysis highlight ethnic segregation to be a more substantial phenomenon in Italian cities rather than socio-economic segregation. Non-EU member states migrants and those coming from other continents are more prone to experience contact with other minority groups in the hinterland rather than in urban cores. On the other side, Italians are experiencing extremely low levels of contact with ethnic minorities. Although surprisingly, this result is also rooted in methodological issues. First, despite segregation measures values of the exposure index P depend on the relative size of the groups being compared. Since figures of ethnic minorities are very small, this is part of the story. Second, all statistics presented in these sections are influenced by the modifiable areal unit problem. *Sezioni* as defined by ISTAT are very small units that vary substantially in population. A more reliable spatial aggregation might provide different results. Finally, additional investigation taking account of a different and more theory oriented partition of ethnic minorities is needed. As an example, the used scheme is not able to distinguish between migrants from rich and poor countries.

Table 4.7 provides detailed information on indices of segregation and exposure for all the municipalities included in the sample, along with relative frequences of these groups. Four indices have been selected: the Dissimilarity for upper class *vs* all other socio-economic groups, the Exposure index of upper to lower class, and the Dissimilarity and Exposure for immigrants *vs* Italians. The Information Theory index has proved to flatten and sometimes alter the hierarchical ranking of the urban space, thus it has been kept out from this summary table. Moreover, the selected partitions might better summarize the characteristics of the selected municipalities, describing the segregation of the upper class and its distance from the lower class, the segregation of immigrants and how much the share space with Italians. In the upper part of the table are listed the the six cities with more than 500,000 inhabitants, while the lower part presents results for medium and big cities of the sample. The former group presents three cities of the northern part of the country (Torino, Milano, Genova), the capital city (Roma) and two cities from the south (Napoli and Palermo). This group is characterized by substantial level of segregation for both the upper class and non-Italians, although in line with results from other municipalities. In this group, Napoli and Palermo are the most segregated cities, followed by Torino, Genova, Milano, and Roma.

City	D_{upp}	P^*_{upp}	D_{imm}	P^*_{imm}	% immigrants	% upper class	Geo. Area
Genova	.376	.376	.419	.820	.068	.162	NW
Milano	.350	.295	.365	.756	.139	.270	NW
Napoli	.471	.331	.503	.864	.043	.138	S
Palermo	.432	.730	.601	.816	.060	.169	I
Roma	.359	.323	.364	.783	.097	.200	C
Torino	.374	.418	.363	.773	.126	.202	NW
Agrigento	.393	.285	.580	.867	.035	.166	I
Asti	.408	.482	.406	.794	.089	.125	NW
Bari	.413	.310	.487	.900	.028	.167	S
Bologna	.328	.339	.284	.828	.116	.238	NE
Brescia	.363	.405	.388	.719	.144	.191	NW
Caltanissetta	.457	.278	.674	.777	.065	.092	I
Catania	.513	.345	.575	.878	.041	.152	I
Crotone	.400	.345	.479	.916	.037	.158	S
Firenze	.291	.374	.273	.818	.119	.240	C
Frosinone	.322	.427	.323	.903	.061	.147	C
Imperia	.299	.503	.311	.825	.115	.162	NW
L'Aquila	.230	.414	.340	.909	.039	.150	C
Latina	.369	.401	.402	.884	.046	.158	C
Mantova	.375	.336	.342	.814	.104	.201	NW
Modena	.346	.426	.410	.749	.119	.169	NE
Olbia	.424	.469	.432	.833	.069	.108	I
Padova	.326	.321	.361	.777	.119	.256	NE
Perugia	.331	.428	.333	.810	.113	.202	C
Piacenza	.365	.421	.367	.751	.135	.175	NE
Prato	.333	.564	.393	.722	.144	.135	C
Ragusa	.440	.356	.677	.724	.054	.140	I
Reggio nell'Emilia	.342	.496	.390	.726	.136	.157	NE
Rovigo	.366	.419	.414	.858	.067	.164	NE
Savona	.308	.468	.289	.875	.075	.163	NW
Teramo	.266	.458	.226	.934	.052	.172	S
Udine	.304	.360	.357	.780	.117	.223	NE
Venezia	.399	.379	.408	.807	.080	.161	NE
Verona	.357	.399	.384	.788	.108	.198	NE
Vibo Valentia	.355	.724	.414	.912	.042	.182	S
Total	.365	.387	.410	.820	.086	.174	

TABLE 4.7: Mean segregation and exposure levels of cities included in the sample.

The table shows that cities of the south are more segregated on both dimensions than cities of the northern part of the country. Socio-economic segregation is higher in municipalities with lower shares of upper class, suggesting concentration of this group in some parts of the town. As previously discussed, the upper class is not completely separated from other groups. In fact, medium levels of segregation (around 30-40%) are accompained by medium levels of exposure to the lower class. A particular case is Palermo, where the upper class lives in areas with high probabilities to get in contact with members of the lower class. An interesting case that call for specific future work on the spatial disposition of socio-economic groups.

Immigration is a relatively recent phenomenon in Italy, and especially in southern cities. The presence of a more dynamic labour market attract migrants to the northern municipalities, where the share of immigrant population is usually above the symbolic threshold of 10%. In the overall sample, only 8.6% of the Italian population in 2011 was foreigner. Immigrants are experiencing high levels of segregation, the dissimilarity index highlights high shares (about 40%) of the immigrant population to be redistributed among areal units in order to obtain a more homogeneous distribution. These numbers are far from values of ethnic segregation recorded in studies on American metropolitan areas, but are consistent with literature on other European metropolis (Malheiros 2002; Arbaci 2007). Moreover, immigrants are experiencing really high levels of exposure to Italians, a fact due to the relative figures of the two populations. It is probably true that immigrants resides in specific locations of the urban space, but exposure indices suggest that immigrants are sharing areas with the majority group. A particular case is represented by port cities, where migrants seems to concentrate in the old port zone (Barbagli and Pisati 2012; Scarpa 2015; Mazza, Gabrielli, and Strozza 2017) probably due to a more favorable housing market. The high levels of ethnic segregation in Palermo and Naples (and even Genoa) might be the result of processes of this kind. Industrial cities of the north, like Turin and Milan, are instead characterized by a more scattered distribution of migrants in the urban space, this might result in comparative lower, but still substantial, levels of segregation.

The average levels of segregation and their variation across the country calls for a more detailed investigated of spatial patterns both on metropolises and big cities, with attention to be devoted especially to southern cities. In fact, at the time of writing, little is known of spatial location processes of towns of southern Italy.

4.3.3 Structural determinants of urban segregation

Farley and Frey (1994) have proposed an ecological approach to the investigation of determinants of ethnic segregation that relies on the structural characteristics of the urban space. The rationale of this framework is rooted in Tauber and Tauber's (1969) ecological assumption that the structural characteristics of a urban environment, in conjunction with demographic trends, are able to explain levels and trends

in segregation. We rely on the original model, that is characterized by seven predictors: (i) geographic region; (ii) functional specialization; (iii) age of metropolitan area; (iv) percentage of buildings recently built; (v) population size; (vi) percent of minority population and minority income as a percent of majority group income. The framework has been shown to have validity for tracts and block groups (Lee et al. 2008), but has not yet been applied to the Italian case nor to residential segregation by socio-economic characteristics. The model is presented in the following equation:

$$y_{ji} = \alpha + Z_j\beta + e_{ji}, \tag{4.1}$$

where y_{ji} is the vector of segregation for municipality j and group of interest i, and Z_j is a vector of municipality characteristics. Following the discussion in previous sections, segregation is computed using the Dissimilarity index in its dichotomized form. Segregation of each group is evaluated against a reference category: the upper class for socio-economic groups, Italians for ethnic groups. According to previous research (Logan, Stults, and Farley 2004; Iceland and Scopilliti 2008; Lee et al. 2008), and to information available in the ISTAT census database, the Z vector includes log of population, percentage of over 65 years old, percentage of minority groups, percentage of blue collars, percentage of housing units built in the last ten years.[5] Italian municipalities are characterized by a huge variation in their population size. Big municipalities are an exception in the Italian landscape, where small towns are the dominant urban typology. However, the model fits better with municipalities that are able to grant at least 1000 members per group of interest. In our case this means no to be able to run the regression on segregation measures between Italians and other ethnic groups due to small sample issues. An alternative solution is to run the model on the sample built of municipalities with more than 20,000 inhabitants.[6] This means using a sample of 507 municipalities whose characteristics are described in Table 4.8. A final note: for eight municipalities was not possible to calculate indices of ethnic segregation. Thus, the sample is reduced to 499 units for the application of the model to ethnic segregation.

Following existing literature, we expect geographic area and population size to be relevant in contributing to segregation. Thus, we firstly test for these predictors. More specifically, following from analysis in the previous section, the effect of being a municipality in the South or in Islands is expected to be significant and positive as well as population dimension. Both variables are the only predictors in Model 1, while all other covariates are added in Model 2 (Table 4.3.3). Consistently with our hypothesis, all Model 1 show that geographic areas (reference category: North-West cities) and municipality's population are determinant in defining levels of segregation between social and ethnic groups. On average, the lower and middle classes are more segregated with respect to the upper class in municipalities on Islands. The

5. Variables that contribute to the computation of segregation measure are omitted in the model.
6. Further analysis have shown that results are consistent using both strategies.

Variable	Mean	Std. dev.	Minimum	Maximum
Municipality population	61,414	147,954	20,018	2,617,175
Geographic area				
North-West	.215	–	–	–
North-East	.160	–	–	–
Center	.205	–	–	–
South	.284	–	–	–
Islands	.136	–	–	–
% of housing built in 1991-2010	.144	.067	.002	.438
Functional specialization				
% of retirement	.200	.040	.073	.290
% of upper class	.158	.043	.075	.339
% of blue collars	.285	.060	.125	.534
% of university graduates	.107	.036	.038	.242
Nationalities groups				
% from EU member states	.014	.013	.000	.105
% from non EU member states	.014	.013	.000	.063
% from other continents	.026	.025	.001	.159

Table 4.8: Description of predictors.

pattern holds for ethnic segregation and establishes a strong hierarchy between municipalities and ethnic groups. Segregation of Europeans from EU-member states is about 20% higher in municipalities of the South and Islands, while it is slightly less for those coming from continents other than Europe. When looking to Europeans fron non-EU-member states this figure rises to about 60% in Islands. Moreover, they are also experiencing less segregation to Italians in North-East cities. All the considered groups are severely more segregated in large urban aggregates. A possible explanation is that greater shares of urban space are associated with more possibilities of clustering of groups, thus fostering segregation.

When looking at full models (Model 2) geographic area maintain its effect on segregation in line with Models 1, as well as city size. The net effect of these variables, after controllig for the additional covariates, is slightly lower but still in line with previous models. The recent expansion of municipalities is measured by the share of residential buildings built in the 1990s. Models show that local construction policies in a period of great soil consumption seem to have any relation with levels of segregation. In other words, new buildings are not contributing to shape the spatial disposition of groups in the urban landscape. The functional specialization of the city is measured by the presence of lower class and upper class. In the first case, high shares of lower class are interpreted as a city with a more industrial vocation, in the second high shares of upper class suggest a more specialization towards professional activities. These variables are not significant in models for both socio-economic and ethnic segregation. Finally, municipalities with higher shares of over 65 experience higher degrees of residential segregation in all models, except for the last one. Both the middle and the lower class are substantially more segregated from the upper class

(about +50%) in elder towns. This pattern holds also for those coming from non-EU-member states With reference the the latter group. The possible explanation looks again at the structure of the local labor and housing market. It might be the case that areas with high percentages of retired individuals are characterized by higher housing and living costs that non-EU immigrants can't afford. On the other hand, these might be the more rural areas characterized by few migrants and workers.

Variables	Middle class		Working class		EU		Non-EU		Other	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Region										
North-West (omitted)										
North-East	-.008	.001	-.015	-.002	.014	.024	-.054*	.027	.028	.036*
Center	-.008	.005	-.012	-.000	-.007	.011	.028	.021	.045**	.035*
South	-.003	.017	.019	.037*	.183***	.080**	.127***	.057*	.146***	.097***
Islands	.037***	.051***	.074***	.084***	.196***	.080**	.585***	.394***	.180***	.122***
Municipality characteristics										
Log of total population	.128****	.116***	.184***	.147***	.151***	.171***	.144***	.121***	.092***	.107***
% new buildings (1990s)		-.040		-.162		-.153		-.000		-.094
% lower class		.043		-.121		-.142		-.200		-.020
% upper class		.132		.190		-.571		-.815*		.-.510
% over 65 years-old		.508***		.467**		.129		.965***		.321
% Europeans		-.465		-.291		-5.92***		-1.14*		-1.73***
% non-Europeans		-.567		-.921*		-.975		-7.50***		-.961
% Others		.058		.073		-.635		-.435		-.895**
% university degree		-.183		-.052		-.085		.497		.216
N(municipalities)	507		507		499		499		499	
Log-likelihood	638.72	655.82	536.26	562.79	279.17	346.07	309.20	394.35	410.06	433.92

Note.-***p-value<.001, **p-value<.01, *p-value<.05.

TABLE 4.9: Farley and Frey (1994) model for structural determinants of segregation.

4.3.4 District and school segregation

The aim of this section is to investigate the urban space in which schools are inserted (school districts) by means of measures of segregation and population composition referring to the two sources of the database, to investigate the composition of schools' student population and to inquiry the correspondace between the two. Firstly I discuss districts composition, secondly school composition and then I investigate if district composition reflects in school population. In Section 3.4.4 I've presented the idea to overcome the problem of school districts boundaries definition by constructing egocentric circular areas of different radii and centered in schools. These artificial districts, or catchment areas, identify the portion of urban space that might be served by a single school and define the theoretical reference population. The main shortcoming of this solution is that it might encounter difficulties in covering the entire municipality surface. By construction, they are not mutually exclusive. Let's consider two schools, A and B located in the urban space. Student's residence is exactly in the middle of the two schools, such that home-school distance is equal in both cases. To which school does he/she belong? By relaxing the mutually exclusive condition we are letting this student to be considered in the theoretical population of both schools. It is worth noticing that the overlapping of districts is not only a matter of possibilities for the student, but also of competition among schools since school districts are not relevant to enrolment procedures in Italy.[7] Moreover, a similar feature characterize districts in the real world, where a reference population defined by administrative borders is served by more than one school. I've decided to build districts on four different radii: 200mt, 300mt; 400mt; 500mt, where increasing radius is associated to increasing relaxing of the mutually exclusive condition.

Districts are built only for public schools, that represents 60% of 1,098 sample schools, and can be investigated by using both ISTAT and INVALSI data, althogh results should be interpreted differently. Census data (population counts) allow to describe the urban area in which the school is located, while INVALSI data (individuals) to identify the theoretical reference population of each school. Districts are then analyzed in terms of their socio-economic and ethnic composition. For census data we can investigate the level of segregation in terms of eveness of distribution (D) for each district, while through INVALSI data we can look at the difference in composition of school theoretical and factual population. In order to gain a better understanding of school districts' composition, the ratio between upper and lower class is also computed. Finally, school composition is investigated through analysis of the ESCS score, an index of socio-economic and cultural status of pupil's family according to decomposition of variance presented in Section 3.4.1. For simplicity, I recall here that the score is standardized with mean [0] and variance [1]. High values are associated with high levels of socio-economic and cultural capital.

7. See discussion in Section 3.2

Group combination	D_{200}	D_{300}	D_{400}	D_{500}
Upper vs Middle	.199 (.136)	.226 (.113)	.234 (.094)	.240 (.082)
Upper vs Low	.275 (.167)	.315 (.135)	.332 (.119)	.343 (.111)
Upper vs Excluded	.202 (.144)	.233 (.127)	.246 (.113)	.255 (.109)
EU vs Italians	.362 (.253)	.435 (.221)	.464 (.198)	.479 (.187)
Non-EU vs Italians	.340 (.307)	.465 (.298)	.510 (.275)	.538 (.260)
Others vs Italians	.284 (.192)	.343 (.166)	.365 (.150)	.381 (.141)
	200mt	300mt	400mt	500mt
Upper/lower ratio	1.90 (2.41)	1.73 (1.77)	1.62 (1.50)	1.57 (1.41)
Immigrants/Italians ratio	.111 (.123)	.116 (.171)	.114 (.136)	.114 (.134)

TABLE 4.10: Profile of school districts (ISTAT data).

Table 4.10 presents average D values and related standard deviation for the upper class *vs* other groups and ethnic minorities *vs* Italians across districts. The figures show the classical ranking between socio-economic and ethnic groups. The upper class is more segregated with respect to the lower class than the middle class and excluded. Ethnic groups are substantially more segregated than socio-economic ones, here Europeans are systematically more segregated than those from other continents. These patterns hold for increasing dimension of districts. On average, smaller districts are less segregated than bigger ones, but results are in line with those highlighted in previous sections. No relevant disproportion in the ranking of segregation and average levels being symilar to those for the entire municipality suggest that schools are quite evenly spatially distributed in the space. Finally, the table presents two ratios computed for all dimensions. The first is the ratio between upper and lower class, and aims at defining a score of the disproportion between these two groups. Its score is higher than one, indicating an over-representation of the upper class on the lower one, but decreasing over radius dimension. The second indicates the disproportion between immigrants (all categories) and Italians. The ratio assumes a very low level, as we might expect due to the low figures of immigrants in Italian cities, and is constant across radius dimension. In more general terms, results from this table suggest that on average schools are located in moderate segregated districts on both dimensions, district segregation is symilar to average municipality levels but characterized by substantial variation.

A second step in the understanding of districts and school composition is to apply the previous analysis to individual data. Here, the reference population is represented by all grade 8 students residing in the sample municipalities. Again, districts are computed with the same method: four concentric areas defined with different radii. Individuals can be considered as reference population of different schools. Since I'm referring to individual data, the Dissimilarity index of each district can not be computed. Alternative strategies are pursued, aiming at describing the composition of districts. Table 4.11 presents the mean ESCS score for district population,

Group combination	$200mt$	$300mt$	$400mt$	$500mt$
Avg. population	11.06 (9.40)	23.54 (16.65)	39.52 (26.79)	58.43 (38.62)
Min population	1	1	1	1
Max population	73	114	168	241
ESCS	.146 (.801)	.156 (.708)	.144 (.663)	.153 (.640)
Upper/lower ratio	1.23 (1.84)	1.86 (3.01)	1.92 (2.84)	1.97 (2.99)
Imm./Ita ratio	.179 (.349)	.168 (.274)	.166 (.227)	.155 (.187)

TABLE 4.11: Profile of school districts (INVALSI data).

the ratio indicating disproportion between upper and lower class, and that for immigrants *vs* Italians. Districts average population is strongly skewed, as the upper part of the table shows. This fact might be the result of the combination of the spatial disposition of schools and pupils. Some schools might be located in non residential areas, or simply in areas characterized by the absence of grade 8 pupils. On average, the composition of districts is stable across radii and consistent with results from census data. The average ESCS score hovers around the value of .15, but its standard deviation is significantly high and indicates high territorial variation. A similar case for the two selected ratios investigating disproportion between groups. On average, the upper class is more represented in our districts than the lower class. These figures are in line with those from census data, but are also the result of the segregation ranking highlighted in previous sections. This result is not surprising, since the upper class is more present in core municipalities (as those selected in this study) and the lower class in the related hinterland. Symilarly, the great average disproportion of immigrants with respect to Italians is not surprising.

Additional information on districts is provided in figures 4.3 and 4.4. The former presents the mean district ESCS score (y axis) according to variation in district upper/lower class ratio (x axis). The positive relation seems to be somehow quadratic, and districts with high mean ESCS are also characterized by a strong disproportion in the presence of the upper class. More interesting is the relationship between mean district ESCS and immigrants/Italians ratio. Here the association is slightly negative, pushed downwards by outliers. In general terms, it seems that the strong disproportion of immigrants does not have a strong relationship with the mean ESCS, and pushes it towards average levels close to the mean of the distribution. This negative not significant relation is certified by the correlation between all the immigrant/Italians ratios and the average ESCS for the related district. The correlation is slightly negative, but not a significant association.

When coming to school composition, alternative measures are available. Figure 4.5 presents the distribution of public and private schools according to average ESCS. The figure highlights clear differences among the two groups. Public schools have a distribution that is more close to a Gaussian, although not perfectly symmetric. The mean and median hover around zero, with a slightly pronounced tail on the left

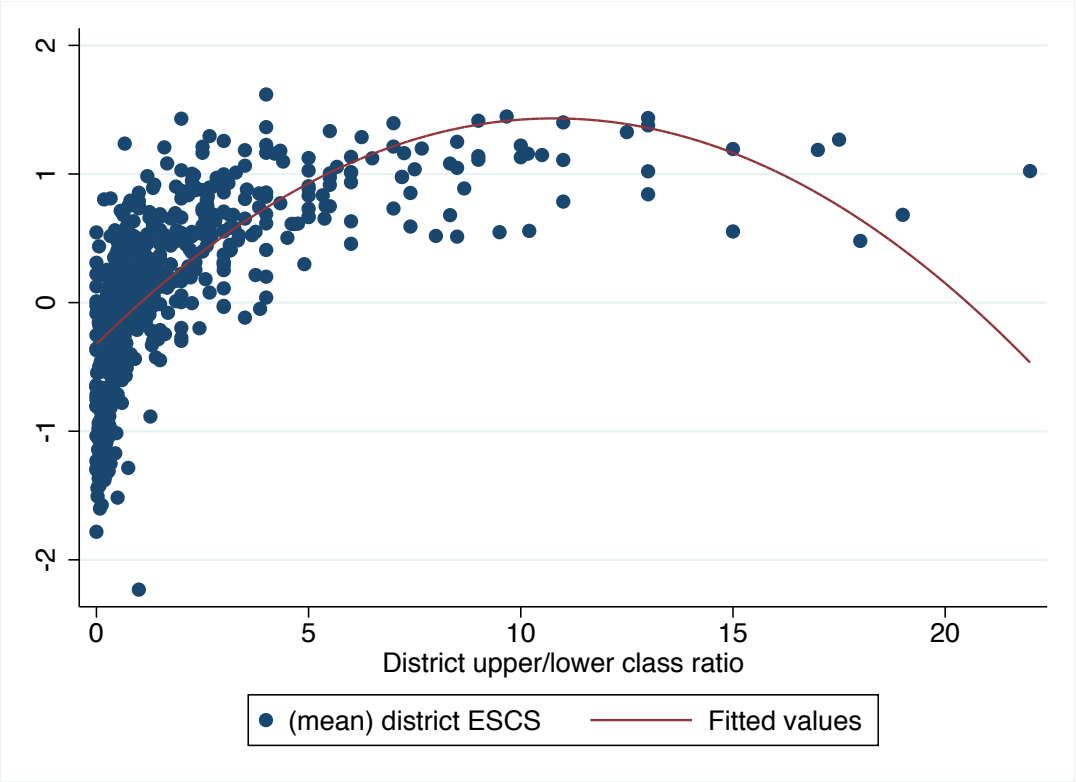

FIGURE 4.3: District ESCS by upper/lower class ratio.

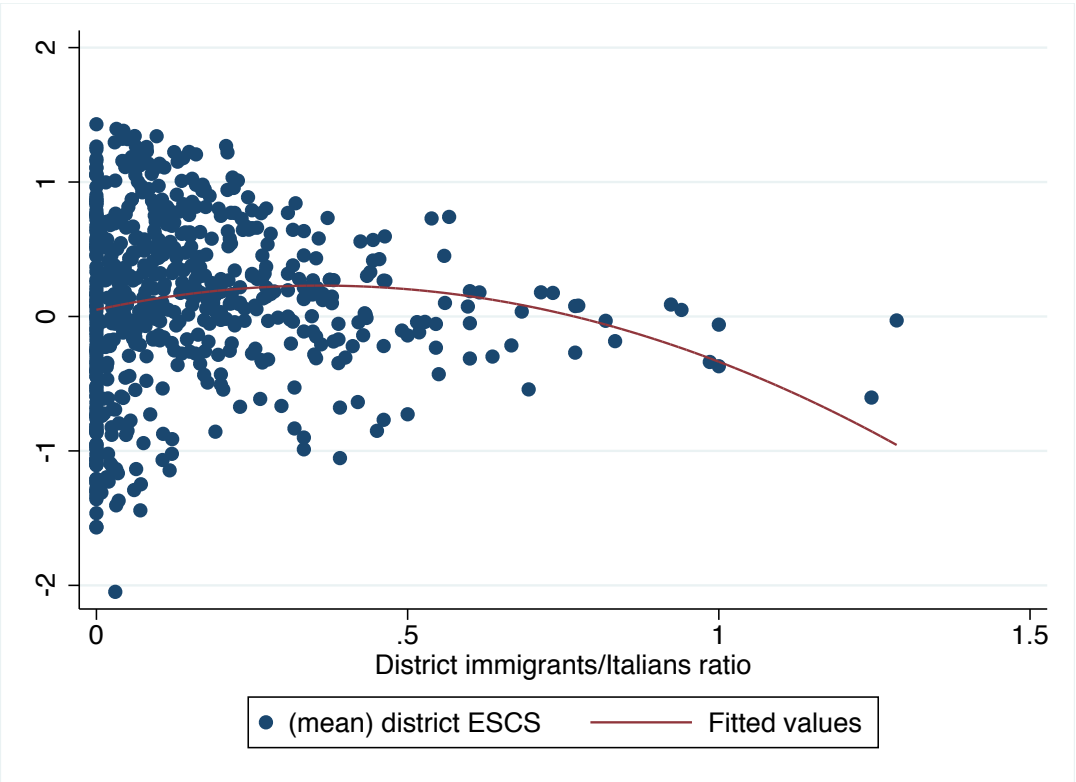

FIGURE 4.4: District ESCS by immigrants/Italians ratio.

side of the distribution. As expected, private schools are characterized by higher values of ESCS. This second aggregate is more homogeneous in terms of average socio-economic and cultural status than the previous one, with less than 70% of private schools with an average ESCS over the mean value of zero.

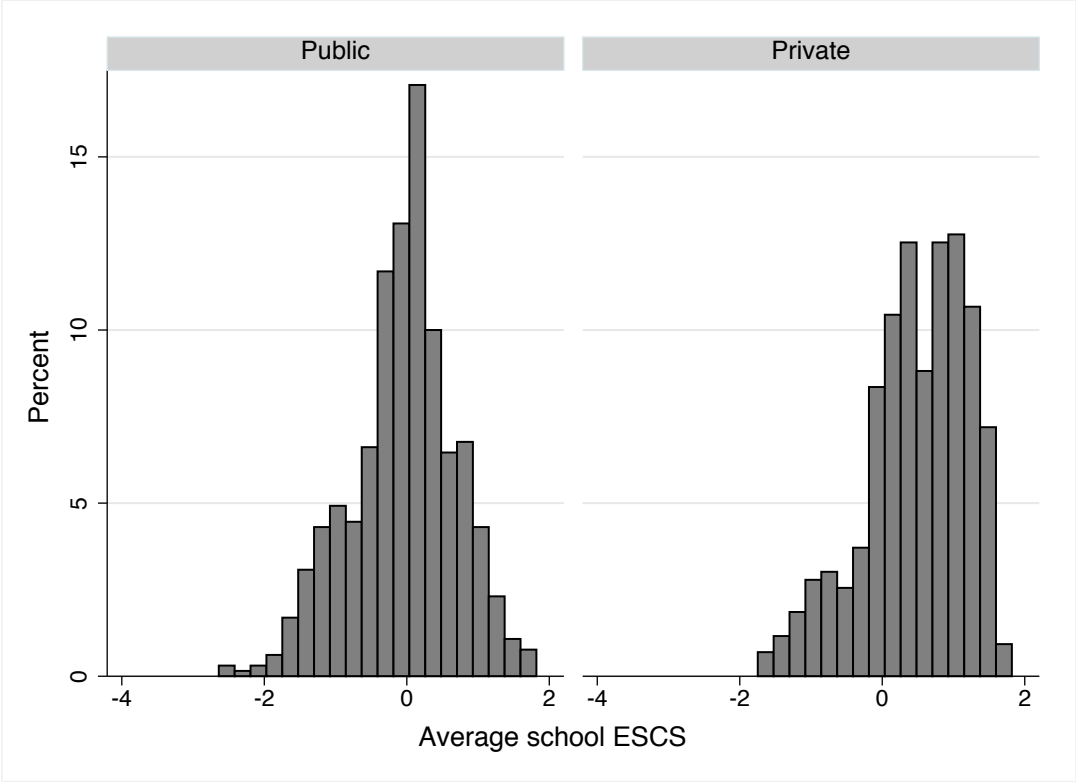

FIGURE 4.5: Average ESCS of schools. Public *vs* private.

Figure 4.6 shows the distribution of public and private schools according to percentage of non-Italians in their student population. Each bar has a width of 2% and the two graphs are presented on the same y axis scale. Both distributions are very skewed and predominantly defined in the range 0-40%. Slightly more than one-forth of schools in both distributions have less than 2% of non-Italian students. Another 55% of public school have a share of foreign pupils ranging between 2-20%, and about 15% in the range 20-40%. The same figure for private schools is about 70%, while only a residual part of schools has more than 20% of non-Italian students. On ethnic bases, private schools are more homogeneous then public ones. Here, three groups can be defined in the distribution. Schools with extremely low shares of immigrants (less than 5%), school with less than 20%, all other schools. Although homogeneous schools in terms of ethnic minorities are hard to be found in the selected sample, substantial differences in the distribution of pupils from ethnic minorities are present.

Average values for variables of interest at the school level hide variability inside schools. In fact, it is not granted that classes are perfectly reproducing the average composition of schools. In order to investigate internal composition I refer to decomposition of variance, a technique presented in Section 3.4.1. School classes are the group of interest, thus it is possible to compute variability due to differences

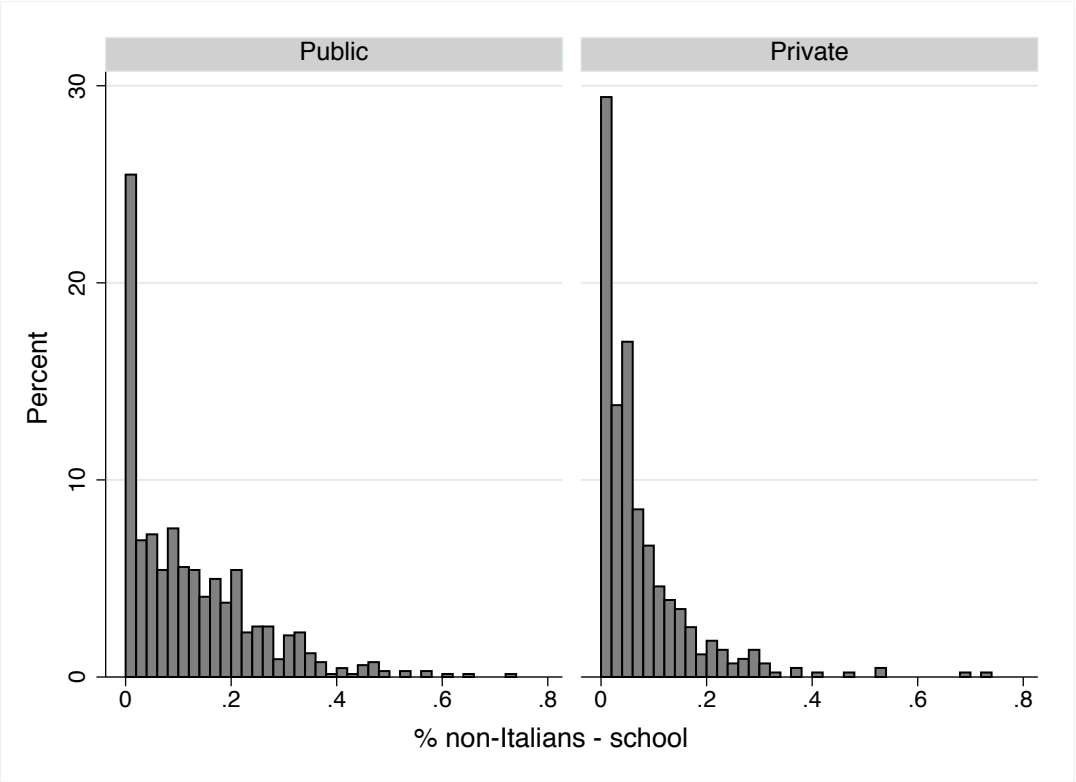

FIGURE 4.6: Schools by percentage of non-Italian pupils.

between classes of the same school and variability within each class. These two measures allow to identify the degree of homogeneity of each class, but also the degree of homogeneity between classes of the same school. Each school can then be plotted according the *between* and *within* component, thus defining the distribution of schools. The red lines are drawn in correspondence of average variance values and divide the plane in four quadrants. The first quadrant (upper right) is characterized by schools with within and between classes variation above average. It means that classes of these schools are heterogenous in their composition, but there are also differences between classes. Thus, we can think of these schools as those that are experiencing mild levels of segregation. The second quadrant (lower right) is made by schools with high variation within classes, but small differences are present in composition between classes. This group is associated to a more "democratic" partition of students. In a school belonging to this quadrant all classes are quiet heterogeneous but there is small variation between them. The third quadrant (lower left) is build by schools with homogeneous classes. These schools are experiencing a particular case of segregation. They are not segregating between their classes, but are themselves a segregated environment with respect to other school. Finally, the fourth quadrant (upper left) is made by strongly segregated schools. This group is characterized by classes that are homogeneous in their composition, but there is variation between classes of the same school. This is the most evident case in which segregation configures itself as the phenomenon of social distance expressed in physical separation of space.

Figure 4.7 presents schools according to their within and between classes ESCS variation. Private schools are taken out of the picture, although they are all distributed in the third quadrant (homogeneous schools). Private schools are again considered in subsequent figures. Quadrants are defined by the red lines drawn in correspondence of average values of both within and between variation. The distribution spreads on the right in the first quadrant, that is also the most populated one. Schools in this portion of the Cartesian space show a high degree of differentiation. In general terms for this group, the within classes variation is smaller than the between component. This element strenghten the idea that these schools are experiencing low levels of segregation. From the figures clearly emerge that strongly segregated schools, included in the fourth quadrant, represent the less numerous aggregate. This group can be interpreted as schools that are actively distributing pupils among classes according to their socio-economic status. However, the group is relatively small suggesting the marginality of this phenomenon.

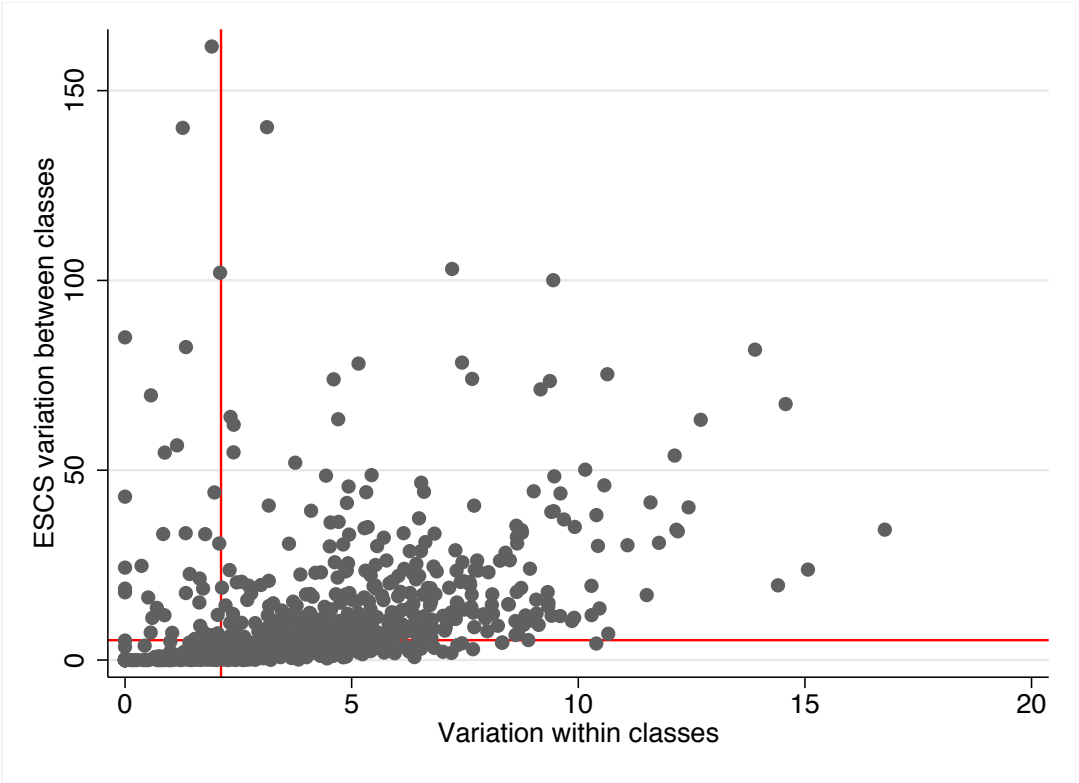

FIGURE 4.7: Schools by within and between variation of ESCS.

Figure 4.8 presents the same data, but schools are weighted by their share of non-Italian population, thus dots dimension highlights different portions of immigrants students. A strong pattern of differentiation of schools according to percentage of non-Italian students does not seem to emerge.

Finally, figures 4.9 and 4.10 present the distribution fo schools in a different way. Schools are divided by quadrants according to variation in ESCS, with the addition of private institutions, and plotted according to their average ESCS level and percentage of non-Italians. In Figure 4.10 schools are wheighted by the between classes

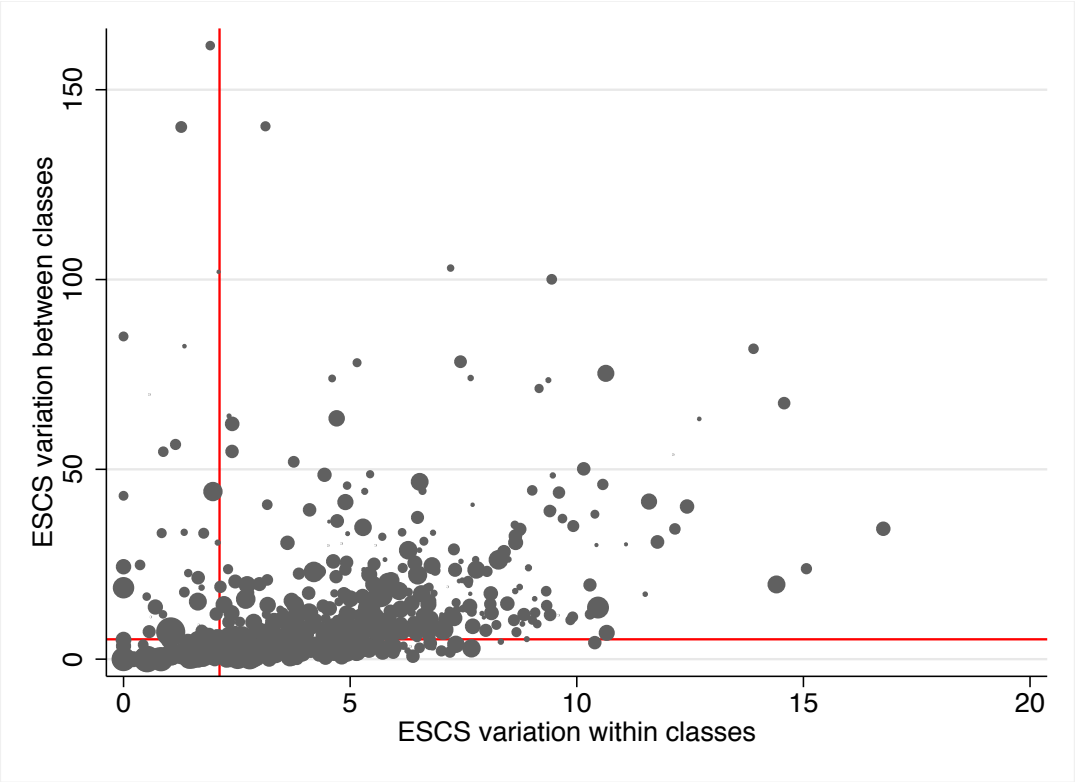

FIGURE 4.8: Schools by within and between variation of ESCS, weighted by percentage of non-Italians.

variation of percentage of non-Italians. Apart for private schools, all groups are quite evenly distributed with respect to the mean of ESCS distribution, although some of them have few cases and distributions are not consistent. The distribution of democratic schools is likely to be splitted in two parts, a lower tale characterized by low ESCS and less than 20% of non-Italian pupils, and a central body with both higher mean ESCS value and share of immigrant pupils. However, no substantial differences emerge between democratic schools and those with weak segregation. The group of homogeneous schools presents several outliers with respect to the central body of the distribution. The distribution of this group must be read with that of private schools. Private schools are highly homogeneous in their composition and strongly characterized high levels of average ESCS. Indeed, they are a particular case of homogeneous schools: the ones serving predominantly upper class pupils. On the other side, homogeneous schools might be those serving more the middle class and characterized by different degrees of non-Italians. Finally, strongly segregated schools expresses low levels of average ESCS and high levels of foreign pupils. These are schools characterized by strong internal variation in terms of average ESCS. From Figure 4.9 seems that these schools are segregating between the middle and lower class, since their average ESCS is hardly higher than the standardized average (0).

Figure 4.10 presents the same results but with the adjunction of between classes variation in the share of non-Italian pupils. Larger circles are associated with higher

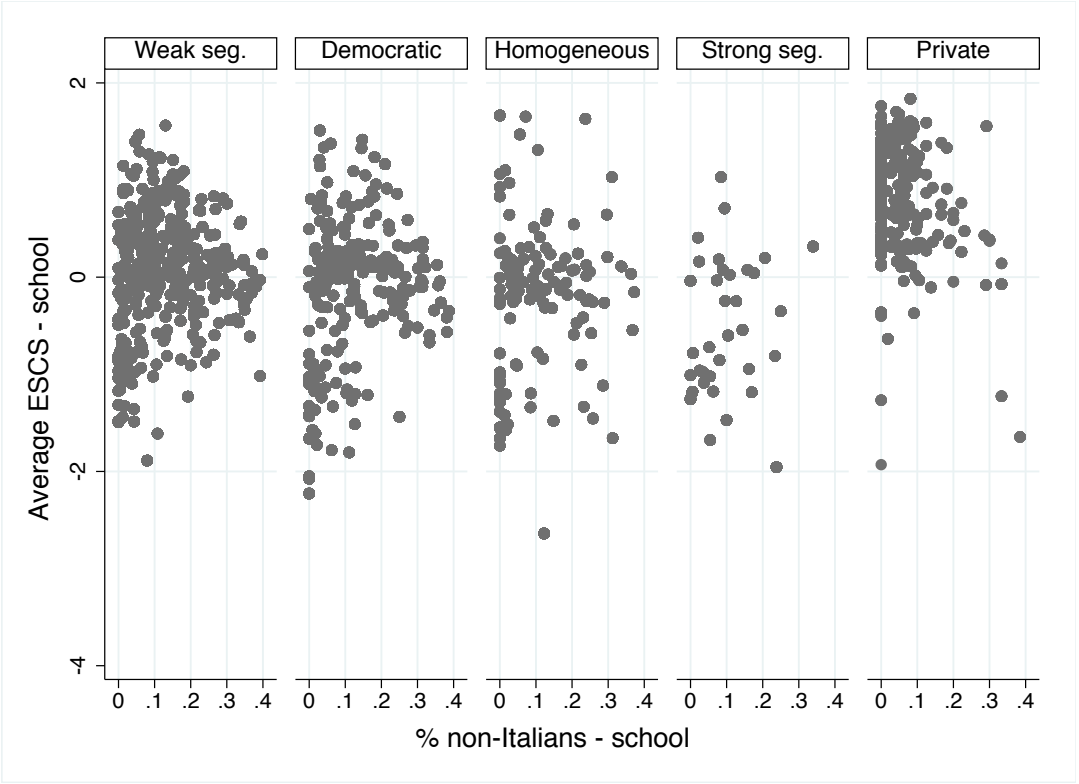

FIGURE 4.9: Schools by segregation profile, average ESCS and percentage of non-Italians

variation between classes, a graphical sign that can be interpreted as higher differentiation on ethnic bases between classes. As the share of non-Italians increases, the circles expand. Likely, this highlights the following pattern: when immigrants are a small portion of the total student body, they are distributed among classes. However, when the share of non-Italians increases above about 10%, the between classes variation increases suggesting that above a certain threshold foreign pupils are concentrated in classes. The phenomenon is consisteng in all categoris, although private schools present also opposite results. This might be the case for special schools serving foreigner students.

Average levels of segregation measured by both D and H indices are reported in Table 4.12. Indices are computed at both the school and municipal level. In the first case, the Dissimilarity index expresses the share of group m that must be relocated between classes of the same school in order to obtain an even distribution, while the Information theory index is measuring the diversity of classes from the average composition of the school. It is then a measure of the extent to which each school divides its students. In the second case, the Dissimilarity index expresses the share of group m that must be relocated in the municipality in order to obtain an even distribution between classes of schools from the same city, while the Information theory index express the average diversity of classes at the local level.

Apart for the excluded group, the rank highlighted in previous sections appears also in schools and municipalities: the upper class is the more segregated group,

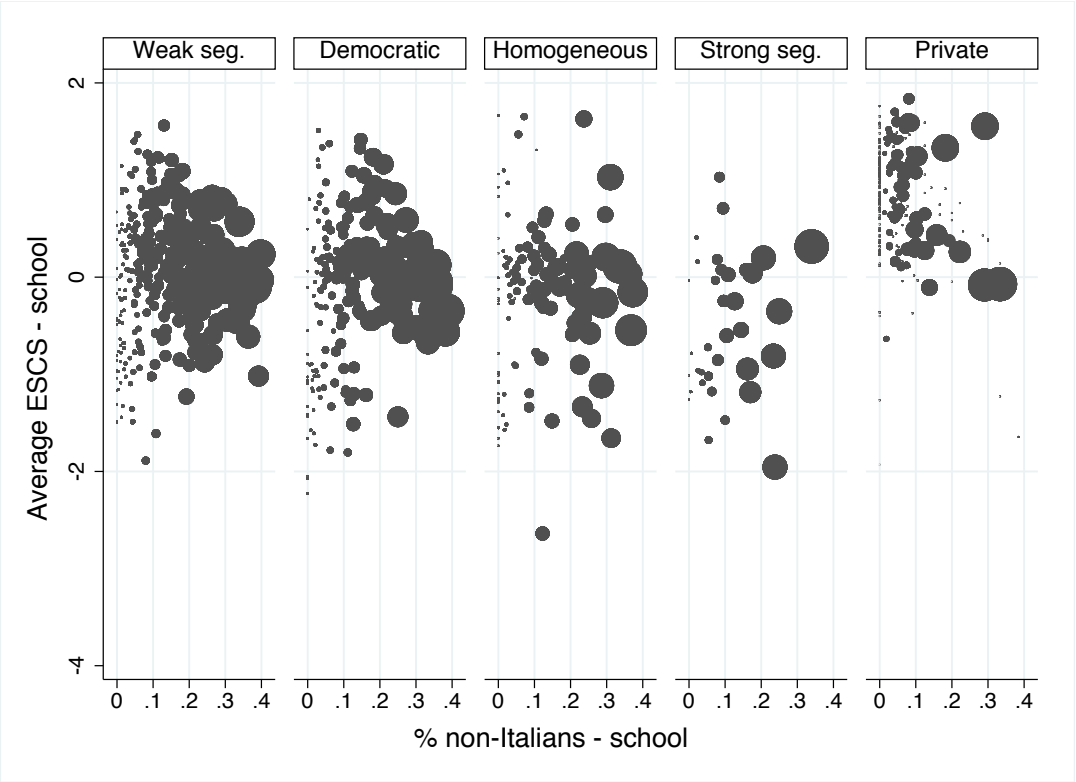

FIGURE 4.10: Schools by segregation profile, average ESCS and percentage of non-Italians, weighted by between class variation of non-Italians.

followed by the lower class and the middle one. The average proportion of pupils that have to be relocated between classes is usually less or around 20% when considering school segregation, while it is about the half for municipal segregation. When coming to segregation by ethnicity, individuals from continents other than Europe are the most segregated group at school level, but not at the municipal level. Here, the first place is hold by Europeans from EU-member states, that with Europeans from non-EU member states are experiencing more segregation between classes of the same municipality rather than between classes of the same school. Concentration of this group happens in specific schools, and is probably the case of special private schools oriented to the offspring of Europeans. The French school is an example of this case. In fact, levels of municipality segregation for Europeans are higher in private schools than public ones.

Table 4.13 presents the two matrices of average P values for both socio-economic (upper part) and ethnic groups (lower part). Results are not comparable to those presented in Table 4.5 for municipalities, as the units we are considering are substantially different (*sezioni* in one case, classes nested in schools in the other one). Medium levels of exposure are found for the upper class to other groups, with the exception of excluded. Indeed, this is a very small category that, as discussed in Chapter 3. Interestingly, the middle class is experiencing low levels of exposure, as probabilities to share a class with other groups are less than 40%. Finally, the lower class has

	D_{sch}	H_{sch}	D_{mun}	H_{mun}
Complete sample				
Upper class	.231	.071	.147	.030
Middle class	.164	.040	.087	.013
Lower class	.195	.054	.106	.016
Excluded	.297	.098	.158	.029
Italians	.274	.094	.160	.033
EU	.291	.107	.316	.078
non-EU	.217	.078	.287	.070
Other	.291	.103	.235	. 056
Public schools				
Upper class	.290	.092	.092	.009
Middle class	.193	.047	.046	.003
Lower class	.235	.066	.061	.005
Excluded	.376	.127	.077	.008
Italians	.342	.119	.098	.011
EU	.398	.146	.174	.031
non-EU	.307	.110	.182	.029
Other	.375	.136	.118	. 017
Private schools				
Upper class	.229	.069	.092	.010
Middle class	.176	.043	.046	.003
Lower class	.231	.068	.070	.007
Excluded	.326	.110	.093	.011
Italians	.347	.127	.107	.013
EU	.291	.139	.199	.036
non-EU	.229	.085	.219	.038
Other	.358	.133	.136	. 020

TABLE 4.12: Average segregation levels of sample schools and municipalities.

	(1)	(2)	(3)	(4)
(1) Upper class		.594	.468	.271
(2) Middle class	.356		.358	.176
(3) Lower class	.423	.564		.218
(4) Excluded	.539	.665	.597	
(1) Italians		.022	.018	.071
(2) EU	.439		.140	.285
(3) Non-EU	.349	.128		.228
(4) Other	.579	.130	.102	

TABLE 4.13: Average exposure index P, sample schools.

medium levels of exposure to other groups (around 50%). Possibilities of Italians to get in contact with students from different ethnic backgrounds are really low, about 2% for both kinds of Europeans and 7% for those coming from other continents. This results is influenced by the strong disproportion in absolute figures of majority and minority groups. However, minority ethnic groups are not exposed one another. The average exposure level for these between group combinations are usually lower than 15%.

Are schools more or less segregated than related districts? Figure 4.11 compares average ESCS of both schools and the related districts build with radius of 400mt. Blue dots are schools whose average ESCS is higher than that of the related district, red dots are schools whose average ESCS is lower than the ESCS of their district. The first group is made by 207 public schools, while the second 456. Moreover, the correlation between school and district ESCS is high, about .828. Figure 4.11 suggests that district composition does not directly reflects into school composition. This means that schools are drawing their population from a bigger portion of the urban space than that defined by their own districts. A result calling for further investigation on movement on pupils and school choice. This result is in line with empirical evidence from Pacchi and Ranci (2017) on Milan, where between districts movement has been detected. Figure 4.12 presents a similar analysis, rooted in differences between the share of non-Italian pupils in schools with respect to districts. Also here, blue dots are schools with more foreign pupils than those present in their district build with 400mt radius, while red dots are schools with less foreigners that the related district. Here there are more schools in the upper part of the graph (351) than those in the lower part (216), while 96 schools have the same share of foreigners than the related district. This figure suggests similar conclusions from those drawn previously. With respect to the related district, some schools are likely to expell immigrant students, while other are receiving.[8]

8. These figures call for further investigation on the movement of pupils between schools, this inquiry is included in the next chapter.

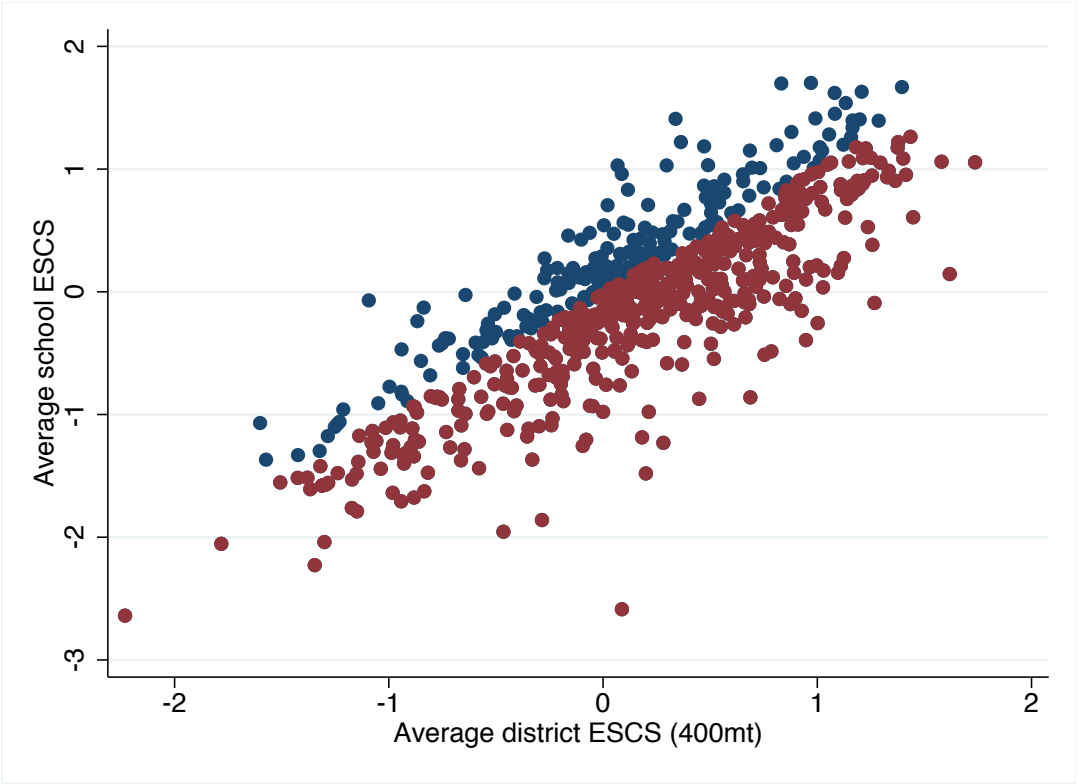

FIGURE 4.11: Comparison of school and district composition - ESCS.

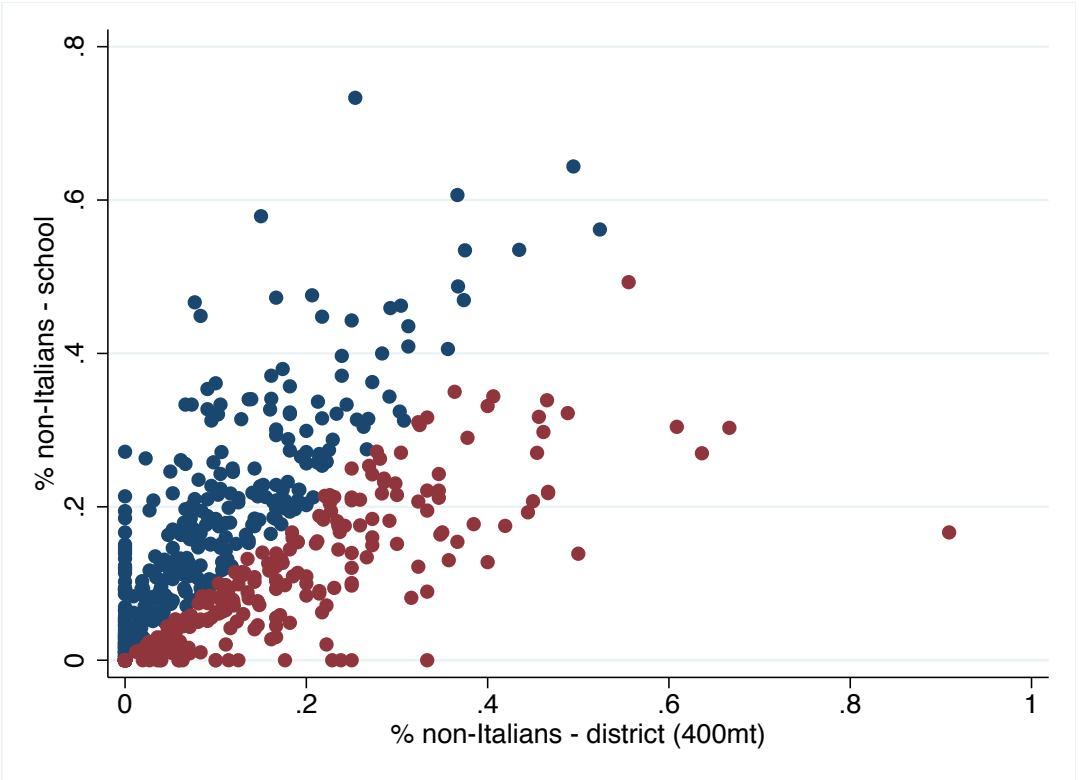

FIGURE 4.12: Comparison of school and district composition - shares of non-Italians.

4.4 Conclusions

This chapter has been devoted to the investigation of levels of segregation in Italy, its structural determinants, and the level of segregation of schools and districts. Existing literature has shown little interest in the Italian case, and this contribution aims to add some evidence on the selected topic. This section presents the main results for each research question, discussing limitations and possible strategies for further work.

Which is the level of residential segregation by socio-economic and ethnic characteristics in Italy?

Socio-economic and ethnic segregation of municipalities have been investigated by means of three segregation indices: the Dissimilarity index (D), the Information Theory index (H), and the Exposure index (P). The combination of these indices has been used in order to depict a profile of Italian municipalities. Italian cities are characterized by medium levels of segregation, although variation has been found according to city dimension. When looking at eveness of distribution, the upper class is the most segregated socio-economic group (scoring about .30 in D), followed by the lower class, the middle class and the excluded. Exposure between socio-economic groups is also characterized by medium levels, with two noticeable facts. The upper class is sharing space with other socio-economic groups, but the opposite is not true. A possible interpretation is that the upper class is not present in the entirety of the urban space (uneveness), but shares areas with other groups as a minority.

Italian cities are characterized by medium-high levels of ethnic segregation. Europeans from non-EU-member states are the most segregated group, followed by Europeans of EU-member states and those coming from other countries. The levels of ethnic segregation are higher than those for socio-economic one across all urban sizes. When looking at exposure, all minorities are poorly exposed one another, while are highly exposed to the majority group (Italians). When both forms of segregation are investigated by geographic variation, cities in the South and Islands usually show higher values. However, the difference is more pronounced for ethnic than socio-economic segregation.

Here, the main limitations of the analysis presented reside in methodological choices. Further work should address the solution of the modifiable areal unit problem. Possible roads involve using spatial indices of segregation, that are able to overcome the classical limitations of aspatial versions by implementing spatial proximity and distribution of individuals; to aggregate *sezioni* in order to produce a new unit that should solve the problems related to low population counts and missing values; investigate segregation through the spatial composition of space as suggested by Préteceille (2003).

Which are the structural determinants of segregation in the urban space?

Determinants of segregation have been investigated using the structural ecological model proposed by Farley and Frey (1994). The idea behind this model is that segregation is the result of ecological elements of the urban space. The functional specialization of urban areas, as well as the presence of new buildings. The original model includes geographic region, functional specialization, age of metropolitan area, percentage of buildings recently built, population size, and percent of minority population and minority income as a percent of majority group income. The model has been applied in its original shape, with the exclusion of the minority income share predictor, following Lee et al. (2008) between groups decomposition. Five models have been computed for many dependent variables each of which is the municipality segregation level (expressed by Dissimilarity index in its classical two-group version) for middle class, lower class, individuals from EU-member states, individuals from non-EU-member states, migrants from other continents.

As expected, results show that municipality's total population has a positive and significant effect on by-group levels of segregation. Also geographic area is positively correlated with segregation, with municipalities in the North-East, South, and Islands characterized by higher levels. The share of over 65 years-old is a significant predictor of segregation between upper and middle class, as well as between upper and working class. This result might suggest that portion of urban areas populated by retired individuals are pretty homogeneous in terms of socio-economic composition. In other terms, middle and working class retired live separately from the upper class. Two interesting results involve segregation between Italians and individuals from non-EU member states. Percentage of upper class is negatively associated with segregation of these two groups, while presence of over 65 years-old it is positively. These interesting results might be the consequence of the occupational structure of this immigrant group and its consequential relation with spatial disposition. An in-depth analysis might be useful in disentangling alternative explanations.

Which is the level of segregation of classes, schools, and school districts by socio-economic and ethnic characteristics in Italy?

Are private schools more or less segregated than public schools? Results on school composition and segregation presented in this chapter suggest that dynamics of segregation of pupils by socio-economic and ethnic background are present in Italian lower secondary. On average, lower secondary school is characterized by medium levels of segregation, although selected highly segregated cases exist. The first distinction is between public and private education, with the latter being highly homogeneous in terms of socio-economic composition. The private school appears to be the segregative strategy of part of the upper class, that exit from public schools in order to find a more homogeneous environment.

Public schools have been analyzed by decomposition of ESCS variance within and between classes, producing a four category scheme. The first component defines variation of social status inside each class, the second describes variation of classes of a same school from the school mean. Thus, we have schools with weak segregation,

with heterogenous classes and with small differences between classes; schools with high variation within classes, but small differences between classes, this group is associated to a more "democratic" partition of students; schools that are homogeneous on both dimension; strongly segregated schools, characterized by class homogeneity but substantial variation between clases. The mild levels of segregation in lower secondary are driven by the great number of weakly segregated and democratic schools.

Further work should be addressed at refining the categorization of schools and analyzing in great detail the spatial disposition of these schools, and their connections with the urban space. This is also leading to case studies analysis.

Does public schools population reflect district composition?

Our investigation of district segregation has highlighted methodological difficulties for a comprehensive investigation at the national level due to the absence of a public repository of district boundaries[9]. In order to overcome this problem, an egocentric local environment has been created using schools as centre of circles with different radius. Subsequently, segregation indices have been computed for these districts using census (aggregate) and INVALSI (individual) data. If a process of school segregation is in place, we should expect that schools' population is more homogeneous than the related district on both the socio-economic and ethnic dimensions. However, segregation for schools and districts is not comparable as they are computed using different units. If we look at composition of schools and districts it is clear that school population does not reflect district population. This means that schools are drawing their population from a bigger portion of the urban space than that defined by their own districts. This result is in line with empirical evidence from Pacchi and Ranci (2017) on Milan, where between districts movement has been detected.

Although informative, this analysis should face the threaths posed by the absence of proper definition of school districts. On one side, single school catchement areas could be computed by recurring to Voronoi polygons thus defining a single and unique portion of the urban space served by that school, on the other, we should focus on specific municipalities for which there is information on districts' boundaries.

9. Districts' boundaries are defined at the municipal level.

Chapter 5

School choice

Chapter 4 has shown that schools' composition is different from that of the respective district, suggesting the existence of a process of redistribution of pupils among schools. In Italy, moving between schools is a practice allowed by the actual institutional context (as discussed in Section 3.2). The adoption of this kind of policies is relatively recent, and has been a historical goal of right wing parties. The theoretical foundations of school choice policies have to be looked for in the United States, and have been discussed in Section 2.1. Here I just recall the main principle of this movement: school choice is a tool to reduce economic segregation among students in public schools by creating access for lower-income families to schools outside their neighbourhood. Although, for the Italian case school choice support was argued upon liberal values of freedom of choice and market competition among schools.

The American literature about school segregation has always highlighted the role of school movers as the main cause of concentration of individuals by group characteristics. For example, the so called "white flight", predicting that whites flee when minorities enroach into their school, has been historically addressed as the main cause of school segregation in the US (Coleman, Kelly, and Moore 1975). In the last decade the European literature has started to investigate school choice and its determinants, leading to a number of selected case studies concentrated in Northern Europe cities. As noticed by Oberti (2007b), few studies investigate school choice determinants in Southern Europe. Rangvid (2007) has analysed the flows of lower secondary school pupils in Copenhagen, finding low levels of residential segregation but high levels of school segregation. In this case, determinants of segregation are identified in immigrant background, parental education and family income. Oberti (2007b) has investigated Paris, finding a relevant role of the institutional arrangement (district definition and other rules) and of the territorial distribution of school provisions. Karsten et al. (2003) has investigated school choice in Dutch primary school finding that ethnic school composition is a relevant point in the school choice of parents, with Dutch parents more interested in social and cultural background matching than ethnic minority parents. Finally, Krinsten (2008) has investigated Essen primary school choice among families through an *ad hoc* survey. As in previous research, the author found substantial differences in school choice between native and ethnic minority parents.

The cause is rooted in the process of school selection by families. Krinsten (2008) conceptualized school choice as a sequential process made of three stages: (i) perception of different alternatives; (ii) evaluation of perceived alternatives; (iii) access to school. As individuals may differ in their perception and evalutation of alternatives (Becker 1957), they are limited in the process of school selection. In general terms, those with more resources will be more likely to make a more precise evaluation, and perhaps a choice that better fit their needs.

Coming to Italy, to the author's knowledge two investigations on determinants of school choice have been published. The first is a work on the determinants of school choice between the public and private sector (Checchi and Jappelli 2003). The authors concentrate on perceived school quality as a determinant of school choice, and especially as a discriminant for choosing private education, a hypothesis hardly supported from empirical evidence. Data from Bank of Italy have been used in this study, these data contain information on individual's perceived quality of public services, included education. Researchers test the effect of several predictors on the probability to enroll in private upper secondary schools: city size, family education and socio-economic background, number of siblings. They found that perceived quality of public schools is a relevant determinant, increasing levels of public school quality reduce probabilities to enroll in private education. The authors include also provincial dummies in order to account for omitted variables that might be relevant in school choice. The models' fit improve, but not the substantial results.

The second contribution is a recent work by Pacchi and Ranci (2017), devoted to inquiry school segregation in primary and lower secondary in the Milanese municipality. The authors concentrate their efforts in clarifying the complex structure of school segregation, investigating movement of pupils between districts defined through administrative data obtained at the local level. Results highlight the movement of pupils of upper and middle class from periurban areas towards schools in the centre, with upper class pupils characterized by higher probabilities to enroll in private schools. They highlight also the relevant role of school-home proximity in the process of school choice, as well as the perception of local alternatives. Differently prom Checchi and Jappelli, authors can not recurr to any self reported measure of school quality. This issue is also shared by this dissertation and clearly represents one on the main limitations in the analysis of school choice.

5.1 Research questions and hypothesis

This chapter is aimed at investigating the following research questions:

- Is there a substantial trend of pupils enrolling in public schools other than the closest to home? What about enrolment in private schools?

  Given previous results, we test the null hypothesis of pupils enrolling in their natural school, expecting to strongly reject it. This sort of "movement" of pupils is expected both in the public sector and between public and private sector.

- Are those from higher socio-economic status more likely to enroll in a school different from the closest to home? Are those from ethnic minorities more likely to enroll in a different school than Italians? Are school quality and school composition relevant in enrolment choices? What is the probability of being enrolled in a school different from the closest to home?

  For this set of research questions we test the validity of the following model.

  The prevailing idea in the literature is that, when given a set of school alternatives, individuals choose the one that seems more promising in light of their interest (Krinsten 2008). The Relative Risk Aversion model (Breen and Goldthorpe 1997) posits that this interest is avoiding downward social mobility. Despite the model having been developed for school transitions, it can be easily applied to school choice. Here, families choose the school minimizing the risk of social demotion. Clearly, it is difficult to imagine that families are able to determine how much a school will minimize the risk of downward mobility with respect to another one, and consequently choose the better costs-benefits combination. This might be the case when tracking is present, but not in non-tracked school levels like primary or lower secondary school. Alternatively, families might look for a proxy, or a combination of proxies, to estimate the best choice. On one side, perceived school quality. Parents may "vote with their feet", enrolling their offspring in schools with higher perceived standards. Since school quality is a multi-dimensional concept that builds on the quality of the several inputs of the educational process (economic resources, personnel, facilities, etc.) and their optimal combination, this evaluation is partial and costly. Despite the difficulties in its estimation, it might be the most relevant element in enrolment decisions. On the other side, school socio-economic and ethnic composition. Socialization with peers for both pupils and parents is a key element in school life. Families might try to take advantage of the positive externalities of socializing with individuals with a status higher than their own (or from a majority group), as well as their offspring. Socialization means being inserted in social networks, thus having access to information here available and making further evaluations less costly, and being exposed to values and beliefs of the peers.

  In a RRA model, the evaluation of the best (transition) choice is informed by the associated costs and the ability of the pupil. Access to full information plays a pivotal role in this model, as availability of information is associated to a more precise evaluation. In the particular application presented above, the evaluation of costs is the same from the original model. Families evaluate actual costs, like tuition fees (especially for private education), transportation costs, possible extra costs (i.e. private tutorship). The offspring's ability is evaluated in the light of past performances, but also in relation to the standards of the possible receiving school. If we consider school choice as an investment, whose dividend is

represented by minimizing the risk of downward mobility, the associated risk is represented by the risk of failure.

According to this formulation, families will evaluate a limited set of alternative schools where the number of alternatives is defined by family constraints (costs, house-school time travel, quality standards, etc.). The assumption of the model is that families will look for schools characterized by high educational standards (quality) and a population composition allowing for socialization with individuals from a higher status. Since higher risks of downward mobility are associated to the upper class, this group is expected to be the more aggressive in its strategies. Different groups might have different strategies to cope with minimization of risks, or being victims of strategies from other groups. This *caveat* reminds of the limitations of RRA models in interpreting segregation.

5.2 Data, variables and methods

The database presented in Chapter 3 is used for the investigation of the school choice phenomenon. Here, information about the population census is not needed, as we focus mainly on the part of the database from INVALSI data. The database includes 113.700 pupils enrolled in the last year of lower secondary (grade 8) in 2014/15 schooling year, the sample represents about the 20% of the total population of that year' students. The database is enriched with the precise geographic location of schools and pupils' residence.

Variables of interest describe the socio-economic condition of the household of origin, the belonging to a specific ethnic or nationality group, school quality, and the composition of its student body. The ethnic group variable includes four groups: Italians, individuals from EU-member states, individuals from non EU-member states, individuals from other continents. An alternative measure is represented by citizenship status, a variable operationalized in three categories: Italians, first generation foreigner, second generation foreigner. School quality is measured through the only available indicator: the mean INVALSI score of each school. The score is derived from 2011/12 wave of the test, that is the wave prior to enrollment of the cohort of interest. The database presents some information on school quality as measured by the characteristics of the teaching staff: the percentage of staff with fixed term contracts, the percentage of staff aged 55 or more. Unfortunately, this information is available only for public schools, a fact that makes it hardly usefull in our analysis.

I define alternative school choices as those made by pupils enrolled in a school different from the public one minimizing home-school travel distance. The nearest public school to home is defined as the natural choice, and represents the benchmark in the process of evaluation of alternative schools. Two kinds of school changers can be identified: those enrolling in a different public school; those enrolling in a private school. In order to identify members of these two groups, I started from two separate database, one with the geographic coordinates of pupils residence, the other with the

geographic coordinates of schools. Through the STATA command [geonear] (Picard 2010) it is possible to generate a database with the n pairs of closest neighbors, starting from the position of each object in the geodetic space. The following step is to add information on each pupil and school. The former is taken from the INVALSI database and includes all the information on pupils and family characteristics. The latter is derived from a database provided by the *Ministero dell'Istruzione dell'Università e della Ricerca* including few information on school characteristics. From this source we derived whether the school is public or private, the percentage of fixed term contracts in teaching staff, the percentage of teaching staff older than 55. Some threats to the procedure described are posed by pupils with the following situation, residence in region "A" and attended school in region "B". These asymmetry was already present in the original databases provided by INVALSI and is likely to be related to the fact that some pupils do not live in the residence address. It is the case of pupils living with separated parents. An alternative explanation is that the address collected by the school is completely misspecified. However, this group is small (about 3.5% of total sample) and we decided to exclude it from these calculations and avoid outliers inconveniences.

As a final step, the routing distance and time of travel have been computed for each pair of pupil residence/attended school. The procedure is implemented in the STATA command [osrmtravel] (Huber and Rust 2016) based on OpenStreetMap data. No problems of comparability emerges from using the routing distance instead of the geodetic one previously computed, as both are highly correlated. In fact, the school ranking is substantially invariant, but have been adjusted in the few cases needing a reordering. For the present analysis I've preferred to use the rounting distance and time of travel computed on OSM data, as it is a more reliable proxy of the distance parents account for when evaluating school alternatives.

Once the population of movers has been defined, schools can be analysed with reference to how many pupils choose a "non natural" school, that is a school different from the closest public school to home. By using the procedure highlighted above, I have implicitly designed school districts as the environment of pupils for whom a specific school is the closer one. The population of each school can be divided in two groups: a base quota, a variable quota. Each school is the natural choice for a group of pupils, those enrolled in their natural school represent the base quota of the school. However, a given amount of pupils will decide not to enroll in that school. These students belong to the variable quota. Thus, each school can be analysed according to the characteristics of the base quota, of the exiting pupils and the receiving pupils.

5.3 Empirical results

5.3.1 Who makes alternative school choices?

Table 5.1 is computed on 60,622, the sample of individuals who have no missing in all variables presented. From descriptive results emerges that making alternative school

choices is a usual condition for lower secondary students in Italy. About 54.8% of our sample students is not enrolled in its "natural school", a higher percentage that devide quite evenly the sample. On average upper class pupils are more prone to make alternative school choices than pupils from other social strata, as well as pupils from higher educated parents. Surprisingly, pupils from the lower class have the same probabilities to be a school mover than pupils with both parents belonging to the excluded category. The excluded background group is relatively small (less than 3.000 units), and the result might be due to its composition. The group is made by unemployed and never worked (especially housewifes), thus it comprehends individuals that are both "permanently" and temporary excluded. However, a check on the average ESCS for students from the excluded socio-economic background enrolled in natural or alternative schools does not highlight relevant differences. Both average values of ESCS are negative and of relevant size, the score is about -1.20. The distribution of pupils between natural and alternative schools according to their citizenship status and ethnic background highlights small differences with Italians. In particular, first and second generation migrants are choosing alternative schools with the same rate of Italian pupils. Similar results for ethnic groups. Italians are the most mobile group, but the gap with other groups is very small.

When looking at the territorial variation of the phenomenon two elements appears. First of all, the phenomenon is less strong in the North-East, the only geographical aggregate where those choosing the natural school are a majority. Second, the share of pupils enrolling in alternative schools peak in Islands municipalities. Finally, the table reports the average house-school travel time in minutes for up to ten alternative schools for each individual. Travel time by walk for up to the fourth closer school to home is in the range 10-30 minutes, with an average delta of about 6 minutes between each alternative. Thirty minutes by walk seems a reasonable discriminant threshold for school choices, if we consider that a school day generally start between 8-8.30 am. Schools higher in the home-school distance ranking show substantial increase in the house-school travel time. These school choices highlight a different situation. Despite their great distance in terms of home-school time by walk, individuals choosing these schools might use alternative transportation, reducing the related travel time.

In general terms, this table highlights some ranking in the phenomenon of school choice. On average, chosing an alternative school is the prerogative of the upper class, and parents with tertiary education but no substantial differences are found between Italians and minority ethnic groups. It seems that the phenomenon of school choice is primarily social, however this is in contrast with existing literature on educational preferences and require great attention. In addition, the population characteristics of those chosing alternative schools are constant along spatial dimensions. When checking for scale-dependence changes in characteristics of the reference population, no substantial differences are found.

When investigating type of education, school movers have a slightly pronounced

Variable		Natural	Alternative
Social class	Upper class	38.2%	61.8%
	Middle class	46.0%	54.0%
	Lower class	53.2%	46.8%
	Excluded	51.5%	48.4%
Parental education	Primary	50.6%	49.4%
	Lower secondary	50.5%	49.5%
	Upper secondary	46.0%	54.0%
	Tertiary education	38.9%	61.1%
Citizenship	Italians	44.5%	55.5%
	1^{st} generation foreigner	48.9%	51.1%
	2^{nd} generation foreigner	46.2%	53.8%
Nationality group	Italians	44.9%	55.1%
	EU-member	47.8%	52.2%
	non-EU-member	50.1%	49.9%
	Other continents	47.3%	52.7%
Geographic area	North-West	45.1%	54.9%
	North-East	51.5%	48.5%
	Centre	44.6%	55.3%
	Southern	44.3%	55.7%
	Islands	38.7%	61.3%
Total		45.02%	54.8%
Home-school travel time	1		10.6
	2		18.1
	3		23.7
	4		30.2
	5		53.9
	6		72.9
	7		86.5
	8		94.3
	9		102.9
	10		132.3

TABLE 5.1: Percentage of movers by relevant variables.

preference for public schooling, in fact the 51% of pupils choosing alternative schools enroll in another public school, while the remaining 49% chooses a private one. However, moving from the public to the private sector might have different motivations than moving exclusively among public schools. First of all, it might answer to parental beliefs like religion or others. Second, the former transition might assume the characteristics of a pulling process, where families use their resources to obtain access to private education, while the latter might be a pushing process, where low status families are forced to move among public schools due to more comfortable school-work arrangements or lack of resources for private education.

In Chapter 4 I've discussed school composition in terms of between and within class variation of the socio-economic and cultural index. Schools have been divided in five groups: the four quadrants, according to combinations of high or low between and within class variation, and private institutions. We can investigate the enrolment in alternative schools by applying this categorization. More specifically, it is possible to construct mobility tables that symultaneously account for the tipe of natural and alternative school, where rows are the starting point and columns the destination. Results are presented in Table 5.2, where the first number of each cell is the by row frequency, while the number in parenthesis is the by column percentage.[1] Column indicated by (5) represents private schools. According to this structure, it is possible to read the distribution of flows. The columns of higher interest are the first and the last. By row values in the first column shows that the 51% of pupils whose natural school is in any category are moving towards schools with weak level of segregation. The figure is higher for those coming from democratic schools (55%). Numbers in parenthesis tell about the composition, for each category, of the population of pupils chosing an alternative school. The composition of pupils is symilar across all schools, although strongly segregated schools (4) present higher shares of students from homogeneous ones rather than all other types.

Despite informative, additional information is needed in order to better grasp the nature of these flows. Thus, a second part has been added, providing information on the averager ESCS of flows between schools. The same table has been drawn, but not presented, by using population counts and relative frequences by social class, but it does not help to get a clear and synthetic idea of the phenomenon. In the bottom part of the table, cell numbers are the mean ESCS of pupils for each flow. Three interesting facts. First, the average ESCS of those going to private education (5) is very high, close to one standard deviation.[2] Weakly segregated schools are receiving, on average, students with higher status from other weakly segregated schools, and medium status pupils from other kinds of institutions. The slightly negative scores for democratic schools (2) tell that these are especially the destination of the middle class. Surprisingly, average ESCS for those coming from homogeneous schools is substantially lower. Homogeneous schools (3) intercept especially middle and lower classes. The

1. Cell numbers are relative frequences and can be interpreted as percentages.
2. ESCS is standardized with mean [0] and standard deviation [1]

former group is prevalently moving from weakly segregated and democratic schools, while the latter from homogeneous and strongly segregted institutions. While homogeneous schools are receiving groups with the lower ESCS from other homogeneous and strongly segregated schools, strongly segregated schools (4) are receiving groups with the lower ESCS from weakly segregated and democratic schools. Flows from (1) and (2) to (4) are probably leading to more segregation of the lower class, as it should have gone to weakly segregatd or democratic schools. Finally, it is clear from this table that upper class is somehow self segregating from other social groups in private education.

Finally, a third part of the table addresses the percentages of non-Italian students in the between school flows. The most noticeable fact are the extremely low shares of immigrants going to private education. Apart from this result, a clear pattern of movement of foreigners among schools is difficult to be highlighted. Moreover, when looking at absolute numbers of migrants choosing alternative schools there are no substantial clues suggesting any path dependency. Clearly, and as previously highlighted, the flows from (1) and (2) to (3) and (4) are leading foreigners from weakly segregated to highly segregated schools. In terms of the RRA model presented at the beginning of this chapter, it seems that the upper class is taking advantage from his resources in order to reduce, in broad terms, the risks associated to enrolment in public schools. The upper class is also moving across other groups, but a check on flows composition does not highlights substantial differences. However, the table is not able to clarify two important causes of choice of alternative schools: are pupils simply moving with their parents on the home-work direction? Are pupils simply moving because schools are discriminating among applications?

		Variable			Destination school	
		(1)	(2)	(3)	(4)	(5)
Natural school	Weak seg. (1)	51.2 (41.4)	22.3 (44.0)	8.6 (42.0)	3.4 (30.3)	14.4 (49.4)
	Democratic (2)	55.0 (34.1)	21.5 (32.6)	7.0 (26.4)	4.15 (28.4)	12.3 (32.4)
	Homogeneous (3)	51.6 (19.5)	19.87 (18.3)	11.07 (25.3)	8.7 (36.2)	8.7 (14.0)
	Strong seg. (4)	51.5 (4.86)	22.2 (5.1)	10.8 (6.2)	4.8 (5.0)	10.6 (4.2)
Average ESCS	(1)	.251	-.055	-.182	-.442	.935
	(2)	.156	-.044	-.030	-.405	.918
	(3)	.083	-.208	-.243	-.084	.842
	(4)	.047	-.009	-.345	-.084	1.00
Percentage non-Italians	(1)	12.9	13.9	13.3	11.1	4.6
	(2)	13.1	11.4	15.2	10.8	6.1
	(3)	11.3	11.9	12.1	13.7	7.2
	(4)	8.1	9.4	18.9	4.5	3.8

TABLE 5.2: Enrolment flows between schools and their characteristics.

The analysis of flows from natural to alternative schools calls for additional investigation on the composition of pupils population. Schools can be described according to the changes between their theoretical population (district population) and actual population (those enrolled in a given school). As previously discussed, the closest public school to home can be seen as a natural option for families, as it minimizes costs. Those enrolling in their natural school are the baseline population of that school. However, a given amount of pupils may decide to enrol in alternative schools. These students can be ascribed to a variable quota in the school population, in fact they are those moving in the schooling system. According to this distinction each school can be analysed in terms of the base quota, exiting and receiving pupils characteristics.

In Figure 5.1 each dot is a public school. On the horizontal axis the average ESCS index of the base quota, those choosing the natural school. On the vertical axis, the average ESCS index of the exiting quota, those who are enrolling in a different school instead of their natural choice. The situation is characterized by some degree of variability. Schools above the dashed line are losing pupils with a higher economic, social and cultural status than those remaining. Conversely, schools under the dashed line are losing pupils with a lower ESCS than those in the base quota. Schools above the 45° line are more numerous than those under it, certifying a sort of exodus and reorganization of the distribution of high status pupils.

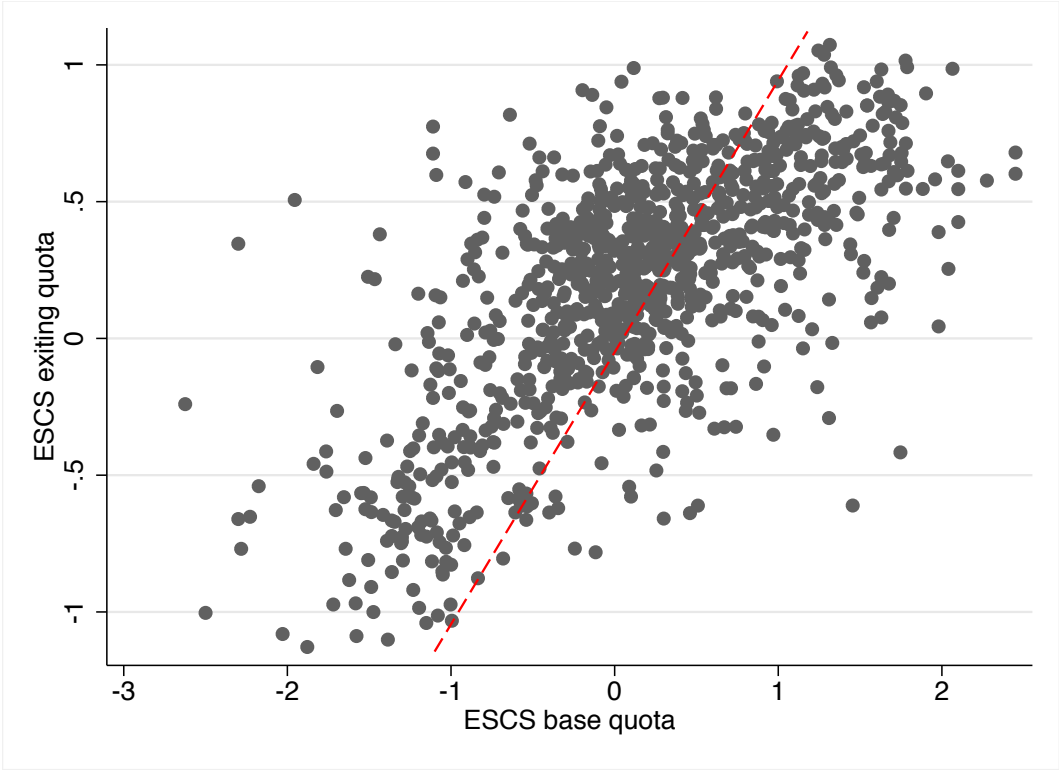

FIGURE 5.1: Public schools by mean ESCS index of exiting quota.

Figure 5.2 shows where exiting pupils are going. The dots' cloud is more compact and the two groups of schools are similarly numerous. Similarly to the previous figure, those above the line are receiving students with a higher status than the base quota. The comparison between the two figures highlights the process of homogeneization of schools on socio-economic ground. The process touches all kind of schools, as the flattening of the cloud happens indistinctly for schools from all quadrants.

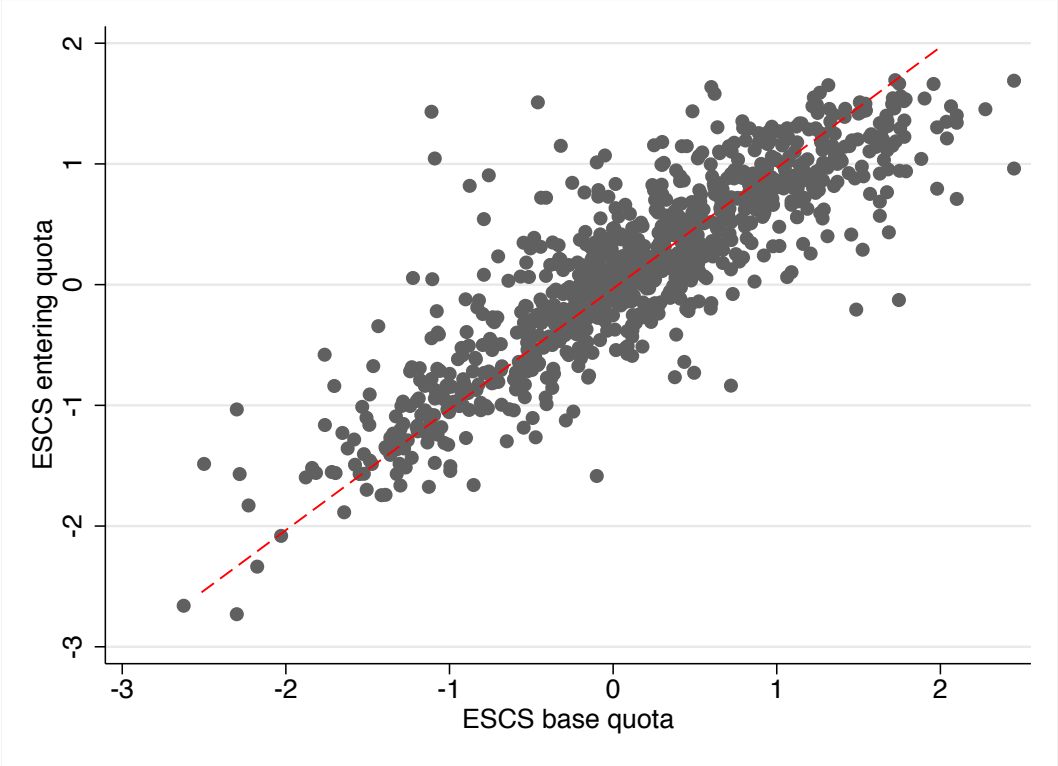

FIGURE 5.2: Public schools by mean ESCS index of entering quota.

We have already seen in Table 5.2 that non-Italians are moving between schools with no difficulties. The two aggregates of interest in this case are represented by the base and entering quota of minority groups, as they bring additional information on the phenomenon. Figure 5.3 presents the distribution of schools according to the ratio between the absolute figures of non-Italians chosing that alternative school (entering quota) and that of non-Italians enrolling in that natural school (base quota). The distribution is skewed on the right, with few schools characterized by a very disproportionate ratio. About the 65% of schools is receiving from other catchment areas up to two times the number of non-Immigrants chosing the natural school, but only 40% of schoos have a ratio lower than one. The share of entering non-Italians is not strongly connected with the average ESCS of the base quota.

The distribution of schools according to these two sub-populations is plotted in Figure 5.4. The distribution is slightly skewed, with few cases separated from the main body, and characterized by a positive correlation of .338. The two tales of the distribution highlight a symilar story: schools with extremely polarized base quota

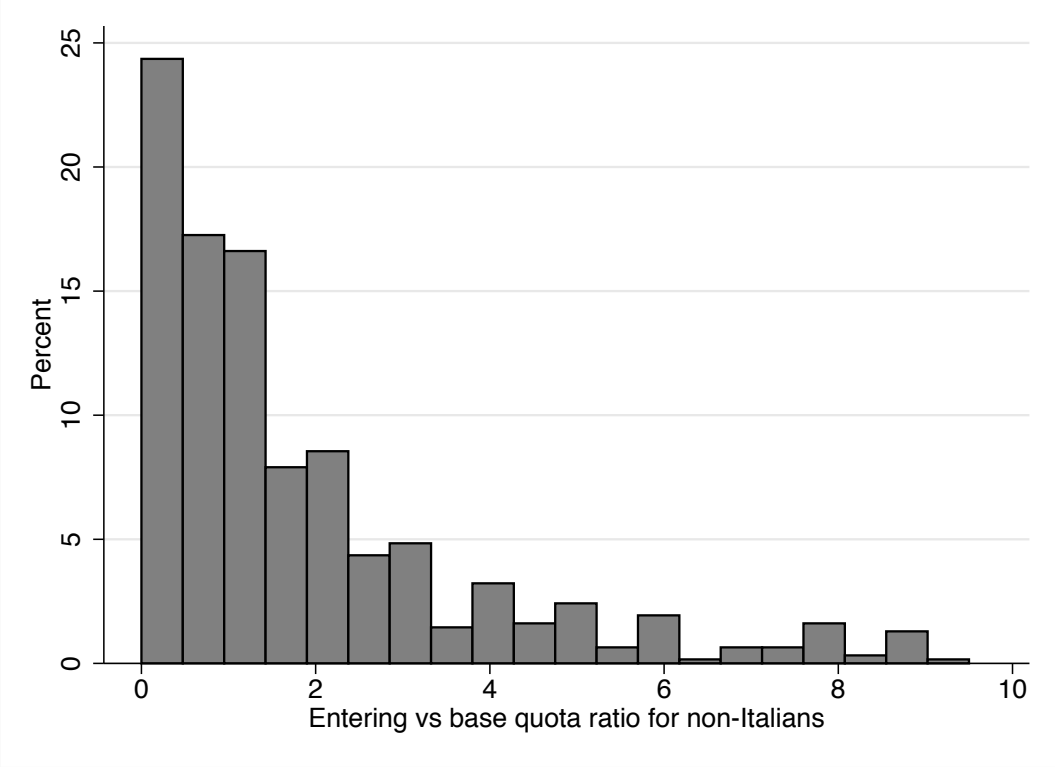

FIGURE 5.3: Entering *vs* base quota of non-Italians ratio.

ESCS, those schools that probably are more homogeneous in their composition, are receiving only small shares of non-Italians with respect to those with medium values of base quota ESCS. Since there is some degree of correspondence between the distribution of ESCS and social classes, as presented in Section 3.4.2, the story can be seen as low numbers of non-Italians are enrolling in non-natural schools characterized by high shares of lower or upper class.

5.3.2 Individual probabilities of chosing non-natural schools

The analysis presented in the previous section highlights a substantial process of redistribution of pupils between non-natural schools. The phenomenon is characterized by several facets that might catch our attention. The first aspect is the general mobility of students between alternative schools, the second is that of pupils, especially from the upper class, chosing a private instead a public school, the third involves those enrolling to a homogeneous or strongly segregated school instead of other alternatives. For each of this phenomenon it is possible to investigate individual probabilities according to a general set of covariates.

First of all, individual and family characteristics such as gender, ethnic and socio-economic background. Enrolling in alternative schools might be the result of several factors, among which the search for a school more adherent to personal beliefs, the choice of a school characterized by higher quality standards, or simply a school where

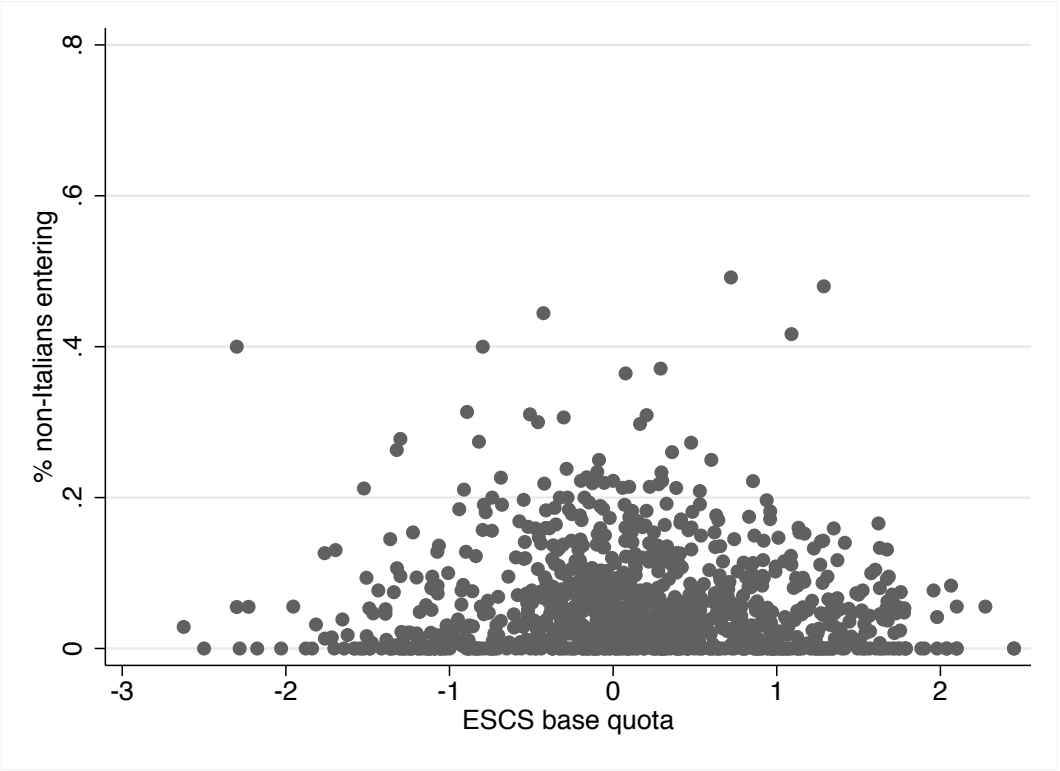

FIGURE 5.4: Schools by average ESCS of the base quota and percentage of entering non-Italians.

there are pupils from the same social extraction. According to the presented theoretical framework, families will evaluate a limited set of alternative schools where the number of alternatives is defined by family constraints (costs, house-school time travel, quality standards, etc.). The model is rooted in the assumption that families will look for schools characterized by high educational standards (quality) and a population composition that allows for socialization with individuals from a higher status. Secondly, school quality proxies. School quality can be measured in different ways, however few indicators are at our disposal to accomplish this task. We can use information on past performance, although the formal publication of clear and confrontable information on school performances and added value is not an established practice in Italy, and its likely to do not shed light on the phenomenon. Another proxy of quality is the average cheating factor of the school. This is a sort of correction that INVALSI uses for the computation of the average school score. It accounts for cheating practices by teachers by comparing the distribution of scores of each class with that of other classes of symilar composition. Thirdly, average characteristics of the student population of the natural school. Here Fourthly, variation of the phenomenon might be related to local specificities. On one side, the specificity related to the spatial characteristics of each municipality, such as the presence of natural barriers or the availability of public local transportation. On the other side, the variation in structural elements of schools, since the educational provision is coordinated and maintained at the provincial level. In this direction, Checchi and Jappelli (2003) have

shown a substantial variation at the provincial level in structural elements of schools that might be relevant in family choices.

Equation 5.1 sums up the structure of the model, computed using logistic regression as link funtion since the dependent variable is a dichotomized variable assuming values [0,1]. Logit models assume that for each unit, given a set of covariates or attributes, there is a defined probability to observe a certain phenomenon, usually conceptualized as success of observation. The model is applied to the following outcomes of interest: (i) probability to choose an alternative school; (ii) probability to choose private education *vs* enrolling in the natural school; (iii) probability of choosing a segregated school *vs* enrolling in the natural school.

$$\begin{aligned} Pr\{y_i = 1|X_i\} = \alpha + \beta_1 Gender + \beta_2 ParEduc + \beta_3 Class + \beta_4 Immigrant \\ + \beta_5 Siblings + \beta_6 Books + \beta_7 Traveltime \\ + \beta_8 School.past.performance + \beta_9 Cheating \\ + \beta_{10} nonItalians.within.dev + \beta_{11} Type.segregation \\ + \beta_{12} ESCS.base.quota + \beta_{13} Geographic.area \\ + \beta_{14} Provincial.dummies + e_i. \end{aligned} \tag{5.1}$$

The model includes several covariates. Family background variables (parental education, socio-economic class, immigrant background, number of siblings, number of books in the house), home-school travel time, school characteristics (school performance as 2011 INVALSI score, cheating score, within school variation in the share of non-Italians, type of segregation of the school, base quota ESCS), geographic area and provincial dummies. Model1 refers to probabilities of enrolling in an alternative public school, while Model2 refers to probabilities of chosing private education. Average marginal effects are presented in Table 5.3 for the complete models, along with model fit information summed up by pseudo R^2, sensitivity and specificity indices. The number of observations is relatively low with respect to the complete sample, as already discussed this issue is due to attrition on different variables.

Average marginal effects for Model1 highlight the relevance of family socio-economic and cultural capital in shaping probabilities to enroll to alternative schools. Pupils whose parents hold educational titles lower than tertiary degrees are more likely to be enrolled in the natural school. The maximum penalty (-.142) is registered for pupils whose highest parental educational title is primary education. Symilarly when looking at social position. Compared to the upper class, the lower class is experiencing a substantial penalty (-.184) in possibilities to enrol in alternative schools. Also number of books in the house has a positive impact on the outcome variable of Model1. In general terms, pupils from lower class families and lower educational titles are more likely to be enrolled in their natural school. Results suggest that being enrolled in alternative schools is associated with high socio-economic and cultural status. This seems to corroborate the hipothesis of strategic action of higher strata with respect to

education. Finally, the coefficient associated to being non-Italians is not significantly different from zero, suggesting that on average probabilities of alternative enrolment for non-Italians are not different from those of the majority group.

The effect of number of siblings is negative, but not always statistically significant. This result is in line with average marginal effects from Checchi and Jappelli (2003). The increasing number of siblings can be viewed as increasing costs on several assets, especially transportation. Thus, more siblings increase the costs of chosing alternative schools. The effect of home-school travel time by walk for the natural school is positive and tells that an increase of one minute in the time needed to go from home to the natural school increases the probability to chose an alternative school by .004. The higher the home-school time by walk, the higher the propensity to choose an alternative school.

About characteristics of schools, information on past performance, cheating and within school variation in the amount of non-Italians are not relevant determinants of school choice. In general, the effect is not statistically different from zero. However, school composition seems to provide a more relevant contribution to the cause. The average marginal effect associated to type of school as defined by the four quadrants discussed in the previous chapter is positive and significant. Probabilities of chosing alternative education are higher by a factor of .311 when the natural school is strongly instead of weakly segregated, .356 when it is a homogeneous school and .092 when it is a democratic one. From these results emerged that choosing an alternative school is also a matter of the nature of the natural school. On average, families prefer to move from strongly segregated and homogeneous schools, although this analysis is not able shed light on destinations. Finally, the socio-economic composition of the base quota has a negative and significant effect. On average, the higher the status of those staying in the natural school, the lower the probabilities of chosing an alternative school. This result represents another clue pointing at movement between schools as a socially driven phenomenon.

Model2 and Model3 analyse the effect of the same set of covariates on two particular choices: private education and strongly segregated and homogeneous schools. These choices are of interest as the more extreme in the distribution. About Model2, results from previous discussions have shown that private school is a sort of self-segregation of the upper class from other social groups, in line with previous research. Thus, I analyse probabilities of making an extremely selective choice. Model3 is about moving from any natural school to strongly segregated or homogeneous schools, that are the most segregated public schools. Previous analysis have shown that part of these schools is characterized by segregation of the lower class, making it a phenomenon of interest.

About Model2, significance of coefficients are symilar to Model1. On average, males have slightly lower probabilities to choose private education, although the coefficient is close to zero. Average marginal effects highlight the classical ranking in socio-economic and cultural capital: pupils from the upper class are those enrolling

more in private schools, as well as those from parents with tertiary education. However, the penalties in average probabilities associated to the belonging of lower social classes are small: -.046 for the middle class, -.081 for the lower class. The penalty involves also non-Italians, characterized by a negative factor of about .02. Finally, about family characteristics, the number of siblings has the expected negative value. Increasing numbers of siblings are associated to decreasing probabilities of being enrolled in a public school. The negative coefficient highlights that budget considerations are taken into account by families. When coming to school characteristics, the only relevant contribution is that of base quota ESCS. The average marginal effect is negative, significant although with small size. In general terms, an higher socio-economic and cultural status of the natural school is confirmed as a factor influencing parental choices towards the staying option.

About Model3, that does not include provincial dummies due to perfect prediction of several observations, no one of the relevant variables are significant and with large effects. The only significant variable is type of natural school, characterized by a small positive effect in all cases. It seems that the model is not able to explain choices of segregated schools. This might be related to model misspecification, as enrolling in highly segregated alternative schools might be just the result of factors not intercepted by the model. On the other hand, differences between groups are not significant.

Finally, the probabilities of enrolling in any public alternative school by social class and type of natural school are presented in Figure 5.5, while those choosing private education are presented in Figure 5.6. In the first figure, the effect of type of natural school varies in intensity between social classes. While the effect of homogeneous and strongly segregated schools by social class is very strong and remain somehow constant, that for weakly segregated and democratic schools is stronger for upper and middle class than lower and excluded. This might be related to the strategic nature of enrolling in alternative schools. In the second figure, the same effects are plotted for the probability to enroll in private education. Here, the by class effect of nature of natural school is smaller than the previous and particularly significant for upper and middle class. The results seems to corroborate the hypothesis of relative risk aversion models, upper classes are looking for educational provision minimizing, by means of several factors, the risk of downward mobility.

When looking at the marginal effect of base quota ESCS of the natural school on probabilities of being a mover or enrolled in private education, according also to type of segregation of the natural school, the situation is somehow flattened as presented in Figure 5.7 and Figure 5.8. Apart for the upper class, other socio-economic groups have same marginal effects for base quota characteristics across type of school. In the first figure, the probability of being a mover for the upper class is reduced on average by .08 for an increase of one standard deviation in ESCS base quota of weakly segregated schools while it is only .05 for strongly segregated ones. It appears that the upper class is more responsive to characteristics of the base quota if the natural school

Variables	Model1	Model2	Model3
Gender (males)	.003	-.006*	.003*
Parental education (ref. Tertiary education)			
Primary	-.143***	-.055***	-.013*
Lower secondary	-.134***	-.043***	-.005
Upper secondary	-.083***	-.020***	-.002
Social class (ref. Upper class)			
Middle class	-.088***	-.046***	-.002
Lower class	-.155***	-.081***	.001
Excluded	-.184***	-.093***	-.005
Immigrant background (ref. Italians)			
non-Italians	.003	-.020***	-.002
Number of siblings (ref. 0)			
1	-.036***	-.024***	.001
2	-.035**	-.022***	.004
3	-.038*	-.005	-.007
4 or more	-.002	.017	.009
Number of books at home (ref. 0-10)			
11-25	.009	.007	.001
26-100	.028*	.012*	-.002
101-200	.050**	.013*	-.005
more than 200	.083***	.015*	.001
Home-school travel time	.004***	-.000	.000
2011 school INVALSI (past performance)	-.001***	.000	-.000*
2011 school cheating score	.035	-.002	.007
Within school non-Immigrants deviation	-.114	-.013	-.022
Segregation of natural school (ref. Weakly seg.)			
Democratic	.092***	-.000	.006**
Homogeneous	.356***	.014**	.022***
Strongly seg.	.311***	.032*	.022**
ESCS base quota	-.080***	-.008**	.003
Geographic area (ref. North-West)			
North-East	.044	.501	-.006
Centre	.189	.189	.008
South	-.085	.254	.007
Islands	.826	-.025	.458
Provincial dummies	YES	YES	NO
Model fit			
N	26,880	25,607	19,731
Pseudo R^2	.100	.127	.092

Note.-***p-value<.001, **p-value<.01, *p-value<.05.

TABLE 5.3: Average marginal effects for model presented in equation 5.1.

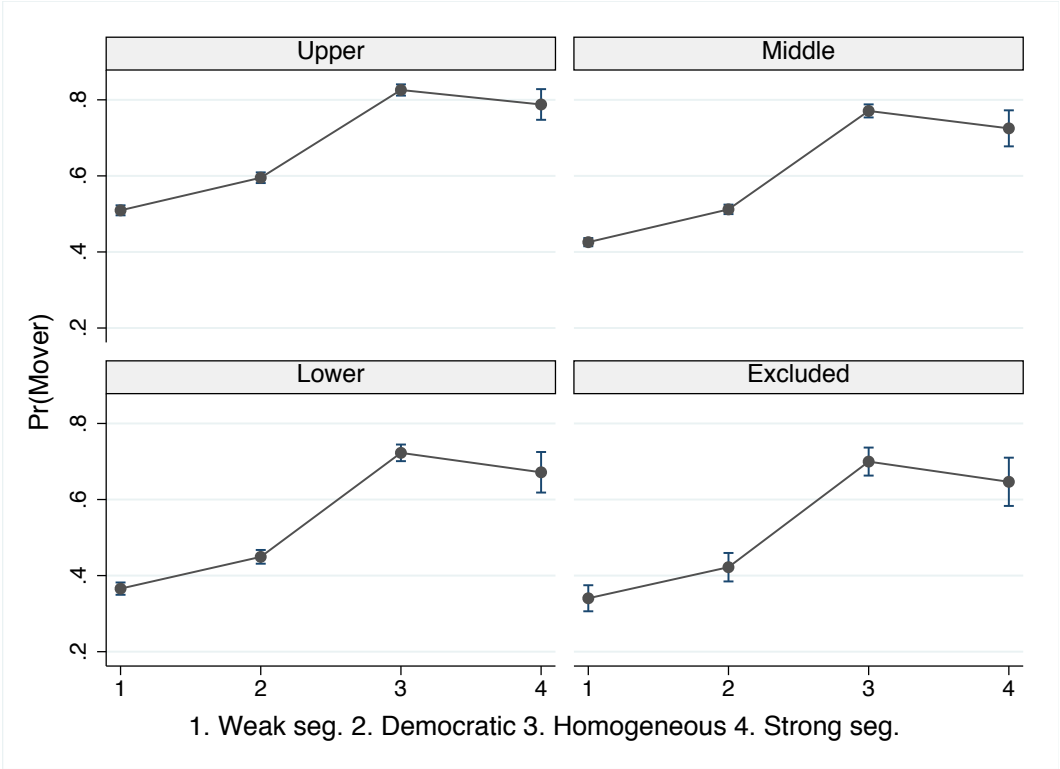

FIGURE 5.5: Probabilities of being a mover by class and type of school.

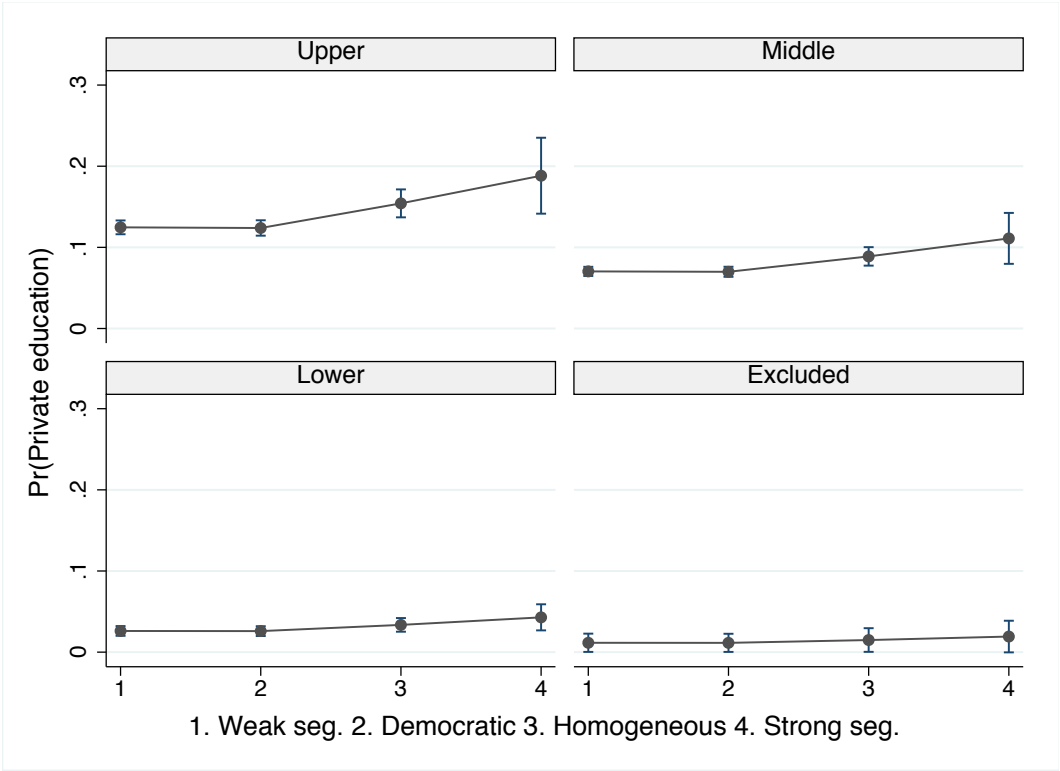

FIGURE 5.6: Probabilities of being enrolled in private education by class and type of school.

si weakly segregated than homogeneous or highly segregated ones. This is clearly

related to composition of schools, as homogeneous ones are usually more populated by lower class pupils. However, weakly segregated schools probably better satisfy the preferences of the upper class in terms of composition and possibilities to get in contact with certain groups. The second figure shows the marginal effect of base quota ESCS on probabilities of being enrolled in private education by considering type of segregation of the natural school. The effects are very small and no significant differences are found between classes. Clearly, the upper class appears as the more responsive social group to the composition of base quota of natural schools, however the average penalty is close to -.02.

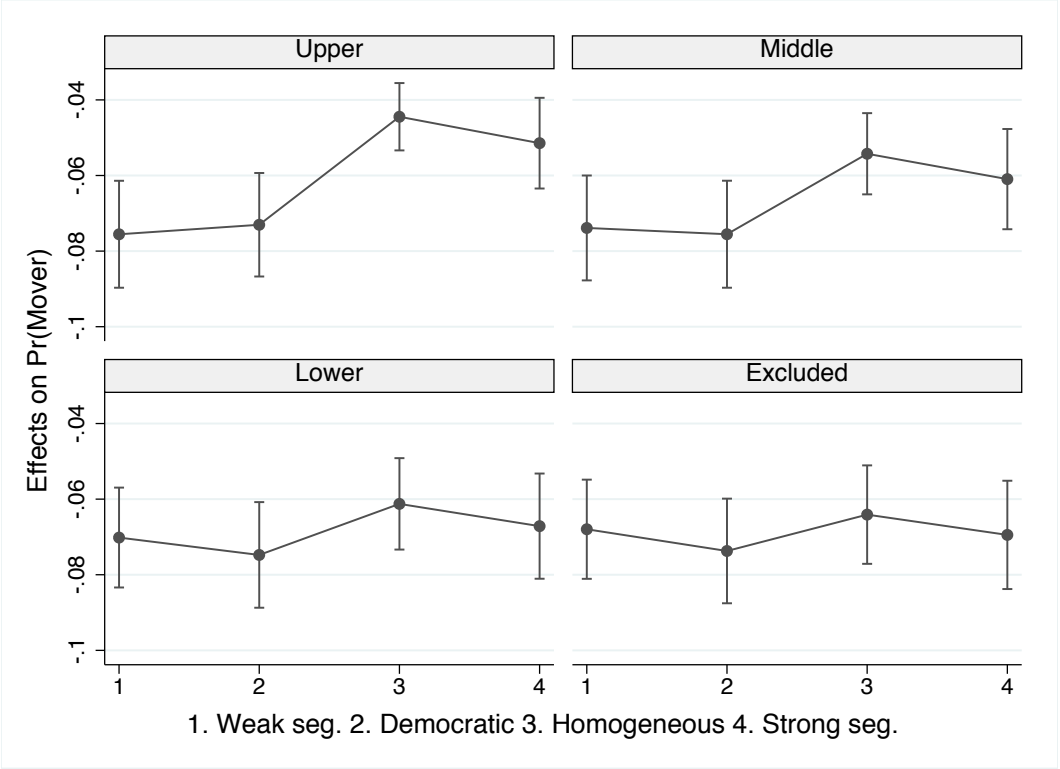

FIGURE 5.7: Effect of natural school's ESCS base quota on individual probabilities of being a mover by class and type of school.

5.4 Conclusions

This chapter contributes to the literature by investigating school choice dynamics in lower secondary school and its determinants. The movement of pupils is analyzed with respect to enrolment, and the school minimizing home-school distance is expected to be the natural choce for families. Enrolment in alternative schools is analyzed with reference to natural schools. The research questions of this chapter aim at investigating flows of pupils from natural to alternative schools at the time of enrollment, by analyzing the actual composition of schools as the result of redistribution of pupils. Upper class pupils are those enrolling the most in alternative schools, while no relevant differences have been found between Italians and ethnic minorities.

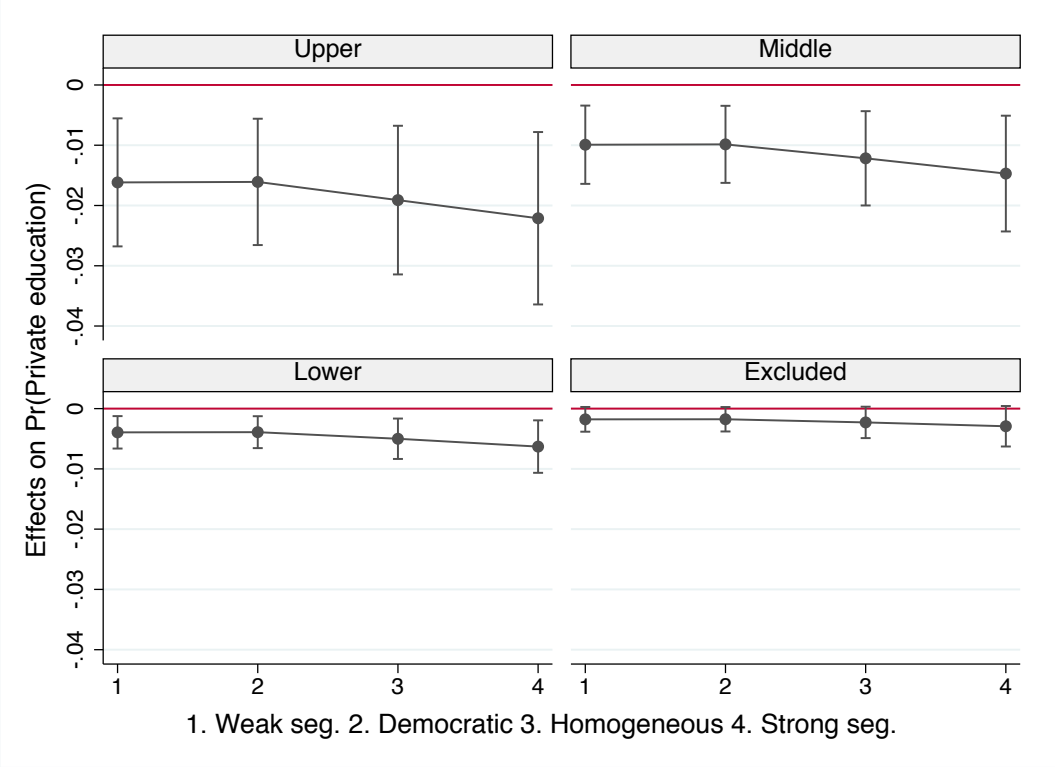

FIGURE 5.8: Effect of natural school's ESCS base quota on individual ESCS on probabilities of being enrolled in private education by class and type of school.

The redistribution of pupils appears to happen especially on socio-economic bases. Two analysis of flows of pupils have been carried out. The fist involve mobility tables and students' enrolment is analyzed by looking at the characteristics in terms of segregation of both natual and destination schools. The flows show that, on average, lower ESCS are moving especially towards homogeneous and strongly segregated schools, in particular those whose natural school is weakly segregated or democratic are going to strongly segregated schools, while lower ESCS from homogeneous and strongly segregated schools are going to homogeneous schools. When this phenomenon is analysed by ethnic composition, no relevant differences are found.

By means of logistic regression, the probabilities to choose an alternative school; to choose private education *vs* enrolling in the natural school; or to choose a segregated school *vs* enrolling in the natural school have been analyzed. The model includes several covariates. Family background variables (parental education, socio-economic class, immigrant background, number of siblings, number of books in the house), home-school travel time, school characteristics (school performance as 2011 INVALSI score, cheating score, within school variation in the share of non-Italians, type of segregation of the school, base quota ESCS), geographic area and provincial dummies.

In general terms, family background characteristics are relevant in shaping probabilities of moving in general and moving to private schools, as well as parental education and cultural capital measured by number of books at home. Number of siblings negatively affect probabilities of being enrolled in alternative schools, as additional siblings increment costs associated to transportations, tuition fees in private education, and changes the opportunity cost of each evaluation. When looking at school characteristics, school perceived quality as measured by past performance has a very small effect. Here, the characteristic capturing variation seems to be type of segregation inside the school: the stronger the segregation of natural school the higher the probabilities of going to another school. Interestingly, the middle class seems to behave similarly to the upper class. This might be related to imitative strategies, suggesting that the hypothesis of relative risk aversion is somehow corroborated.

This analysis is limited by the lack of several relevant information that might help in reducing variation of results. The most relevant element is represented by absence of information (of any kind) on parental job place. In fact, at the moment it is not possible to distinguish false positives: those pupils that are moving because they are attached to parental place of job instead of moving because of quality or strategic decision. The analysis of determinants of segregation is also affected by measurement errors due to difficulties, at the moment of writing, to properly estimate school quality. The information available is considered as proxy of perceived school quality (i.e. school past performance) but is difficult to consider it as a reliable measure. Perceived school quality as measured by proper scales should contribute more in disentagling enrolment decisions.

Chapter 6

Effects of school composition

In Chapter 1 we have defined segregation as a more or less institutionalized form of social distance that translates into physical separation of space, producing homogeneous environments with respect to some individual characteristic. A relevant stream of research has been devoted to investigate effects of living, growing and being educated in a segregated social environment on individual outcomes and behaviours. Concerning research on effects of school segregation on school performance (usually measured through standardized tests) a number of studies on the US and little research on European countries are available.

Hanushek (1972) and Hanushek, Kain, and Rivkin (2009) have investigated the effects of racial composition on test scores using Texas' schools. After controlling for family background, the results show that black enrolment shares adversely affects achievement of both black and white students and that the penalty is twice as large for blacks as for whites. Similar results have been provided by Bankston III and Caldas (1996) for Louisiana's schools, while Vigdor and Nechyba (2007) provided evidence of the effects of peer characteristics on student achievement in North Carolina elementary schools. As a general result, minority concentration exercises a negative impact independently of behavioural patterns and socioeconomic composition of schools. Hoxby (2000) has investigated the effect of both ethnic and sexual composition of classrooms in the US, finding evidence of a positive effect of the share of females in class and confirming previous evidence of the negative effect on test scores of a high proportion of black students.

An alternative way of looking at the effect of the racial composition of schools is to investigate the effects of desegregation policies. It has been documented that black students as a group, in the US, perform worse than white students. This stream of literature is devoted to investigate whether black students in desegregated schools are performing better than their similars in segregated schools. This approach has lead to several studies summed up in a review by Crain and Mahard (1978). The authors suggest the presence of a positive but small effect of desegregation policies on average gain in performance for blacks in the US. Mixed evidence on both achievement and differences in behaviours and attitudes toward school and among blacks and whites is also the result found by Schofield (1995) in his literature review on desegregation effects in primary and secondary schools. Similar results were found by

Simmons et al. (1978) for grade 6 and 7. The key element in evaluating the effects of desegregation is that such practices can't be considered as laboratory-controlled experiments identical across different settings. The uniqueness of each policy plan and the context in which it is applied is determinant in shaping the results.

While evidence of the effects of school racial composition is mixed, the one for socio-economic composition is more clear (Rumberger and Palardy 2005a; Palardy 2013), although its effects may differ for attainment outcomes as compared with achievement outcomes. We refer to *attainment* when talking about obtaining an educational title at the end of a grade, while the term *achievement* usually refers to the educational performance. In this case, achievement refers to grades (marks) and scores in standardized tests. The reason of such a difference might be the presence of different mechanisms at work (Rumberger and Palardy 2005b). The average socio-economic background of peers is likely to substantially influence school achievement (Kahlenberg 2001; Hanushek et al. 2003). Moreover, the composition effect might not be the same across ethnic or socio-economic groups. In their review, Mayer and Jencks (1989) found little evidence on this issue. The average socio-economic school composition may affect black students more than whites and higher-status students more than lower-status ones for high school, but there is no substantial effect for primary school. Here, the black box of socialization might hide the explanation. With respect to attainment, studies have focused on the effects of school composition on college enrolment (Perna and Titus 2005) and dropout (McNeal 1997; Rumberger and Palardy 2005b). While a positive relation has been highlighted for college enrolment (Engberg and Wolniak 2010), no effects have been found for dropout rates (McNeal 1997).

Coming to Europe, literature provides few cases. Ammermueller and Pischke (2009) found socio-economic composition of classrooms to be associated with PIRLS scores in grade four in six countries (Germany, France, Iceland, Sweden, Norway, the Netherlands). Using PISA data Brunello and Rocco (2013) investigate this association for ethnic composition in secondary school in cross country comparison. They identify a negative correlation between higher shares of immigrant pupils and achievement tests of natives. However, the effect is small and varies according to socio-economic status: natives with a disadvantaged socio-economic background suffer more than their advantaged-background class mates, a result in line with Mayer and Jencks's (1989) prediction. Using Swedish data, Szulnik and Jonsson (2007) found a small and negative effect of schools' ethnic composition on grades.

Mixed evidence is available for France. Felouzis (2003) estimates the gap in students' performance due to concentration of students with immigrant background in about 0.4 points on a scale ranging from 0 to 20. Using French secondary school data Cebolla-Boado (2007) finds no effect of the share of foreigners on various educational outcomes in lower secondary schools. However, estimates vary by the estimation technique used and caution is suggested in interpreting these results. Similar evidence is available for primary school in Spain (Cebolla-Boado and Medina 2011).

Studies on peer effects are concord in showing that once social individual characteristics are controlled for, the average effect of ethnic classroom composition disappears. However, it might be possible that the effect persists given a certain threshold in ethnic composition. Cebolla-Boado and Medina (2011) investigate this particular issue finding that the effect of school composition is significant only if immigrants represent at least one fifth of the student body. To the author's knowledge, no other studies address the threshold issue in composition effects.

Coming to Italy, Contini (2013) has investigated the effect of the concentration of students with immigrant origin on student learning in primary and lower secondary schools using 2010 INVALSI data. Consistently with previous literature, the author identifies a weak negative effect of the concentration of ethnic minorities on educational performance of pupils. Moreover, when considering classrooms, the concentration of first generation immigrants has a larger impact on students of immigrant origin than on natives. This is especially true for the Italian test. Among natives, those who are negatively affected the most are those from the lowest socio-economic background. Cardone, Falzetti, and Severoni (2015) run a similar analysis on pupils enrolled in grade 5 (last year of primary education) in 2012. They show that students enrolled in schools with segregated classes[1] perform worse than peers in not (or less) segregated classes. The gap is pretty the same across the distribution of ESCS, thus high ESCS students are affected as much as low ESCS peers. Interestingly, this result is not in line with previous research and may be a specificity of Italian primary school, or being the consequence of some degree of model misspecification. For example, the model presented in this study accounts only on structural characteristics of schools and does not control for individual characteristics of pupils.

6.1 Research questions and hypothesis

In the previous chapters the multidimensionality of the school segregation phenomenon has been addressed by investigating variation in school composition, the characteristics of pupils flowing from natural to alternative schools and probabilities associated to certain kinds of movement conditional to family background and school characteristics, expressed by quality and students characteristics. The observed school composition is the result of the redistribution of pupils between schools, and that of strategic behaviour of families, especially those from the upper and middle class. The most evident case is that of private education, characterized by strong self-segregation of the upper class in this track. The aim of this chapter is to investigate the effect of school composition, as analyzed in previous chapters, on individual performance (INVALSI test score) and school transitions to upper secondary education. The following research questions are addressed:

1. The authors identify segregated schools as those in which the between-classes variability of the index that measures socio-economic and cultural background (ESCS) is higher than the national mean.

- Is there an effect of class composition and school segregation on the INVALSI test at the individual level?
- Is there an effect of class composition and school segregation on the probability to do any school transition?

Researchers have highlighted two competing mechanisms to explain variability in the effect of school composition on the educational performance of pupils. The first mechanism is related to structural characteristic of schools. Schools in which pupils with immigrant and disadvantaged social background are more represented differ significantly to other schools in their structural characteristics (Cebolla-Boado and Medina 2011). In general terms, schools in poor neighbourhoods may be as deprived as the surrounding environment, thus negatively influencing results of pupils. However, there is little and mixed evidence about the strength of this argument. Similarly to school effects, some authors have suggested the importance of class-level effects. Duru Bellat and Mingat (1997) indicate that in this case, teachers might adapt their requirements and demands to the average level of pupils. Thus, controlling for school and teacher characteristics the effect of segregation might diminish.

The second mechanism is *peer to peer* interaction, or the social interaction between individuals in the same environment. In its seminal work, Mansi (1993) distinguishes between endogenous peer effects (those that derive from peers' outcomes) and exogenous peer effects (those that derive from peers' background). Researchers have traditionally encountered difficulties in the proper identification, or estimation, of peer effect due to the difficulty to disentangle it from family and school characteristics (Hanushek et al. 2003). Moreover, peer effects might vary among individuals. The usual and simplest strategy adopt by researchers on this topic is to consider average peer effects, while a more demanding strategy requires the collection of precise information on school relationships among peers.

6.2 Data, variables and methods

Information from the INVALSI partition of the database is used. Variables of interest describe the socio-economic condition of the houshold of origin, migratory background, and compositional variables at the school level. Consistently with previous analysis, socio-economic condition of pupils is defined by social class, while variables for cultural capital (parental education, number of books in the house) are added in order to have a complete picture. Migratory background is registered by a dummy variable. School segregation is measured by the between classes component o of a same school, while school and class composition is measured by means of percentages of non-Italians and average ESCS. The dependent variables are individual math and italian performance (INVALSI test score) and type of upper secondary for school transitions.

A substantive strand of literature identifies multilevel models as a useful tool for the isolation of compositional effects. The starting point of this literature is that individual outcomes are the result not only of personal characteristics, but also that of factors influencing the individual. Following this broad area of research I opted to implement a multilevel model that accounts for both school and classes characteristics, in the attempt to consider both differences between and within schools. Thus, a three-level linear model is implemented following this formulation:

$$y_{ics} = \beta_0 + \beta_1 z_{ics} + \beta_2 \bar{z}_{cs} + \beta_3 k_s + \mu_{cs} + \mu_s + \varepsilon_{ics}, \tag{6.1}$$

where y_{ics} is the individual INVALSI math or Italian test score. Each individual i is nested in classrooms c, nested in school s. z_{ics} are individual characteristics: gender, family background expressed as socio-economic class, immigrant background, cultural capital measured as parental education and number of books at home. $\bar{z_{cs}}$ are average class characteristics: percentage of non-Italian students, percentage of females, average ESCS score. Finally, k_s is a parameter capturing school composition. More specifically, this parameter expresses the position of the school in the quadrant scheme presented in Chapter 3 and defined according to variation of ESCS within and between classes of the same school. The error term is made by three components, all capturing the random part of the model. ε_{ics} captures variation of the pupil belonging to class c in school s from the class mean, μ_{cs} captures the deviation of class mean from its school mean, while μ_s captures the deviation of the school meand from the grand mean of the sample. The error components can be considered as elements capturing unobserved characteristics. Unobservables at the school level are factors related to organizational aspects, quality of structures, additional activies and curriculum organization. Unobservables at the class level as usually teachers' quality. According to Agasisti et al. (2016), the expectation is that class level explains a relevant part of variation, higher that school level variation. Corroboration of this hypothesis means that variation inside schools is more relevant in shaping individual performances than variation between schools. In other terms, attending certain classes is more relevant than attending certain schools. At the individual level, males are expected to perform better in mathematics while females in Italian, and socio-economic background is expected to have a significant impact fostering performances of high status pupils. On the contrary, non-Italians are expected to suffer substantial penalties in individual performances with respect to natives. According to Contini (2013), the share of non-Italians at the class level is expected to have a negative impact on individual performances with heterogeneous effets over socio-economic groups: lower class pupils suffer ethnic concentration, while upper class pupils are expected to benefit.

Probabilities of a certain school transition to upper secondary are investigated through a multinomial logit model where the dependent variable expresses the track to which one can enroll: (i) *Liceo*; (ii) *Tecnico*; (iii) *Professionale*.

$$Pr\{y_i = 1,2,3|X_i\} = \alpha + \beta_1 Gender + \beta_2 ParEduc + \beta_3 Class + \beta_4 Immigrant + \beta_5 Books + \beta_6 Type.segregation + \beta_7 Math.score + \beta_8 ClassESCS + \beta_9 \%non.Ita + \beta_{10} \%Females + \beta_{11} Type.segregation + \beta_{12} Geographic.area + e_i. \quad (6.2)$$

The model is the same of that presented in equation 5.1 in Chapter 4, with the addition of individual lower secondary exam score and class average characteristics from equation 6.1: percentage of non-Italian students and average ESCS score. In order to address the problem of school selection bias I refer to clustered errors at the school level. This means relaxing the assumption that observations are independent at the school level, allowing for intragroup correlation. Here, family background is expected to have a strong effect on transitions. Upper class families are expected to have higher probabilities of transition to *Liceo*, the more academic oriented track, in line with existing research on the association between social origin and education. If socialization is a mechanism through which individual outcomes are influenced by peers outcomes, behaviours and expectations, individual transition is likely to be affected by class composition.

6.3 Empirical results

It is well established that the INVALSI test score is associated to individual characteristics. Previous literature have highlighted the role of immigrant and socio-economic background, and found territorial differences. Figure 6.1 presents the math test score stratified by geographic area and two background characteristics: immigrant background and social position. In general terms, children with immigrant background and from lower social strata are associated with lower scores. From the boxplot emerge some differences between socio-economic groups, the upper and middle class are characterized by similar distributions as well as the lower class and the excluded. Non-Italians are associated to lower scores in math, a penalty persistent across all social classes. Differences in the distribution of test scores by immigrant background are more evident in the South and Islands, typically places of first arrival of migrants from Africa. It is worth noticing the case of Centre Italy, where the distributions of test scores by immigrant and socio-economic background are very similar. Possible explanations for such territorial differences are rooted in unobservable school characteristics like teachers' quality, or socialization processes difficult to measure.

Table 6.1 reports correlation among average INVALSI test scores at class level and percentages of pupils according to background characteristics of interest. The results

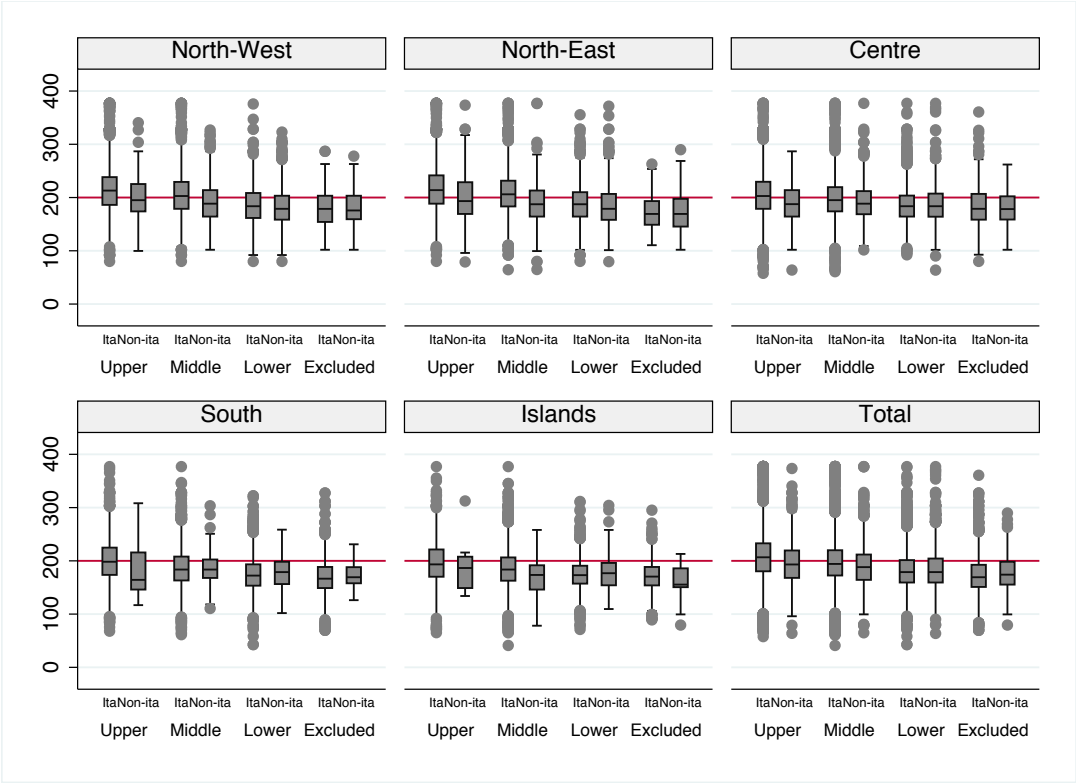

FIGURE 6.1: Boxplot INVALSI score over immigrant and socio-economic background by geographic area.

are for public schools. First of all, there is a positive and significant correlation between the school test score in math and Italian, in line with positive correlation of the same size at the individual and school level. The presence of higher shares of upper class is positively associated with average test scores in mathematics and Italian. Both correlations are substantial in size and statistically significant. The percentage of middle class is poorly associated with test scores, while classes score worst when there are higher shares of lower class and excluded. Onlu average class scores in mathematics seem to be not correlated with the percentage of non-Italians. When checking this figures for school composition, results are confirmed.

Figures 6.2 and 6.3 visually presents results of Table 6.1. The former figure presents the distribution of classrooms of public schools according to their composition and average math score, with the overlapping of linear fitted values. The latter figure verify the consistency of the same distributions at the school level. The relationships identified hold for all combinations.

6.3.1 Effects of school and class composition on individual performance

Model in equation 6.1 is estimated via maximum likelihood, where all the error terms are assumed to be independent and to follow a normal distribution. For this exploratory analysis the sample is represented by 40,925 individuals, nested in 3,306

	1	2	3	4	5	6	7
1. Av. math score	1.000						
2. Av. Ita score	.411***	1.000					
3. % Upper cl.	.423***	.156***	1.000				
4. % Middle cl.	.029	.093***	-.416***	1.000			
5. % Working cl.	-.337***	-.169***	-.529***	-.388***	1.000		
6. % Excluded	-.282***	-.171***	-.273***	-.303***	-.012	1.000	
7. % non-Italians	-.080***	-.025	-.118***	-.094***	.274****	-.066***	1.000

Note.-***p-value<.001, **p-value<.01, *p-value<.05.

TABLE 6.1: Correlation matrix of average school test scores and school composition.

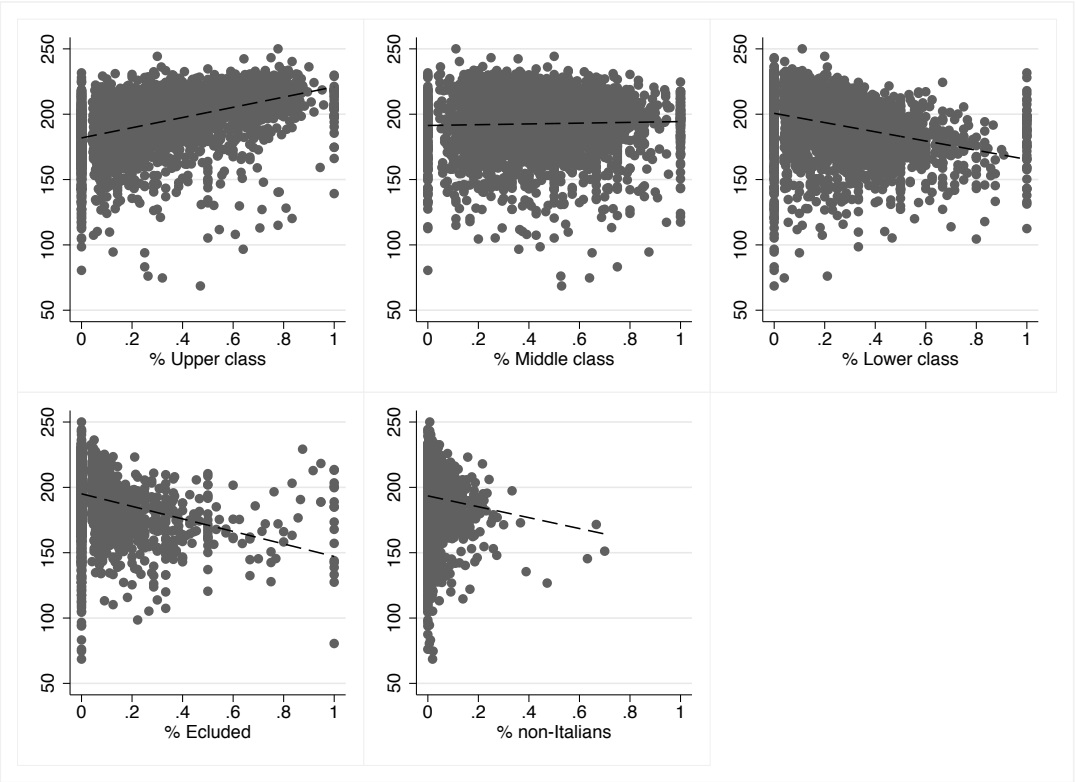

FIGURE 6.2: Average class score in math by school composition.

classes in 671 public schools. The dependent variable, standardized test score, is centred in the value of 200. Thus, effects have to be read as changes in the absolute score. In line with results from previous research on this theme, females experience a penalty in math, while are doing better than males in the Italian test. The usual ranking between socio-economic groups is found in both regressions. Performance for the middle class is not statistically different from that of the upper class, while lower class pupils and those with parents in the excluded category are performing worst. Children with high cultural capital, measured by means of parental education and number of books in the house, are performing better. All coefficients are substantial in size and statistically different from zero. Finally, the achievement of native students is higher than that of non-Italians. However, the penalty is not very large,

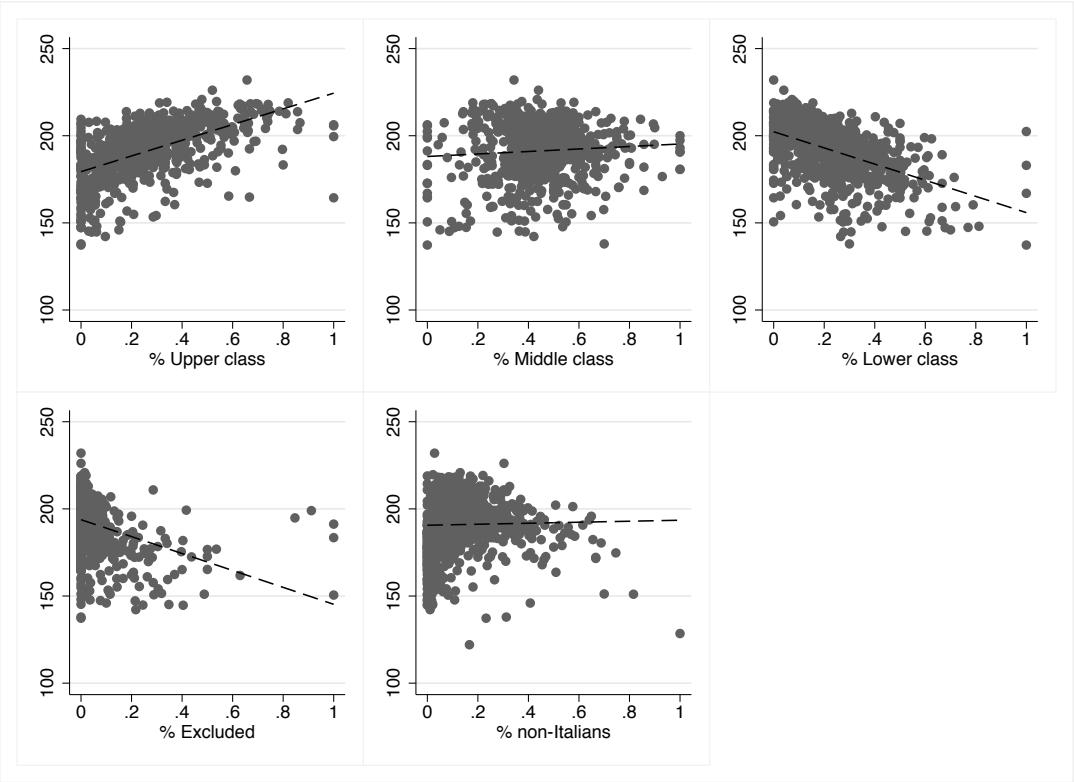

FIGURE 6.3: Average school score in math by school composition.

although statistically significant.

When coming to composition effects, the average class level ESCS affects individual performance on both math and Italian. However, while the impact is positive as expected on math score, it is negative for the Italian test. The result might hide substantial variation. Agasisti et al. (2016) highilighted a negative contribution of mean class and school ESCS to individual performances, but their analysis is preformed on different models for north-centre-south of Italy. On the contrary, the national level analysis carried out by Contini (2013) found no significant effect for mean ESCS at the class level. The percentage of non-Italians in classrooms has a significant effect on the achievement of mathematics, but a non significant effect on that of Italian, an inverse result with respect to that from Contini. Finally, type of school segregation is substantially not significant. This term is categorizing the variation of between and within classes ESCS for schools. Everything being equal, being enrolled in schools that are weakly segregated, democratic, homogeneous or strongly segregated does not affect individual performances. The information from this term is probably captured by class composition, although homogeneous schools have a slightly significant effect for the performance on the Italian test. Finally, geographic area variable captures the between school variation at the geographic level. The substantial result from this variable is that schools from the South and Islands are negatively contributing to individual performance on both tests.

Variables	8th Math		8th Ita	
	Null	Full	Null	Full
Gender (F)		-6.968***		8.282***
Social class (ref. Upper)				
Middle		-.519		-.613
Lower		-5.470***		-5.779***
Excluded		-7.662***		-8.852***
Parental education (ref. Tertiary)				
Primary		-20.220***		-22.516***
Lower secondary		-17.282***		-16.366***
Upper secondary		-10.339***		-8.418***
N. of books at home (ref. 0-10)				
11-25		3.865***		4.252***
26-100		10.533***		11.818***
101-200		15.119***		17.762***
more than 200		19.687***		22.797***
Non-Italian		-4.985***		-12.887***
Mean ESCS		2.468***		-5.760***
% non-Italians		-20.781*		16.716
% Females		-1.137		-7.292
School segregation (ref. Weak seg)				
Democratic		-.917		-1.916
Homogeneous		-2.023		-5.935*
Strong seg.		-5.028		-4.152
Geographic area (ref. North-West)				
North-East		2.699*		4.708**
Centre		-5.376***		-1.918
South		-12.139***		-9.787***
Islands		-12.982***		-12.869***
ICC School	10.35%	2.04%	6.71%	3.19%
ICC Class	21.75%	14.48%	35.81%	36.93%
N° individuals	40,925	40,925	40,925	40,925
N° classes	3,306	3,306	3,306	3,306
N° schools	671	671	671	671

Note.-***p-value<.001, **p-value<.01, *p-value<.05.

TABLE 6.2: ML estimates of model in equation 6.1.

Relative frequences	Liceo	Tecnico	Professionale
Sample	58.64	27.42	13.94
Social class			
Upper	79.25	16.46	4.29
Middle	59.53	29.16	11.32
Lower	34.10	40.22	25.68
Excluded	31.26	33.49	35.25
Immigrant background			
Italians	62.72	25.28	12.00
Non-Italians	36.96	41.25	21.79
Geographic area			
North-west	57.46	30.41	12.13
North-East	56.88	31.47	11.65
Centre	65.61	24.28	10.11
South	56.46	25.97	17.57
Islands	56.39	23.73	19.88

TABLE 6.3: Distribution of individuals in upper secondary track according to selected characteristics.

6.3.2 Effects of segregation on school transitions

Thanks to the collaboration of INVALSI it was possible to link information on preferences for upper secondary education to the existing database. These preferences correspond to pre-enrolment, expressed in the period january-february of the year in which pupils are taking the final examination. The distribution of preferences in our sample is reported in Table 6.3 and confirms already known patterns. The classical ranking among social classes is present. Pupils from the upper class select more frequently a *Liceo* rather than other types of education, while moving towards lower social strata increases the percentages of enrolment in technical and vocational education. Non-Italian pupils are largely under-represented in the more academic track of upper secondary education, about half than Italians, while they are over represented in technical and vocational education. Finally, geographic sorting of preferences are characterized by small territorial differences, with the academic track always leading enrolment preferences but with technical education more preferred in the north and vocational education in the south and islands.

The analysis of upper secondary preferences is made by means of multinomial logistic regressions, that allow to apply logistic function to a categorical non ordered outcome. As already discussed, the dependent variable is articulated in three categories. The model is fitted by using *Liceo* as the reference or "base" category, used for baseline comparison group. In order to account for school effects, errors are clustered at the school level. The model is run on 41,266 individuals and has a Pseudo R^2 of .216.

The ratio of the prbability of choosing one outcome category over the probability of choosing the baseline is referred as relative risk. Relative risk ratios represent

the relative risk of an outcome for a unit change in the predictor variable. Table 6.4 presents the relative risk ratios for preferring *Professionale* or *Tecnico* to the academic track. Individual background variables are statistically significant and strongly correlated with the outcome. Individual performance has a negative, significant and substantial effect on probabilities of choosing tracks other than the reference category. Lower classes have higher ratios of choosing vocational or technical education instead of the more academic track. Also parental education plays a role in shaping preferences. The offspring of higher educated parents prefers the academic track rather than other options. Results from background variables are in line with previous research on enrolment and somehow corroborate hypothesis from risk aversion models and effectively maintained inequality theory in which upper classes are more prone to invest in education in order to limit the risk of downward mobility. Coming to ethnic background, the effect is significanti only for technical education. This suggests that the vocational track is equally populated by Italians and ethnic minorities, although variation should be present at the regional level. Moreover, the small effect and little statistically significance of ethnic backgrund suggests that the socio-economic position is more relevant in determining preferences for educational transitions rather than belonging to any ethnic group.

When coming to class and school composition variables, only the effect of the mean class ESCS is significant. The relative risk ratios are below one, thus an increase of one standard deviation in class ESCS is pushing preferences towards the baseline outcome. School segregation is not significant for both outcomes. This variable is capturing the variation in the distribution of pupils between and within classes, suggesting that repartition criteria of pupils are not directly influencing educational preferences. Indeed, the effect of such a variable should be partially absorbed by class composition variables.

Table 6.5 reports marginal effects for the compositional variable capturing class mean ESCS by social class and ethnic background. The effect of average socio-economic and cultural status composition of the class is negatively affecting the probabilities to choose vocational or technical education. Being inserted in a high ESCS class affect the most lower class pupils and those whose parents are simultaneously excluded from the labour market, rather than middle and upper class. Although differences are reducing while passing from probabilities of an outcome to another, the ranking persists. When looking at the effect of class background on Italians and non-Italians the situation is slightly different. The effect is smaller for non-Italians on preferences to enroll in *Professionale*, however the order changes for the other two outcomes. Thus, non-Italians are benefitting more than Italians from average class ESCS. The effect of average class ESCS on predicted probabilities by class is presented in the following Figure 6.4. Here, the blue line is the upper class, the red line the middle class, the green line the lower class and the yellow one the excluded. On the y axis of each graph there are the predicted probabilities of choosing that specific track plotted against the average ESCS of the classroom. The graph shows the distribution of the effect along

Variables	Professionale	Tecnico
Math score	.973***	.989***
Female	.410***	.263***
Non-Italian	1.152	1.383***
Social class (ref. Upper class)		
Middle	1.379***	1.312***
Lower	2.344***	1.672***
Excluded	2.610***	1.675***
Parental education (ref. Tertiary ed.)		
Primary	13.029***	3.837***
Lower Secondary	6.489***	3.060***
Upper secondary	2.195***	
N. of books at home (ref. 0-10)		
11-25	.618***	.739***
26-100	.451***	.634***
101-200	.311***	.524***
more than 200	.278***	.380***
Mean class ESCS	.510***	.571***
% non-Italians	2.564	2.223
% females	.608	2.035*
School segregation (ref. Weak seg)		
Democratic	1.090	1.134
Homogeneous	.790	.888
Strong seg.	619	.709*
Private education	.770	.565***
Geographic area (ref. North-West)		
North-East	1.058	1.145
Centre	.626***	.617***
South	.342***	.368***
Islands	.456***	.356***
N	41,266	
Pseudo R^2	.2167	

Note.-***p-value<.001, **p-value<.01, *p-value<.05.

TABLE 6.4: Multinomial logit for the effect of individual characteristics and class composition on school transition.

Variables	Professionale	Tecnico	Liceo
Upper class	-.027	-.067	.094
Middle	-.029	-.069	.099
Lower	-.035	-.066	.102
Excluded	-.038	-.064	.103
Italians	-.030	-.065	.095
Non-Italians	-.027	-.070	.098

TABLE 6.5: Marginal effects of class ESCS on probabilities of get any outcome.

different values of average ESCS. Increasing values of the average ESCS affect similarly all classes. This result suggests also that family strategies aiming at enrolling the offspring in classes with high ESCS are paying some dividends in the long run (educational transitions). However the observed effects might be spourious and additional investigation is needed.

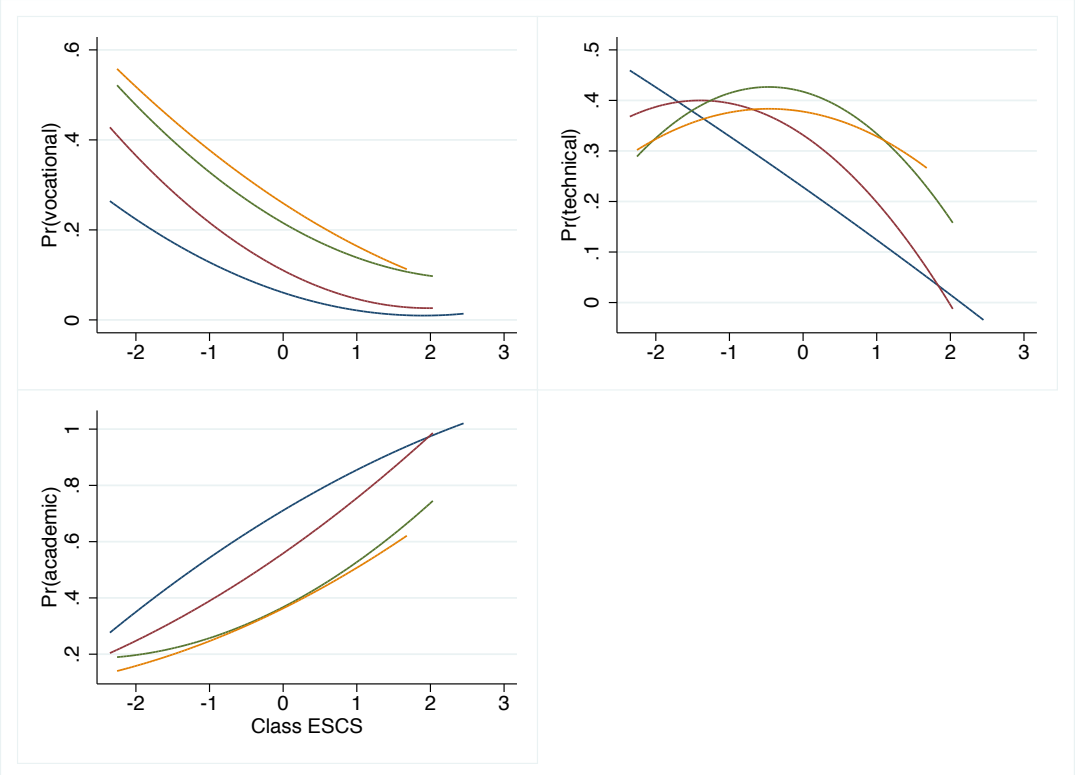

FIGURE 6.4: Average school score in math by school composition.

6.4 Conclusions

This chapter has been devoted to the investigation of effects of segregation on test score performance and educational transitions. In order to answer the research questions I recurred to multilevel models and multinomial logit regressions. It contributes to the existing literature by replicating existing analysis on individual performances

and adding new evidence on effect of class composition on preferences for upper secondary educational transitions.

The effect of class composition is found to be statistically significant on both individual performances and preferences for upper secondary education. The results on individual performances are in line with existing literature, although some differences are present. These might depend on differences in the specification of the model, especially in choice of the dependent variable. Here, I'm using the individual math and Italian score from INVALSI, while Contini (2013) is using the percentage of correct answers in the test. Somehow counter-intuitive results are found for the multilevel model on Italian score. Here, the average ESCS of the classroom is having a negative effect on individual performance. Since the model is run only on public schools, the result is difficult to interpret.

When coming to preferences for educational transitions, background characteristic are significantly relevant. Pupils from families from higher socio-economic and cultural status are more likely to enroll in the academic upper secondary track rather than any other one. Here, the effect of class composition is significant and affecting lower socio-economic groups and non-Italians the most. Although a counter-factual is not present, since class composition is largely the result of movement of pupils between schools it is also affecting individual performances and transition preferences.

The investigations presented in this chapter still require furhter inquiry on methodological ground. At the moment it is not possible to clearly highlight the contribution of segregation to individual performances, attitudes or behaviours, nor is possible to state how much of the present class composition is the result of segregation. Moreover, class composition true effect in the multinomial regressions is likely to intervene both as direct and indirect effect. In this last case, by the mediation of individual performances. Moreover, the multinomial logit regression is not accounting for the complete hierarchical structure of the data. The adopted solution has been that of clustering errors at the school level, another possible solution might be to recurr to structural equation models.

Chapter 7

Conclusions

In this dissertation I've attempted to contribute to the discussion about school segregation in Italian lower secondary, its causes and effects. This study aims at contributing to the little existing literature on Italy in several ways. First, it tries to address the issue of school segregation with a national perspective, not present in the sociological debate on Italy until now. In fact, several contributions have focused on selected cases, usually single municipalities as in the case of Pacchi and Ranci (2017), and providing basis for a first understanding of the phenomenon. Second, a theoretical interest. The phenomenon of segregation by ethnic and social background in educational systems characterized by a tracking structure is well established in the literature. What is less known is how segregation works in systems with no tracking, like Italian lower secondary school. I argued that the absence of tracking in presence of an institutional context that allows for diversification of the educational supply results in hidden tracking on school quality basis, and thus sorting of pupils among schools according to their family background. Such an arrangement of the institutional context usually transforms good schools in a scarce resource, thus favouring competition for the educational provision among families. The theoretical discussion has been devoted to investigate these mechanisms in determining school segregation. In lights of the results, we should reckon that segregation is just one of the strategies that families play when enrolling their offspring to a school. Probably, segregation is only the aggregate result of several individual behaviours whose aim is not producing segregation, but granting better education to their sons and daughters. Third, the interest in segregation was due to investigation of its effects in terms of school performance and educational transitions.

On a theorethical ground, the dissertation aims at linking segregation to existing approaches to inequalities, connecting the phenomenon to social closure dynamics. The discussion starts from reckoning that available contributions usually see school segregation as a natural phenomenon and address "white flight" as the main "natural" cause. However, they do not consider enough the role of the schooling system, how it is perceived by ethnic and social groups, its structure and organization. According to Collins (2000), schooling in modern capitalist societies is characterized by a bureaucratic and hierarchical structure, where educational institutions grant formal degrees as certification of possession of some knowledge. The possession of the

highest degrees allows economic returns in the labour market (*via* premium wages or access to professions), but represents also a *status*, the key to access the highest positions in the social structure. Given the advantages deriving from the possession of educational titles, along with the mass structure of schooling systems of modern capitalist societies and the scarcity in the supply of high-level degrees, families and pupils compete to gain access to educational paths and institutions that grant this asset. Social closure dynamics between groups may emerge as a main consequence of competition, thus determining segregation of groups between educational paths, schools, classes. Social closure dynamics are discussed with reference to structuralist approaches and rational action theory, in particular to relative risk aversion models.

The dissertation has been devoted to investigate four research questions:

- Which is the level of residential, school district and school segregation in Italy? Which are the determinants of segregation?
- How is district composition reflected in school composition? What makes families more prone to enroll their offspring in a school different from the closest public school?
- Is there a significant link between the composition of schools and classes on the performance in achievement tests? Is it heterogeneous over pupils?
- Are those coming from segregated schools more inclined to make any school transition? Is it heterogeneous over pupils?

These questions contributes to the existing literature by adding evidence on the Italian case, combining information from two databases already existing databases that have been enriched with spatial information. In particular, the dissertation contributes by opening possibilities on research on school segregation through the combination of administrative data of pupils. This combination allows for the exploitation of data at different levels. At the aggregate level, for the investigation of trends of segregation, flows of pupils between schools and their repartition in classes. At the individual level, determinants of chosing an alternative school rather than the natural one, effects of these preferences on individual performances and school-type choices. Although extensively used, the information contained in these data can be additionally "mined" by recurring to spatial thecniques, in order to give more value to residence information.

The database used is the result of a linking process between two sources. The first is the *2011 Population Census* provided by ISTAT, the latest census wave available. Census data are collected every ten years at the household level but provided in aggregate form. The elementary geographical units for both collection and output are the *sezioni di censimento*, the smaller territorial partition at which data are aggregated. Census data include information on residential population, its distribution among different classes of age, education, occupation and nationality, along with information on buildings' condition. The second source of data is the *2015 Survey on*

Learning - INVALSI. INVALSI is the governmental agency for the evaluation and monitoring of the educational system, each year it runs a survey aimed at investigating the skills and knowledge of pupils. The survey has been introduced with the law 25/10, 2007, n. 176, and administered for the first time in the 2007-2008 along with the final examination of grade 5.

School segregation is analyzed by means of segregation indices of eveness of distribution (Dissimilarity index, Information theory index) and exposure (Exposure index). Moreover, mobility tables have been produced in order to understand the flows of pupils between schools. The determinants of probabilities of school choice have been investigated through logistic regression, while multilevel models have been used for in order to explorate the effect of classroom composition on individual performances. Finally, multinomial regression has been used to investigate probabilities of educational transitions.

The main results highlight Italian lower secondary schools as characterized by medium levels of segregation, similar to those of residential segregation, although the two aggregates are not comparable. However, the distribution of schools according to levels of segregation highlights selected cases characterized by high internal homogeneity or strong segregation. If we consider that classes are the group of interest, it is possible to compute variability due to differences between classes of the same school and variability within each class. These two measures allow to identify the degree of homogeneity of each class, but also the degree of homogeneity between classes of the same school. Each school can then be plotted according the *between* and *within* component, thus defining the distribution of schools. Tresholds are drawn in correspondence of average variance values dividing the plane in four quadrants. The first quadrant (upper right) is characterized by schools with within and between classes variation above average. It means that classes of these schools are heterogenous in their composition, but there are also differences between classes. Thus, we can think of these schools as those that are experiencing mild levels of segregation. The second quadrant (lower right) is made by schools with high variation within classes, but small differences are present in composition between classes. This group is associated to a more "democratic" partition of students. In a school belonging to this quadrant all classes are quiet heterogeneous but there is small variation between them. The third quadrant (lower left) is build by schools with homogeneous classes. These schools are experiencing a particular case of segregation. They are not segregating between their classes, but are themselves a segregated environment with respect to other school. Finally, the fourth quadrant (upper left) is made by strongly segregated schools. This group is characterized by classes that are homogeneous in their composition, but there is variation between classes of the same school. This is the most evident case in which segregation configures itself as the phenomenon of social distance expressed in physical separation of space. Finally, private education is the most evident case of segregation in Italian lower secondary school. Private education is characterized by high levels of segregation of the upper class and high

homogeneity of schools and classes.

The observed levels of segregation are the consequence of a well established phenomenon of redistribution of pupils among schools. In order to investigate the phenomenon, I distinguish between natural and alternative schools. For each pupil, the natural school is the one minimizing the home-school distance, while all others up to a maximum of ten are alternative schools. On a descriptive basis, those enrolling in alternative schools are espetially those from high status families, while the are no substantial differences between Italians and non-Italians. Going more in detail, the quality of this phenomenon of redistribution has been investigated by applying the previously discussed categorization of schools. Mobility tables that simultaneously account for the tipe of natural and alternative school have been computed. Results show that the 51% of pupils whose natural school is in any category (weak segregation, democratic, strong segregation, homogeneous schools) are moving towards schools with weak levels of segregation. The figure is higher for those coming from democratic schools (55%).

When looking at the composition of these flows by average ESCS, three interesting facts emerge. The average ESCS of those going to private education is very high, close to one standard deviation.[1] Weakly segregated schools are receiving, on average, students with higher status from other weakly segregated schools, and medium status pupils from other kinds of institutions. The slightly negative scores for democratic schools tell that these are especially the destination of the middle class. Surprisingly, average ESCS for those coming from homogeneous schools is substantially lower. Homogeneous schools intercept especially middle and lower classes. The former group is prevalently moving from weakly segregated and democratic schools, while the latter from homogeneous and strongly segregted institutions. While homogeneous schools are receiving groups with the lower ESCS from other homogeneous and strongly segregated schools, strongly segregated schools are receiving groups with the lower ESCS from weakly segregated and democratic schools. About movement for ethnic minoritis, the most noticeable fact is the extremely low share of immigrants going to private education. Apart from this result, a clear pattern of movement of foreigners among schools is difficult to be highlighted.

Movement has been also investigated by looking at how the school ESCS changes. The population of each school has been divided in two aggregates: the base quota, the "premium" quota. The first aggregate is made by those enrolling in their natural school, the second by those entering in a school that is an alternative option to them. When looking at how the composition of schools changes in this analysis, we observe a strong flattening in the distribution of schools. In practice, on average each school is receiving students that are more similar to those in the base quota. The logistic regression for probabilities of enrollment in alternative schools has confirmed the results of previous analysis adding some evidence that taking the choice of enrolling

1. ESCS is standardized with mean [0] and standard deviation [1]

in alternative school is also the result of how the natural school is partitioning the student body between classes.

The redistribution of pupils is seen as the result of competition in a system characterized by free movement of individuals and no formal strict limitations (no districts). While analysis have indicated that the movement of upper class pupils is oriented by strategic behaviour, there are no possibilities to clearly state whether lower class individuals are just acting in the same way or suffering the consequences of upper class redistribution. This incapacity is also rooted in data limitations, an element that will be discussed later on. Moreover, the social closure dynamic is only partially proved for the upper class. It is certainly proved for those of the dominant class moving to private education. Here, the incumbents of upper class are deliberately separating from other classes, creating homogenoeus educational environments in order to take advantage from socialization and social and cultural capital accumulation. However, if the social closure dynamic would be persistent also in the public school, we should observe strong polarization and enrolment of upper class pupils in strongly segregated schools. Although happening, this phenomenon is not characterized by the expected strenght.

The effects of segregation are difficult to capture, especially when there is no possibility to build a counterfactual. Instead, I've decided to investigate the effects of classrooms composition as these are the privileged place in which pupils interact and shape their behaviours and attitudes. Moreover, if a segregative dynamic is at work, the final result is captured by class composition. The effects of individual characteristics, family background and compostional variables on individual performances and preferences for upper secondary education has been investigated. Here, apart for the classical effect of individual and family background characteristics, the interesting effect is that of class socio-economic composition on both outcomes, confirming that peer characteristics are actually playing a role in shaping individual performances and transitions to upper secondary. However, it is not possible to say how much of this effect is the result of segregation, or at least of its main determinant: redistribution of pupils. Further analysis are required, and more methodological work is needed in order to disentangle the proper effect.

As shortly highlighted, this contribution suffers from some limitations involving data and methods. Concerning school segregation, the shortage of complete and accessible administrative data on school characteristics is a major limitation in testing hypothesis of school quality as determinant in the choice of being a school mover. This makes the opportunity costs of doing research on this theme on the Italian case quite high. In fact, at present the time and efforts that must be devoted to the cause of data obtainment, and the additional time and efforts for data cleaning, are hardly comparable to those related to research on the same topic in other coutries. On the methodological ground, limitations are related to the quality of available variables and to choices of the author.

The variable accounting for the ethnic background of pupils in the INVALSI database

is collected using an arbitrary aggregate form, that does not allow to identify immigrants from developed and underdeveloped countries. Moreover, also the classification of occupations might be improved by some form of disaggregation. However, as far as census and INVALSI data are analized together with the aim of comparing results, the chosen scheme is the only available. Alternatively, one might refers only to INVALSI data for each school, by looking at the complete population of that school and its distribution in the urban space. In this way, the socio-economic classification proposed by INVALSI might be sufficient for the analysis. About methodological choices, the main limitation is represented by the sample selection. The sample can be expanded in two directions. The first is to include all the municipalities from the selected Local Labour Systems, and analyze flows of pupils inside these aggregates. The second is to select pupils from all the municipalities, thus working with the complete population. This poses the problem of comparability between local school systems of different size. A possible solution might be that of defining population tresholds for making the phenomenon more comparable.

This dissertation open possible research lines in different directions for future work. The first regards updating available information in the database and improving its quality. At the moment, the main weakness is represented by the absence of reliable and complete information on school characteristics. Although existing, administrative information on schools' characteristics is usually difficult to access as no public repository is available at the moment and requests should pass through ministerial offices. Moreover, reliable information on characteristics of the teaching staff is not available at the moment. Improvement of these two components might substantially help in increasing reliability of information on school quality. Moreover, information on pupil side can be improved for analysis on educational transitions to upper secondary. These are now completely observed for the cohort of interest, and analysis should benefit from a more complete universe of observations. The analysis on flows between natural and alternative schools might benefit on additional information regarding parental place of job. At the moment, it is not possible to clear the analysis from white noise produced by pupils moving between school only because they are following their parents. Unfortunately, there is no available information of this kind for the present sample, while it must be constructed *ad hoc* for future research.

The second regards investigation of trends of segregation over time. To the author's knowledge, no investigation of segregation over time is present for Italy. It would be a valuable tool in order to understand evolution of preferences concerning schools, to evaluate the impact of different policies and to investigate the structure of school segregation. In fact, the whole dissertation relies on the (probably strong) assumption that the situation depicted is constant each year. Investigation of segregation over time might also rely on looking at transitions between primary and lower secondary education, always by means of INVALSI data, and investigating the path of pupils according to characteristics of peers.

The third element regards the solution of the district definition problem. Alternatively to choices make in this dissertation, school districts might be defined as mutually exclusive environments covering the entirety of tht urban space. Here, Voronoi polygons might be useful but they still represents a tentative solution. Alternatively, data on school districts might be gathered from municipal administration, although this represents a highly costly option in time and efforts.

The final stream of possible future research involves the qualitative analysis of school strategies, if any, in shaping possibilities of enrolment, the analysis of the role of information shared by parents and how it is conveyed in friendship networks, the perception of school quality. Although difficult to investigate, these aspects seems to be the more promising in widening knowledge on segregation, as they touch mechanisms that are at the base of the enrolment process.

Bibliography

Abascal, M., and D. Baldassarri. 2015. "Love Thy Neighbor? Ethnoracial Diversity and Trust Reexamined." *American Journal of Sociology* 121 (3): 722–782.

Agasisti, T., C. Masci, F. Ieva, and A.M. Paganoni. 2016. "Does class matter more than school? Evidence from a multilevel statistical analysis on Italian junior secondary school students." *Socio-economic Planning Science* 54:47–57.

Ammermueller, A., and J. Pischke. 2009. "Peer effects in European primary schools: Evidence from the progress in international reading literacy study." *Journal of Labor Economics* 27 (3): 315–348.

Andersson, E., J. Östh, and B. Malmberg. 2010. "Ethnic segregation and performance inequality in the Swedish school system: A regional perspective." *Environment and Planning A* 42 (11): 2674–2686.

Arbaci, S. 2007. "Ethnic Segregation, Housing Systems and Welfare Regimes in Europe." *International Journal of Housing Policy* 7 (4): 401–433.

Arbaci, S., and J. Malheiros. 2010. "De-Segregation, Peripheralisation and the Social Exclusio of Immigrants: Southern European Cities in the 1990s." *Journal of Ethnic and Migration Studies* 36 (2): 227–255.

Archbald, D.A. 2004. "School Choice, Magnet Schools, and the Liberation Model: An Empirical Study." *Sociology of Education* 77:283–310.

Argentin, G., B. Romano, and A. Martini. 2012. "Giocare a scacchi aiuta a imparare la matematica? Evidenza da una sperimentazione controllata." In *Gli scacchi, un gioco per crescere. Sei anni di sperimentazione nella scuola primaria.* Franco Angeli.

Baird, K. 2012. *Trapped in Mediocrity: Why Our Schools Aren't World-Class and What We Can Do About It.* Rowman / Littlefield Publishers.

Ballarino, G., F. Bernardi, M. Requena, and H. Schadee. 2009. "Persistent Inequalities? Expansion of Education and Class Inequality in Italy and Spain." *European Sociological Review* 25 (1): 123–138.

Ballarino, G., N. Panichella, and M. Triventi. 2014. "School expansion and uneven modernization. Comparing educational inequality in Northern and Southern Italy." *Research in Social Stratification and Mobility* 36:69–86.

Ballarino, G., and H. Schadee. 2008. "La disuguaglianza delle opportunità educative in Italia, 1930-1980: Tendenze e cause." *Polis* 22 (3): 373–402.

Bankston III, C., and S.J. Caldas. 1996. "Majority African American Schools and Social Injustice: The Influence of De Facto Segregation on Academic Achievement." *Social Forces* 75 (2): 535–555.

Barbagli, M., and M. Pisati. 2012. *Dentro e fuori le mura. Cittá e gruppi sociali dal 1400 ad oggi.* Il Mulino.

Barberis, E., and A. Violante. 2013. "La frammentazione in azione: la segregazione scolastica in quattro aree metropolitane." In *Il welfare frammentato. Le articolazioni regionali delle politiche sociali italiane,* edited by E. Barberis and Y. Kazepov, 182–204. Carocci Editore.

Barzel, Y. 1997. *Economic Analysis of Property Rights, 2d ed.* New York: Cambridge University Press.

Becker, G.S. 1957. *The Economics of Discrimination.* Chicago: University of Chicago Press.

Bell, W. 1954. "A probability model for the measurement of ecological segregation." *Social Forces* 32 (4): 357–364.

Blau, P.M., and O.D. Duncan. 1967. *The American Occupational Structure.* John Wiley.

Bråmå, Å. 2008. "Dynamics of Ethnic Residential Segregation in Goteborg, Sweden, 1995-2000." *Population, Space and Place* 14:101–107.

Breen, R., and J.H. Goldthorpe. 1997. "Explaining Educational Differentials: Towards a Formal Rational Action Theory." *Rationality and Society* 9:275–305.

Brunello, G., and L. Rocco. 2013. "The effect of immigration on the school performance of natives: Cross country evidence using PISA test scores." *Economics of Education Review* 32:234–246.

Burgess, S., D. Wilson, and R. Lupton. 2005. "Parallel lives? Ethnic segregation in schools and neighbourhoods." *Urban Studies* 42 (7): 1027–1056.

Card, D., and J. Rothstein. 2007. "Racial Segregation and the Black-White Test Score Gap." *Journal of Public Economics* 91 (11): 2158–2184.

Cardone, M., P. Falzetti, and A. Severoni. 2015. "Equità o segregazione scolastica? L'effetto della composizione delle classi sull'apprendimento degli studenti." *INVALSI Working Paper N.25/2015.*

Casper, M. 2001. "A definition of "social environment"." *American Journal of Public Health* 91:465.

Cebolla-Boado, H. 2007. "Immigrant Concentration in School: Pee Pressures in Place?" *European Sociological Review* 23 (3): 341–356.

Cebolla-Boado, H., and L.G. Medina. 2011. "The Impact of Immigrant Concentration in Spanish Schools: School, Class, and Composition Effects." *European Sociological Review* 27 (5): 606–623.

Charles, C.Z. 2003. "The Dynamics of Racial Residential Segregation." *Annual Review of Sociology* 29 (1): 167–207.

Checchi, D., and T. Jappelli. 2003. "School Choice and Quality." *IZA Discussion paper series,* no. 828.

Clark, W.A.V. 1991. "Residential Preferences and Neighborhood Racial Segregation: A Test of the Schelling Segregation Model." *Demography* 28 (1): 1–19.

Clotfelter, C.T. 1998. "Public School Segregation in Metropolitan Areas." *NBER Working Paper Series.*

Cobalti, A. 1990. "Schooling inequalities in Italy: Trends over time." *European Sociological Review* 6 (3): 199–214.

Cobalti, A., and A. Schizerotto. 1993. "Inequality of educational opportunity in Italy." In *Persistent inequality: Changing educational attainment in thirteen countries,* edited by Y. Shavit and H.P. Blossfeld.

Coleman, J.S., E.Q. Campbell, C.F. Hobson, J. McPartland, and A.M. Mood. 1966. *Equality of Educational Opportunity.* Technical report. U.S. Office of Education.

Coleman, J.S., S.D. Kelly, and J.A. Moore. 1975. *Trends in School Segregation, 1968-73.* Technical report. The Urban Institute.

Collins COBUILD - English Dictionary. 2003. HarperCollins Publishers.

Collins, R. 1979. *The Credential Society: An Historical Sociology of Education and Stratification.* Academic Press.

———. 2000. "Comparative and Historical Patterns of Education." In *Handbook of the Sociology of Education,* edited by M.T. Hallinan. Springer.

Contini, D. 2013. "Immigrant background peer effects in Italian schools." *Social Science Research* 42 (4): 1122–1142.

Cortese, C.F., F.R. Falk, and J.K. Cohen. 1976. "Further Considerations on the Methodological Analysis of Segregation Indices." *American Sociological Review* 41 (4): 630–637.

Cousin, B., and E. Préteceille. 2008. "La division sociale de l'espace milanais." *Mélanges de l'école française de Rome* 120 (1): 315–333.

Cowgill, D.O., and M.S. Cowgill. 1951. "An Index of Segregation Based on Block Statistics." *American Sociological Review* 16 (6): 825–831.

Crain, R.L., and R.E. Mahard. 1978. "Desegregation and black achievement: A review of the research." *Law and Contemporary Problems* 42 (3): 17–56.

Dark, S., and D. Bram. 2007. "The modifiable areal unit problem (MAUP) in physical geography." *Progress in Physical Geography* 31 (5): 471–479.

Datcher, L. 1982. "Effects of community and family background on achievement." *Review of Economics and Statistics* 64:32–41.

DiPrete, T.A., and G.M. Eirich. 2006. "Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments." *Annual Review of Sociology* 32:271–297.

Dujardin, C., H. Seldon, and I. Thomas. 2008. "Residential Segregation and Unemployment: The Case of Brussels." *Urban Studies* 45 (1): 89–113.

Dumay, X., and V. Dupriez. 2014. "Educational quasi-markets, school effectiveness and social inequalities." *Journal of Education Policy* 29 (4): 510–531.

Duncan, G., J. Brooks-Gunn, and P. Klebanov. 1994. "Economic Deprivation and Early Childhood Development." *Child Development* 65 (2): 296–318.

Duncan, O.D., and B. Duncan. 1955. "A Methodological Analysis of Segregation Indexes." *American Sociological Review* 20 (2): 210–217.

Duru Bellat, M., and A. Mingat. 1997. "La constitution de classes de niveau dans les colléges: les effets perverse d'une pratique à visée égalisatrice." *Revue Française de Sociologie:* 759–789.

Ellen, I.G., and M.A. Turner. 1997. "Does Neighborhood Matter? Assessing Recent Evidence." *Housing Policy Debate* 8 (4): 833–866.

Elster, J. 1978. "Exploring Exploitation." *Journal of Peace Research* 15 (1): 3–17.

Emerson, M.O., K.J. Chai, and G. Yancey. 2001. "Does Race Matter in Residential Segregation? Exploring the Preferences of White Americans." *American Sociological Review* 66 (6): 922–935.

Engberg, M.E., and G.C. Wolniak. 2010. "Examining the effects of high school contexts on postsecondary enrollment." *Research in Higher Education* 51:132–153.

Erikson, R. 1984. "Social class of men, women and families." *Sociology* 18 (4): 500–514.

Fack, G., and J. Genet. 2016. "Mixité sociale et scolaire dans le lycées parisiens. Les enseignements de la procédure Affelnet." In *Massification scolaire et mixité sociale,* edited by de l'Enseignement supérieur et de la Recherche Ministère de l'Education nationale, 77–100.

Farley, R., S. Danziger, and H.J. Holzer. 2000. *Detroit divided.* Russel Sage Foundation.

Farley, R., and W.H. Frey. 1994. "Changes in the segregation of whites from blacks during the 1980s: Small steps toward a more integrated society." *American Sociological Review* 59 (1): 23–45.

Farley, R., and A.F. Tauber. 1974. "Racial segregation in the public schools." *American Journal of Sociology* 79:888–905.

Farley, R., and K.E. Tauber. 1968. "Population trends and residential segregation since 1960." *Science* 159 (3818): 953–956.

Felouzis, G. 2003. "La ségrégation ethnique au collèges et ses conséquences." *Revue Française de Sociologie* 44 (3): 413–447.

Fiel, J. 2015. "Closing Ranks: Closure, Status Competition, and School Segregation." *American Journal of Sociology* 121 (1): 126–170.

Fielding, T.J. 1989. "Migration and Urbanization in Western Europe Since 1950." *The Geographical Journal* 155 (1): 60–69.

Fix, M., and R.J. Struyk. 1993. *Clear and convincing evidence: Measurement of Discrimination in America.* Urban Institute.

Frankenberg, E. 2013. "The Role of Residential Segregation in Contemporary School Segregation." *Education and Urban Society* 45 (5): 548–570.

Frankenberg, E., and C.M. Lee. 2002. *Race in American public schools: rapidly resegregating school districts.* Technical report. Harvard Civil Rights Project, Cambridge, MA.

Galster, G.C. 1988. "Residential segregation in American cities: A contrary review." *Population Research and Policy Review* 7 (2): 93–112.

———. 1990. "Racial steering by real estate agents: Mechanisms and motives." *The Review of Black Political Economy* 19 (1).

———. 2012. "The mechanism(s) of neighbourhood effects: Theory, evidence and policy implications." In *Neighbourhood effects research: New perspectives*, 23–56. Springer.

Ganzeboom, H. 2010. "A new International Socio Economic Index (ISEI) of occupational status for the International Standard Classification of Occupations 2008 (ISCO-08) constructed with data from the ISSP 2002-2007." In *Annual Conference of International Social Survey Programme*, vol. 1.

Givord, P., M. Guillerm, O. Monso, and F. Murat. 2016. "La ségrégation scolaire entre le collèges. Quelles différences entre public et privé, aux niveaux national, académique et local?" In *Massification scolaire et mixité sociale*, edited by de l'Enseignement supérieur et de la Recherche Ministère de l'Education nationale, 53–76.

Gnaldi, M., and S. Bacci. 2016. "Joint assessment of the latent trait dimensionality and observed differential item functioning of students' national tests." *Quality & Quantity* 50 (4): 1429–1447.

Gotham, K.F. 2000. "Urban Space, Restrictive Covenants and the Origins of Racial Residential Segregation in a US City, 1900-50." *International Journal of Urban and Regional Research* 24 (3).

Gramberg, P. 1998. "School segregation: the case of Amsterdam." *Urban Studies* 35 (5): 547–564.

Grusky, D., and A. Sørensen. 1998. "Can Class Analysis Be Salvaged?" *American Journal of Sociology* 103:1187–1234.

Guest, A.M., and J.A. Weed. 1976. "Ethnic Residential Segregation: Patterns of Change." *American Journal of Sociology* 81 (5): 1088–1111.

Gustafsson, J.E. 1980. "Testing and obtaining fit of data to the Rasch model." *British Journal of Mathematical and Statistical Psychology* 33:205–233.

Hanushek, E.A. 1972. *Education and Race: An Analysis of the Educational Production Process.* D.C. Heath.

Hanushek, E.A., J.F. Fain, J.M. Markman, and S.G. Rivkin. 2003. "Does Peer Ability Affect Student Achievement?" *Journal of Applied Econometrics* 18 (5): 527–544.

Hanushek, E.A., J.F. Kain, and S.G. Rivkin. 2009. "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." *Journal of Labor Economics* 27 (3): 349–383.

Harrison, E., and D. Rose. 2006. "The European Socio-economic Classification." *European Societies* 9 (3): 459–490.

Hoxby, C. 2000. *Peer effects in the classroom: Learning from gender and race variation.* Technical report. National Bureau of Economic Research.

Hoxby, C.M., and G. Weingarth. 2005. *Taking race out of the equation: School reassignment and the structure of peer effects.* Mimeograph.

Huber, S., and C. Rust. 2016. *OSRMTIME: Calculate Travel Time and Distance with OpenStreetMap Data Using the Open Source Routing Machine (OSRM).*

Iceland, J. 2004. *The Multigroup Entropy Index (Also Known As Theil's H or the Information Theory Index.* Technical report. U.S. Census Bureau.

Iceland, J., and M. Scopilliti. 2008. "Immigrant Residential Segregation in the U.S. Metropolitan Areas, 1990-2000." *Demography* 45 (1): 79–94.

INVALSI. 2016. *Rilevazione nazionale degli apprendimenti 2015-2016. Rapporto tecnico.* Technical report. INVALSI.

ISTAT. 1991. *I sistemi locali del lavoro 1991.* Technical report. ISTAT.

Jahn, J., C.F. Schmid, and C. Schrag. 1947. "The Measurement of Ecological Segregation." *American Sociological Review* 12 (3): 293–303.

James, D.R., and K.E. Tauber. 1985. "Measures of Segregation." In *Sociological Methodology 1985,* edited by N. Tuma, 1–32. Jossey-Bass.

Jenkins, S.P., J. Micklewright, and S.V. Schnepf. 2008. "Social Segregation in Secondary Schools: How Does England Compare with Other Countries." *Oxford Review of Education* 34 (1): 21–37.

Johnston, R., D. Wilson, and S. Burgess. 2004. "School segregation in multiethnic England." *Ethnicities* 4 (2): 237–265.

Kahlenberg, R.D. 2001. *All together now: Creating middle-class schools through public school choice.* Brookings Institution.

Karsten, S., G. Ledoux, J. Roeleveld, C. Felix, and D. Elshof. 2003. "School Choice and Ethnic Segregation." *Educational Policy* 17 (4): 452–477.

King, D. 1995. *Separate and unequal: black Americans and the U.S. federal government.* Clarendon Press.

Krinsten, C. 2008. "Primary School Choice and Ethnic School Segregation in German Elementary Schools." *European Sociological Review* 24 (4): 495–510.

Kristen, C. 2008. "Primary School Choice and Ethnic School Segregation in German Elementary Schools." *European Sociological Review* 24 (4): 495–510.

Krysan, M., and R. Farley. 2002. "The Residential Preferences of Blacks: Do They Explain Persistent Segregation?" *Social Forces* 80 (3): 937–980.

Kuper, L. 1968. "Segregation." *International Encyclopedia of the Social Sciences* 14:144–150.

Kushner, James A. 1979. "Apartheid in America: An Historical and Legal Analysis of Contemporary Racial Residential Segregation in the United States." *Howard Law Journal* 22:547–686.

Lazear, E.P. 2001. "Educational production." *Quarterly Journal of Economics* 116:777–803.

Ledoux, G., P. Koopman, and L. Schaap. 1999. *Schoolkeuzen van ouders in het basisonderwijs [School choice in primary school].* SCO-Kohnstamm Instituut.

Lee, B., S.F. Reardon, G. Firebaugh, C.R. Farrell, S.A. Matthews, and D. O'Sullivan. 2008. "Beyond the Census Tract: Patterns and Determinants of Racial Segregation at Multiple Geographic Scales." *American Sociological Review* 73 (5): 766–791.

Lelo, K., S. Monni, and F. Tomassi. 2018. "Disuguaglianze metropolitane: un confronto con Milano e Napoli." In *Politiche urbane per Roma. Le sfide di una Capitale debole,* edited by E. d'Albergo and D. De Leo, 17–35. Sapienza Università Editrice.

Leventhal, T., and J. Brooks-Gunn. 2000. "The Neighborhoods They Live In." *Psychological Bulletin* 126 (2): 309–337.

Lever, W.F. 2001. "The Post-fordist City." In *Handbook of Urban Studies,* edited by R. Paddison. Sage Publications.

Lieberson, S. 1981. "An Asymmetrical Approach to Segregation." In *Ethnic Segregation in Cities,* edited by C. Peach, V. Robinson, and S. Smith, 61–82. Croom Helm.

Littman, M.S. 1998. *A Statistical Profile of the United States.* Bernan Pressk.

Logan, J.R., B.J. Stults, and R. Farley. 2004. "Segregation of minorities in the metropolis: Two decades of change." *Demography* 41 (1): 1–22.

Lucas, S.R. 2001. "Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects." *American Journal of Sociology* 106 (6): 1642–1690.

Ludwig, J., G. Duncan, L. Gennetian, Lawrence Katz, R. Kessler, J. Kling, and L. Sanbonmatsu. 2013. "Long-Term Neighborhood Effects on Low-Income Families: Evidence from Moving to Opportunity." *American Economic Review* 103 (3): 226–231.

Ludwig, J., G. Duncan, and H. Ladd. 2001. "The effect of MTO on Baltimore children's educational outcomes." *Poverty Research News* 5 (1): 13–15.

Lutters, W.G., and M.S. Ackerman. 1996. "An introduction to the Chicago School of Sociology." *Interval Research Property:* 2–6.

Malheiros, J. 2002. "Ethni-cities: Residential Patterns in the Northern European and Mediterranean Metropolises - Implications for Policy Design." *International Journal of Population Geography* 8 (2): 107–134.

Maloutas, T. 2007. "Segregation, Social Polarization and Immigration in Athens during the 1990s: Theoretical Expectations and Contextual Differences." *International Journal of Urban and Regional Research* 31 (4): 733–758.

Mansi, C.F. 1993. "Identification of endogenous social effects: The reflection problem." *Review of Economic Studies* 60:531–542.

Marx, K. 1996. "Capital, Volume I." In *Collected Works,* edited by K. Marx and F. Engels, vol. 35. London: Lawrence & Wishart.

Massey, D.S., and N. Denton. 1987. "Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 1970-1980." *American Sociological Review* 52 (6): 802–825.

———. 1988. "The Dimensions of Residential Segregation." *Social Forces* 67 (2): 281–315.

———. 1989. "Hypersegregation in U.S Metropolitan Areas: Black and Hispanic Segregation along Five Dimensions." *Demography* 26 (3): 373–391.

Massey, D.S., A.B. Gross, and M.L. Eggers. 1991. "Segregation, the Concentration of Poverty, and the Life Chances of Individuals." *Social Science Research* 20:397–420.

Massey, D.S., and J. Tannen. 2015. "A Research Note on Trends in Black Hypersegregation." *Demography* 52 (3): 1025–1034.

Mayer, S.E., and C. Jencks. 1989. "Growing up in Poor Neighborhoods: How Much Does it Matter?" *Science, New Series* 243 (4897): 1441–1445.

Mazza, A., G. Gabrielli, and S. Strozza. 2017. "Residential Segregation of Foreign Immigrants in Naples." *Spatial Demography:* 1–17.

McKenzie, R.D. 1924. "The Ecological Approach to the Study of Human Community." *American Journal of Sociology* 30 (3): 287–301.

McKeow, A. 2004. "Global Migration, 1846-1940." *Journal of World History* 15 (2): 155–189.

McNeal, R.B. 1997. "High school dropouts: A closer examination of school effects." *Social Science Quarterly* 78:209–222.

McPherson, M., L. Smith-Lovin, and J.M. Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual Review of Sociology* 27 (1): 415–444.

Merton, R.K. 1973. *The sociology of science: Theoretical and empirical investigations.* University of Chicago Press.

———. 1988. "The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property." *Isis* 79 (4): 606–623.

Morgan, B.S. 1975. "The Segregation of Socioeconomic Groups in Urban Areas: A Comparative Analysis." *Urban Studies* 12:47–60.

Murphy, R. 1984. "The Structure of Closure: A Critique and Development of the Theories of Weber, Collins, and Parkin." *The British Journal of Sociology* 35 (4): 547–567.

Musterd, S. 2005. "Social and ethnic segregation in europe levels, causes, and effects." *Journal of Urban Affairs* 27 (3): 331–348.

Nye, L.V., B. nad Hedges, and S. Konstantopoulos. 2000. "The Effects of Small Classe on Academic Achievement: The Results of the Tennessee Class Size Experiment." *American Educational Research Journal* 37 (1): 123–151.

Oberti, M. 2007a. *L'école dans la ville. Ségregation - mixité - carte scolaire.* Les Presses de Sciences Po.

———. 2007b. "Social and school differentiation in urban space: inequalities and local configurations." *Environment and Planning A* 39 (1): 208–227.

Oberti, M., and E. Préteceille. 2016. *La ségrégation urbain.* La Découverte.

Oberti, M., E. Préteceille, and C. Rivière. 2012. *Les effets de l'assouplissement de la carte scolaire dans la banlieue parisienn.*

OECD. 2013. *Definition of Functional Urban Areas (FUA) for the OECD metropolitan database.* Technical report. OECD.

Orfield, G. 2001. *Schools more separate: consequences of a decade of resegregation.* Technical report. Harvard Civil Rights Project, Cambridge, MA.

Osterman, P. 1991. "Welfare Participation in a Full-Employment Economy: The Impact of Neighborhood." *Social Problems* 38:475–491.

Pacchi, C., and C. Ranci. 2017. *White flight a Milano. La segregazione sociale ed etnica nelle scuole dell'obbligo.* Franco Angeli.

Palardy, G.J. 2013. "High School Socioeconomic Segregation and Student Attainment." *American Educational Research Journal* 50 (4): 714–754.

Pan Ké Shon, J. 2010. "The Ambivalent Nature of Ethnic Segregation in France's Disadvantaged Neighbourhoods." *Urban Studies* 47 (8): 1603–1623.

Parkin, F. 1979. *Marxism and Class Theory: A Bourgeois Critique.* Taylor & Francis Books.

Parliament, UK. 2008. *Children, Schools and Families - Third Report.* Technical report. UK Parliament.

Pearche, D. 1980. *Breaking down the barriers: New evidence on the impact of metropolitan school desegregation on housing patterns.* Washington DC: National Institute of Education.

Perna, L.W., and M.A. Titus. 2005. "The relationship between parental as social capital and college enrollment: An examination of racial/ethnic group differences." *Journal of Higher Education* 76:486–518.

Petsimeris, P. 1998. "Urban Decline and the New Social and Ethnic Divisions in the Core Cities of the Italian Industrial Triangle." *Urban Studies* 35 (3): 449–465.

Petsimeris, P., and S. Rimoldi. 2015. "Socio-economic divisions of space in Milan in the post-fordist era." In *Socio-Economic Segregation in European Capital Cities*, 210–237. Routledge.

Phillips, K., R. Rodosky, M. Munoz, and E. Larsen. 2009. "Integrated schools, integrated futures? A case study of school desegregation in Jefferson County, Kentucky." In *From the courtroom to the classroom: The shifting landscape of school desegregation,* edited by C. Smrekar and E. Goldring, 19–47. Harvard Education Press.

Picard, R. 2010. *GEONEAR: Stata module to find nearest neighbors using geodetic distances.* Boston College Department of Economics. Statistical Software Components S457146.

———. 2015. *GEOINPOLY: Stata module to match geographic locations to shapefile polygons.* Boston College Department of Economics. Statistical Software Components S458016.

Pickett, K.E., and M. Pearl. 2001. "Multilevel analyses of neighborhood, socioeconomic context and health outcomes: A critical review." *Journal of Epidemiology and Community Health* 55:111–122.

Pinkster, F. 2014. "Neighbourhood Effects as Indirect Effects: Evidence from a Dutch Case Study on the Significance of Neighbourhood for Employment Trajectories." *International Journal of Urban and Regional Research* 38 (6): 2042–2059.

Préteceille, E. 1995. "Division sociale de l'espace et globalisation. Le cas de la métropole parisienne." *Sociétés contemporaines* 22 (1): 33–67.

———. 2003. *La division sociale de l'espace francilien.* Technical report. Observatoire Sociologique du Changement - Sciences Po & CNRS.

———. 2009. "La ségrégation ethno-raciale a-t-elle augmenté dans la métropole parisienne?" *Revue français de sociologie* 50 (3): 489–519.

Rangvid, B.S. 2007. "Living and learning separately? Ethnic segregation of school children in Copenhagen." *Urban Studies* 44 (7): 1329–1354.

Rasch, G. 1961. "On general laws and the meaning of measurement in psychology." In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability,* 4:321–333.

Reardon, S.F., R.T. Brennan, and S.L. Buka. 2002. "Estimating multi-level discrete-time hazard models using cross-sectional data: Neighborhood effects on the onset of adolescent cigarette use." *Multivariate Behavioral Research* 37 (3): 297–330.

Reardon, S.F., and G. Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology* 32 (1): 33–67.

Reardon, S.F., and D. O'Sullivan. 2004. "Measures of spatial segregation." *Sociological Methodology* 34 (1): 121–162.

Reardon, S.F., and A. Owens. 2014. "60 Years after Brown: Trends and consequences of school segregation." *Annual Review of Sociology* 40:199–218.

Reardon, S.F., and J. Yun. 2005. "Integrating neighborhoods, segregation schools: The retreat from school desegregation in the South, 1990-2000." In *School resegregation: Must the South turn back?*, edited by J.C. Boger and Orfield G., 51–69. Chapel Hill: University of North Carna Press.

Reardon, S.F., J.T. Yun, and T. McNulty Eitle. 2000. "The changing structure of school segregation: measurement and evidence of multiratial metropolitan-area school segregation, 1989-1995." *Demography* 37 (3): 351–364.

Ricardo, D. 1951. "On the Principles of Political Economy and Taxation." In *The Works and Correspondance of David Ricardo*, edited by Sraffa P., vol. 1. Cambridge: Cambridge University Press.

Roberto, E., and J. Hwang. 2016. "Barriers to Integration: Institutionalized Boundaries and the Spatial Structure of Residential Segregation." *arXiv preprint arXiv:1509.02574*.

Rugh, J.S., L. Albright, and D.S. Massey. 2015. "Race, Space, and Cumulative Disadvantage: A Case Study of the Subprime Lending Collapse." *Social Problems* 62:186–218.

Rumberger, R.W., and G.J. Palardy. 2005a. "Does Segregation Still Matter? The Impact of Student Composition on Academic Achievement in High School." *Teachers College Record* 107 (9): 1999–2045.

———. 2005b. "Test scores, dropout rates, an transfer rates as alternative measures of school performance." *American Education Research Journal* 42:1–42.

Sacerdote, B. 2011. "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In *Handbook of the Economics of Education*, 3:249–277. Elsevier.

Sakoda, J.M. 1981. "A Generalized Index of Dissimilarity." *Demography* 18:245–250.

Scarpa, S. 2015. "Immigration policy regimes, welfare states and urban inequality patterns: A comparison between Malmo and Genoa." *European Urban and Regional Studies* 23 (4): 862–877.

Schelling, T.C. 1971. "Dynamic models of segregation." *Journal of Mathematical Sociology* 1 (2): 143–186.

Schofield, J. 1995. "Review of research on school desegregation's impact on elementary and secondary school students." In *Handbook of research on multicultural education*, edited by J. Banks and C. Banks. MacMillan Publishing.

Scriven, M. 1967. "The methodology of evaluation." In *Curriculum Evaluation*, edited by R. E. Stake. Rand McNally.

Seitles, M. 1998. "The perpetuation of residential racial segregation in America: Historical discrimination, modern forms of exclusion, and inclusonary remedies." *Journal of Land Use & Environmental Law* 14 (1): 89–124.

Shannon, C.E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (3): 379–423.

Shevky, E., and M. Williams. 1949. *The social areas of Los Angeles, analysis and typology.* University of California Press.

Silver, C. 1997. "The racial origins of zoning in American cities." In *Urban Planning and the African American Community: In the Shadows,* edited by June Manning Thomas and Marsha Ritzdorf. Sage Publications.

Simkus, A.A. 1978. "Residential Segregation by Occupation and Race in Ten Urbanized Areas, 1950-1970." *American Sociological Review* 43 (1): 81–93.

Simmons, R., L. Brown, D. Mitsch Bush, and D. Blyth. 1978. "Self-esteem and achievement of black and white adolescents." *Social Problems.*

Solon, G., M.E. Page, and G.J. Duncan. 2000. "Correlations between neighboring children in their subsequent educational attainment." *The Review of Economics and Statistics* 82 (3): 383–392.

Sørensen, A. 1979. "A Model and a Metric for the Analysis of the Intergenerational Status Attainment Process." *American Journal of Sociology* 85 (2): 361–384.

———. 1996. "The Structural Basis of Social Inequality." *American Journal of Sociology* 101 (5): 1333–1365.

———. 2002. "Foundations of Neo-Ricardian Class Analysis." In *Alternative Foundations of Class Analysis,* edited by Erik Olin Wright.

Szulnik, R., and J.O. Jonsson. 2007. *Ethnic segregation and educational outcomes in Swedish comprehensive schools.* Stockholm University Linnaeus Center.

Tauber, K.E., and A.F. Tauber. 1969. *Negroes in Cities: Residential Segregation and Neighborhood Change.* Atheneum.

Theil, H. 1972. *Statistical decomposition analysis.* Amsterdam: North-Holland Publishing Company.

Theil, H., and A.J. Finezza. 1971. "A note on the measurement of racial integration of schools by means of informational concepts." *Journal of Mathematical Sociology* 1:187–194.

Thompson, M.A. 2000. *Black-white residential segregation in Atlanta.* Russel Sage Foundation.

Torri, R., and T. Vitale. 2009. *Ai margini dello sviluppo urbano. Uno studio su Quarto Oggiaro: On the Fringes of Urban Development. A Study on Quarto Oggiaro.* Bruno Mondadori Editore.

Triventi, M., N. Panichella, G. Ballarino, C. Barone, and F. Bernardi. 2015. "Education as a positional good: Implications for social inequalities in educational attainment in Italy." *Research in Social Stratification and Mobility* 43:39–52.

Uyeki, E.S. 1964. "Residential distribution and stratification, 1950-1960." *American Journal of Sociology* 69 (5): 491–498.

Verdugo, G. 2011. "Logement social et ségrégation résidentielle des immigrés en France, 1968-1999." *Population* 66 (1): 169–193.

Vigdor, J., and T. Nechyba. 2007. "Peer effects in North Carolina public schools." In *Schools and the Equal Opportunity Problem,* edited by L. Woessmann and P. Peterson, 73–101. MIT Press.

Vogel, J. 1992. "Urban Segregation in Sweden Housing Policy, Housing Markets, and the Spatial Distribution of Households in Metropolitan Areas." *Social Indicators Research* 27:139–155.

Weaver, R.C. 1948. *The Negro Ghetto.* Russel & Russel.

Weber, M. 1968. "Economy and Society," edited by G. Roth and C. Wittich. New York: Bedminster Press.

Weiss, M.A. 1987. *The rise of community builders: the American real estate industry and urban land planning.* Columbia University Press.

Welch, F. 1987. *New Evidence on School Desegregation.* Technical report. Commission on Civil Rights, Washington, D.C.

Wessel, R.H. 1967. "A Note on Economic Rent." *American Economic Association* 57 (5): 1221–1226.

Wilkins, A.H. 1956. "The Residential Distribution of Occupation Groups in Eight Middle-Sized Cities of the United States in 1950." *Unpublished PhD Dissertation.*

Yinger, J. 1998. "Evidence on discrimination in consumer markets." *Journal of Economic Perspectives* 12 (2): 23–40.

Zoloth, B. 1976. "Alternative Measures of School Segregation." *Land Economics* 52 (3): 278–298.