Paper for the annual meeting of the Population Association of America 2018

Is childhood neighborhood segregation at the micro-level associated with college attendance? A longitudinal study of a Swedish city 1948-2011

Finn Hedefalk and Martin Dribe
Lund University
Centre for Economic Demography
Finn.Hedefalk@ed.lu.se

Abstract

The effect of socioeconomic background and segregation on education has been extensively studied; however, fewer longitudinal studies have analyzed the role of the socio-spatial segregation at the individual level in childhood for adult educational outcomes. We study the association between neighborhood socioeconomic conditions in childhood and educational choice net of any intergenerational transmission of disadvantage within the family. We analyze children growing up 1948-1967 in a Swedish industrial town (Landskrona), and follow their educational history throughout Sweden, 1968-2011. For the period 1948-1967, 54,000 individuals have been geocoded to address and building level, providing full residential histories of the individuals within the city. We find that the likelihood an individual gets a university degree is strongly associated with both the class and income of the close-proximity neighbors of same age in childhood. The main contributions of our study are the use of data on childhood segregation going back to 1950, and our measures of segregation based on individual neighborhoods (using the k-nearest neighbors at address-level), which is a more realistic approach than previously used measures that have been dependent on administrative borders.

1 Introduction

The social background as well as the socio-spatial segregation affect the socioeconomic and educational outcomes of individuals (Ainsworth 2002; Brooks-Gunn et al. 1993; Burger, 2010; Chetty et al. 2016; Jensen & Seltzer 2000; Crane 1991; Lee & Burkam, 2002; Szulkin & Jonsson, 2007; Wodtke et al. 2011). Similar effects have been found for ethnic and racial segregation in the United States on different socioeconomic outcomes (e.g. Borjas 1995; Cutler & Glaeser 1997). Much research has focused on the United States, which is a special context given the important interaction between race and socioeconomic status. Moreover, there has been an almost exclusive focus in the literature on more recent times, with few studies looking at the association between segregation in childhood and adult outcomes for earlier periods. Sweden, as well as many other countries, faces problems of socioeconomic bias in recruitment to higher education. According to the 2017 *Higher Education in Sweden* status report (Kolm et al., 2017), 69% of those having at least one parent with a higher education had themselves started a higher education at the age of 25, whereas only 22% of those having parents with less than an upper-secondary education had continued to higher education at age 25. Whereas the effect of socioeconomic background on education has been extensively studied, fewer longitudinal studies have analyzed the role of the socio-spatial segregation at the individual level in childhood for adult educational outcomes. That is, most segregation studies have been limited to analyses at the macro- and meso-level (various forms of administrative units such as wards and enumeration districts) (e.g., Sampson et al. 2002; Vartanian & Houser 2010; Bailey et al. 2017; Shertzer et al. 2016; Wodkte et al. 2011; Chetty et al. 2016). Hence, information on the segregation that occur within these aggregated units (see e.g., Logan & Bellman 2016) are often lost. Modern, and sometimes historical, datasets in which individuals are geocoded at the micro-level (e.g., buildings and address points) are available (e.g., Connor 2017). However, they are seldom linked to large individual-level longitudinal demographic databases that cover a long period. Therefore, such datasets are incompatible with historical longitudinal analyses.

We study the association between neighborhood socioeconomic conditions in childhood and educational choice net of any intergenerational transmission of disadvantage within the family. Do individuals from low-class origins benefit from living in a high-status neighborhood compared to living in a deprived neighborhood? Similarly, do individuals from high-class origins make different educational choices depending on the socioeconomic characteristics of their childhood neighborhood? We study these issues for children growing up 1948-1967 in a Swedish industrial port town (Landskrona), and follow their educational history throughout Sweden in the period 1968-2011. We use a geocoded historical database (1948-1967) combined with a modern longitudinal database (1968-2011). For the period 1948-1967, about 54,000 individuals have been geocoded to both address and building level, providing full residential histories of the individuals within the city. In addition, we have created a temporal representation of geographical objects such as buildings and roads for the period; i.e., we have information on when an object started and ceased to exist. Therefore, this geocoding in combination with the historically correct geography of the city enable us to estimate micro-level segregation indices based on individually-based neighborhoods (cf. Östh et al. 2015; Logan & Shin 2016). To analyze the educational outcomes in the adulthood, we use longitudinal register data from Statistics Sweden on the full Swedish population for the period 1968-2011.

The main contributions of our study is: (1) to use data on childhood segregation going back to 1950; and (2) to use a measure of segregation based on individual neighborhoods, which is a more realistic approach than previously used measures that have been dependent on administrative borders. In future work of this study, we will analyze the spatio-temporal patterns and trends of segregation and neighborhood effects, as well as improve the segregation measures by incorporating the city's street network and buildings.

The main hypothesis is that an individual's education level in adulthood is affected by the socioeconomic status of the close proximity neighbors in childhood, net of their own socioeconomic background. For example, an individual with low socioeconomic status growing up in an area with relatively many neighbors of high socioeconomic status are expected to have a higher level of education in adulthood compared to an individual from the same group growing up in a segregated area of low socioeconomic status. We expect that the neighborhood effects in childhood differ depending on the socioeconomic origin of the individual; i.e., individuals from a high socioeconomic background are more robust against the neighborhood effects compared to individuals with a low socioeconomic status. There are several reasons for such neighborhood effects on educational choices, related, for example, to peer influence and adults acting as role models, promoting a set of norms which affect the likelihood of opting for different career paths (e.g. Ainsworth 2002; Crane 1991; Galster 2012; Wilson 1987). Poorer neighborhood may also be exposed to adverse environmental conditions, high crime rates, etc. which could affect school choices both directly and indirectly (e.g. through bad health, or cognitive disorders) (see, e.g. Case and Katz 1991; Galster 2012). Neighborhood effects could also be mediated by school characteristics, especially in contexts where school choice is structured by neighborhood, but the empirical evidence for its actual importance is mixed (see Ainsworth 2002; Wodtke & Parbst 2017).

We believe that our geocoded historical database at the building and address level presents an opportunity to incorporate more sophisticated measures of childhood segregation than have previously been possible in longitudinal historical analyses. By using such measures in a study that covers a relatively long period, we expect to increase the knowledge on the mechanisms behind historical and modern social recruitment biases in education.

2 Study area and data

We use data for the city of Landskrona from the Scanian Economic Demographic Database (SEDD), created by the Centre for Economic Demography at Lund University, in collaboration with the Regional Archives in Lund (Bengtsson et al., 2014). During the post WWII period Landskrona represented a medium-sized Swedish city with a strong manufacturing and shipyard industry. In 1949, it had a population of 25,000, in 1960 28,000, and in 2000 27,500. The primary sources for the database are continuous population registers, and income- and taxation registers. Information on birth, marriages, deaths, as well as in- and out-migrations has also been linked to the data. The historical data (before 1968) have been linked to national longitudinal register data for the period 1968-2011 from Statistics Sweden, The National Board of Health and Welfare, and the Swedish Defense Recruitment Agency. Together these data enable us to follow children growing up in the city to adulthood wherever they reside in Sweden, thereby avoiding bias from only looking at the stayer population.

We have geocoded the population of Landskrona for the period 1948-1967. It was carried out in three process steps: (1) standardization of historical addresses in the demographic data; (2) linkage of the historical addresses to modern geographical address points and buildings (provided by the Swedish mapping, cadastral, and land registration authority); and (3) creation of continuous information on when each building, street segment and other relevant geographic information was created and ceased to exist. Step two was achieved by using modern geographical data provided by the *Lantmäteriet* (the Swedish mapping, cadastral, and land registration authority) and by Landskrona municipality. Step three was achieved by using georeferenced historical maps and aerial photographs in combination with information on creation year of the building and the date when a street was first mentioned in the historical demographic data (i.e., an indication of when the street was created). Approximately 56,000 individuals have been linked to the addresses and buildings they resided in. So far, we have geocoded 95% of the individuals' time at risk (we aim to obtain a 99% geocoding), in which their residential histories within Landskrona are traced. In addition, we have created an object lifeline representation of 90% of the buildings and roads in Landskrona. That is, we have information on when a road and building started and ceased to exist. This allows us to use geographical data that are correct for each time point that we study, which is important when estimating individually-based neighborhoods. For example, Figure 1 shows the main parts of Landskrona for two points in time: 1950 and 1960. The figure also shows the resident locations of a geocoded individual for the two time points. In the year 1950, the individual resides in an apartment block in the center of the city; in the year 1960, the individual has moved to a newly built detached area. Thus, this individual will be exposed to different segregation and environment estimates throughout the childhood.

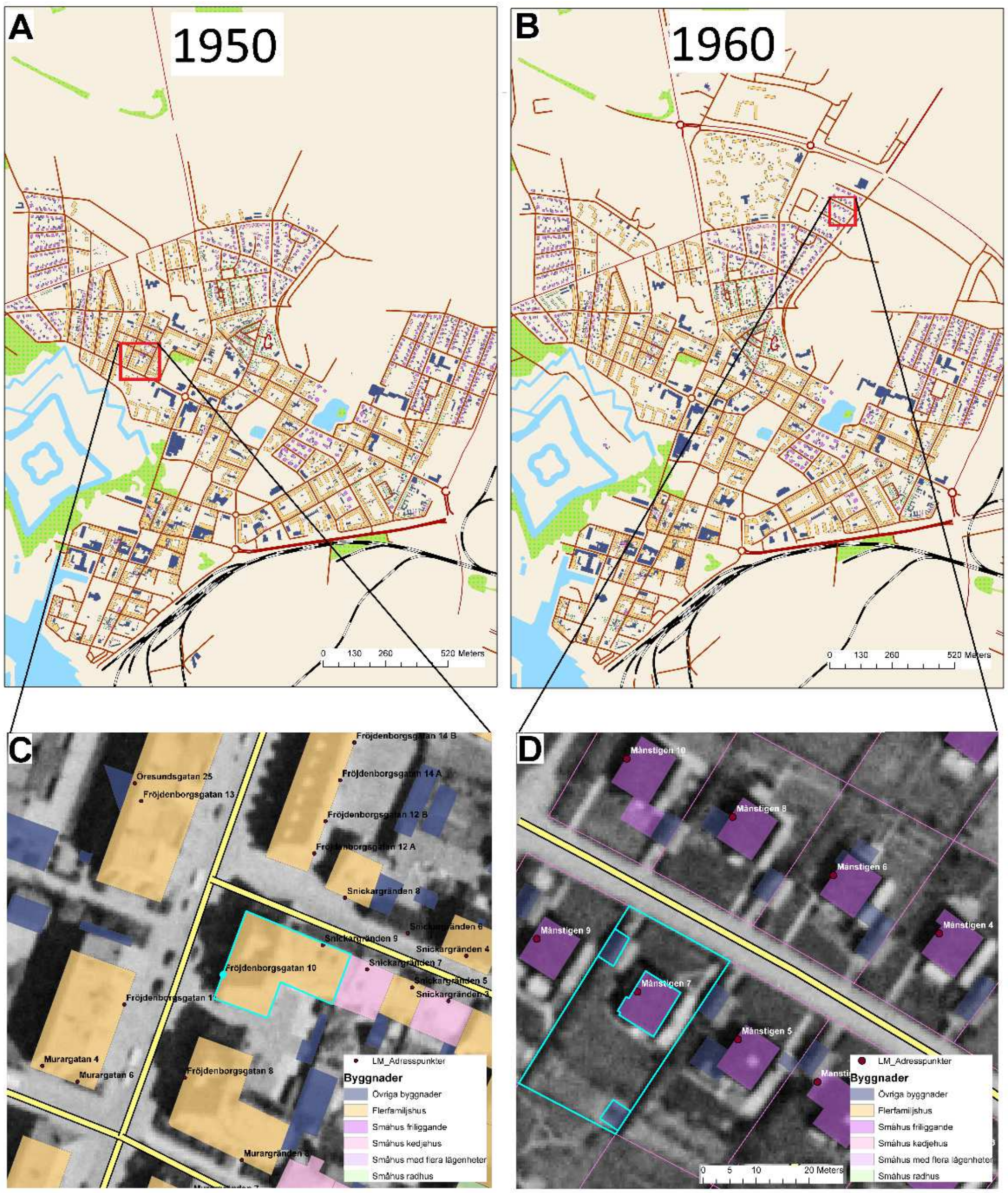

Figure 1: Geocoding example at the micro-level in the city of Landskrona. A, B) Parts of Landskrona in 1950 and 1960 (note the geographic change between the two time points); C, D) The residential location of a geocoded individual in 1950 and 1960.

3 Methods

The overall procedure for analyzing the effects of the socioeconomic context of the neighborhood in childhood on adult educational choices is performed in two main steps.

In the first step, we estimate individual and neighborhood variables for the children aged 13-14 in Landskrona between 1948 and 1967. We focus of the age group 13-14 because peer influence may affect these children's choice of high-school orientation (Roderick, 2003; Calster, 2011). At the individual level, we have detailed economic and demographic information on the social class (based on father's occupation) and family income, as well as on other family specific variables. Therefore, we use the variables social class and family income as measures of socioeconomic status of the children. Socioeconomic status rather than education of the parents are used to predict adulthood educational outcomes because of the low percentage adults (i.e., parents) having a higher education during the period 1948-1967.The great expansion of higher education in Sweden did not take place until the 1960s and 1970s (see, e.g. Stanfors 2003: 154).

In the second step, children growing up in Landskrona 1948-1967 are followed throughout their life-courses, regardless of where they live within Sweden. For these individuals, we estimate logistic regression models to analyze the effects of childhood neighborhood conditions on the education level in the adulthood.

3.1 Quantification of neighborhood variables

Using the geocoded data on address level, we estimate segregation indices based on individual neighborhoods. The first step was to construct annual matrices containing the shortest Euclidean distances between every child aged 13-14. The final dataset contained yearly snapshot information on each individual's neighbors at the end of the year (19xx-12-31), for the period 1948-1967, and the neighbors were ranked based on their proximity to the individual. From these matrices, we created individual neighborhoods from the k-nearest neighbors that did not live within the same household of the individual. In this study, we use the 13-nearest neighbors of same age; however, because of this arbitrary selection, tests were also done on the 25-, 50-, and 100-nearest neighbors of same age. In addition, we performed tests on the 13-, 25-, 50- and 100-nearest neighbors of age 13 and over. Based on the individual neighborhoods, we constructed two variables: Geographically Weighted (GW) income of the neighborhood, called *neighborhood income*, and *neighborhood class.*

The *neighborhood income* variable is based on the square root scale family income (family income divided by the square root of family size) of each neighbor for a specific year. For the k-nearest neighbors of individual *i*, we define the geographically weighted mean income as follows:

$$GW\ mean\ income = \sum_{j=1}^{j=n}(Inc_j \times RelW_{ij}) \tag{1}$$

$$RelW_{ij} = \frac{W_{ij}}{\sum_{j=1}^{j=n} W_{ij}}\ , \qquad W_{ij} = e^{-0.5\cdot\left(\frac{d_{ij}}{b}\right)^2}$$

where Inc_j denotes the square root family income of neighbor j. Moreover, $RelW_{ij}$ is the relative spatial weight between individual i and neighbor j (relative to the spatial weight of all other neighbors), in which W_{ij} is the spatial weight implemented as a Gaussian distance function between individual i and neighbor j (cf. Fotheringham et al. 2003). In the Gaussian distance function, the bandwidth b limits the search of the neighbors and d_{ij} is the Euclidean distance between the address points of individual *i* and neighbor *j*. In this study, we use an adaptive bandwidth, which uses the maximum distance between individual *i* and its k-neighbors. Figure 2 shows an example of the 13 nearest neighbors *j* of same age to an individual *i*, in which each neighbor's family income is weighted according to Eq. (1).

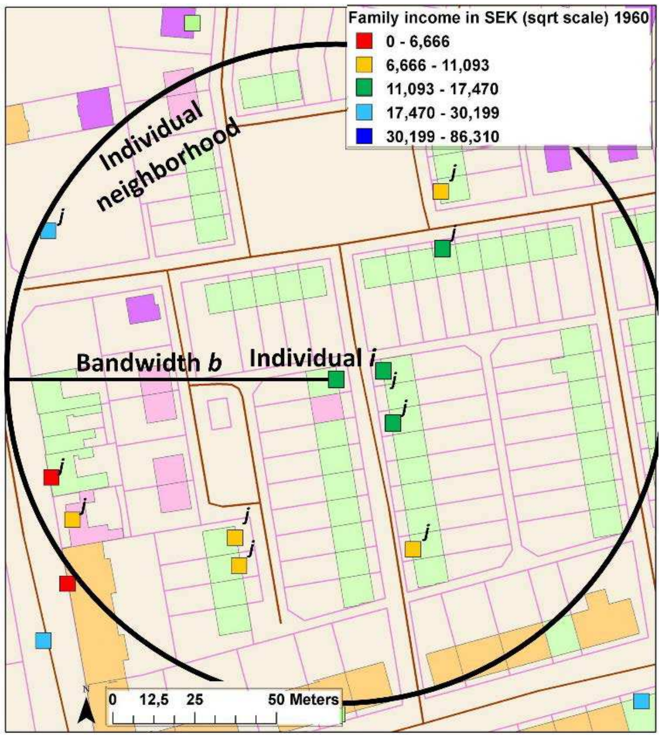

Figure 2: Example of an individual neighborhood for individual *i*, containing the 13 nearest neighbors of ages 13-14 in Landskrona, 1960. The background map shows buildings, property unit borders (purple lines), and streets (brown lines) in 1960. Note: some neighbors resides at the same address point.

The *neighborhood class* variable is based on the social class (HISCLASS) of the nearest k-neighbors. That is, we count how many neighbors from each social class that are currently present within every individual neighborhood and year. This variable is further described in the section 3.2 Statistical analysis.

3.2 Statistical analysis

By estimating seven logistic regression models (Table 1), we analyze how neighborhood conditions affect education in adulthood. The binary outcome variable of interest is whether or not the highest level of education reached at age 40 is at least a 3-year university degree. We include control variables in a stepwise manner to observe changes in the associations compared to the previous models. At each step, we either perform likelihood-ratio tests for the nested models to see if each extended model improve the previous model, or use Bayesian information criterion (BIC) to compare the fit between the non-nested models (e.g., model 1 and 2). Model 7 includes an interaction between the variables class origin and neighborhood class.

Table 1: Logistic regression models used in the study. Variables included in the models are marked by an x

Model	Variables									
	Class origin	Income origin	Neigh. class	Neigh. Income	Older brothers	Household size	Mother pres.	Res. city	Birth year	Sex
1 Class origin	x								x	x
2 Income origin		x							x	x
3 Neighborhood class	x		x						x	x
4 Neighborhood income	x			x					x	x
5 Neighborhood class full	x		x		x	x	x	x	x	x
6 Neighborhood income full	x			X	x	x	x	x	x	x
7 Class origin – neigh. class Interactions	x		x		x	x	x	x	x	x

The variables used in the models and presented in Table 1 are described as follows. Note that all childhood variables are estimated from annual snapshots when the individual was 13 and 14 years old. An average value for the two time points is thereafter used for the continuous variables (e.g., income origin), whereas the most recent value in time is chosen for the categorical variables (e.g., presence of mother).

Education level. The binary outcome variable, which is based on information on the highest level of education reached at age 40 (an age of when most people have finished their education in Sweden). This information is coded according to the national standard Swedish education nomenclature (SUN) for classification of educations. A value of 1 indicates that the individual has completed a 3-year university degree or higher. A 3-year university degree is set as limit to eliminate some of the vocational programs.

Class origin. We measure socioeconomic status based on father's occupation. Data on occupation is obtained from several sources: demographic events, population registers, as well as annual data from the income registers. Occupational notations are coded in an internationally comparable coding scheme for historical occupations (HISCO) (Van Leeuwen et al. 2002) and then grouped into HISCLASS, a 12-category occupational classification scheme based on skill level, degree of supervision, whether manual or non-manual, and whether urban or rural: 1) higher managers; 2) higher professionals; 3) lower managers; 4) lower professionals; clerical and sales personnel; 5) lower clerical and sales personnel; 6) foremen; 7) medium-skilled workers; 8) farmers and fishermen; 9) lower-skilled workers; 10) lower-skilled farm workers; 11) unskilled workers; and 12) unskilled farm workers (Van Leeuven and Maas 2011). We further aggregate the HISCLASS categories into three status groups: high-class (HISCLASS: 1-6), mid-class (HISCLASS 7), low-class (HISCLASS 8-12).

Neighborhood class: We use a simple measure of the relative presence of high-class neighbors in the 13-neighborhood of same age. We categorize the neighborhood class using Location Quotients (LQ) for each year. An LQ is the ratio of a local ratio to the global ratio (cf., Cromley and Hanink, 2012). For a specific year, considering neighbors of same age, the neighborhood class variable is defined as:

$$LQ_i = \frac{h_i/n_i}{H_j/N_j} \tag{2}$$

where h_i is the number of high-class neighbors, n_i is the total number of neighbors, H_j is the number of high-class children of same age in the city, and N_j is the current population of children aged 13-14 in the city.

The following variable groups are created:

- High-class (LQ>=1.2): A 13-neighborhood with at least 20% more high-class neighbors than the city average for a specific year.
- Mid-class (0.8<=LQ<1.2): A 13-neighborhood with a high-class neighbors' presence of at least 80% of the city average and not more than 20% more high-class neighbors than the city average.
- Low-class (LQ < 0.8): A 13-neighborhood with a high-class neighbors' presence less than 80% of the city average for the specific year.

Income origin: We use the square root scale to define the family income for a specific year (family income divided by the square root of family size, see e.g., (OECD 2011)). To account for inflation through time, we use the ratio of the family income compared to the average family income for the current year. Because this ratio is the same measure in principle as the LQ for the neighborhood class, the term LQ is used to facilitate interpretation of the results. Three variable groups are defined:

- High-income (LQ>=1.2): a family income at least 20% higher than the city average for a specific year.
- Mid-income (0.8<=LQ<1.2): a family with income of at least 80% of the city average and at most 20% higher than the city average.
- Low-income (LQ < 0.8): a family income less than 80% of the city average for the specific year.

Neighborhood income: Geographically weighted income of the closest 13 neighbors of the same age (13-14 years old). As for the income origin, we use the ratio of the neighborhood income compared to the average neighborhood income for the current year (defined as LQ, see Eq. (2)). Three variable groups are defined:

- High-inc (LQ>=1.2): A neighborhood income at least 20% higher than the city average neighborhood income for a specific year.
- Mid-inc (0.8<=LQ<1.2): A neighborhood income of at least 80% of the city average neighborhood income for a specific year, and not more than 20% higher than the city average.
- Low-inc (LQ < 0.8): A neighborhood income less than 80% of the city average for a specific year.

Sibling sex composition – Number of older brother in childhood: The presence of older brothers as well as birth order may affect educational outcomes (Butcher and Case, 1994; Haan, 2005; Jacob, 2011). For example, Jacob (2011) found that older brothers negatively affected the probability of their sisters to get a university degree. The evidence of sibling effects is, however, mixed. Therefore, as sensitivity tests, we include number of older sisters, younger sisters and younger brothers in the models as well.

Household size: A continuous variable of the number of members in the household in childhood. As for the number of older brothers, family and household size may in general affect the academic performance of children, for example through resource dilution (Blake 1989; Downey, 1995).

Presence of mother: A binary variable of whether the mother is present or not at the ages of 13-14. Research has shown that mothers have a strong influence on their children's academic success (e.g., Englund et al., 2004). Therefore, having lost a mother may negatively affect the probability of getting a university degree. Sensitivity tests with the presence of father is performed as well.

Currently residing in Landskrona: A binary variable of whether the person resides in Landskrona at age 40 or not. Because Landskrona does not have a university, individuals that aims to get an academic degree need to move away from the city.

Birth year: Year of birth of the individual

3.3 Descriptive statistics

Table 2 shows selection procedure of the individuals that were aged 13-14 in Landskrona, for the period 1948-1967. The selection criteria was as follows. Individuals that were children in Landskrona 1948-1967 had to be linked to the register data for the period 1968-2011 from Statistics Sweden. Because the education level reached at age 40 was the focus of this paper, survival until age 40 was a criterion for inclusion in the dataset. Lastly, the individuals' address in Landskrona had to be recorded; therefore, those with missing address information were excluded from the analyses. Based on this selection procedure, our final dataset contained 7025 individuals, 90% of the individuals that had resided in Landskrona at ages 13-14 (Table 1).

Table 2: Stepwise selection of individuals in Landskrona for the period 1948-1967

Variable	Observed	% the original children sample
Individuals all ages	54488	
Children aged 13-14	7830	100
Children that can be linked to SCB	7764	99.16
Children with address that can be linked to SCB	7455	95.21
Adults survived until age 40 in the SCB database	7267	92.81
Adults survived until age 40 in the SCB database and have address	7025	89.72

Tables 3-4 present the distribution of the observations (in percentages) among the categorical variables and the average values of the continuous variables, respectively. The continuous LQ variables in Table 3 were not used in the estimated models in this paper; however, because the categorical variables are created from them, they are presented here for descriptive purposes.

Table 3: Distribution in percentage of the individuals on the categorical variables for the study sample (N = 7025).

Variable	%
Class origin	
High-class.	39.47
Mid-class	26.72
Low-class	30.27
NA	3.54
Income origin	
High-inc. (>=1.2)	22.67
Mid-inc. (0.8-1.2)	38.57
Low-inc (<0.8)	38.76
Neighborhood class	
High-class (>=1.2)	29.12
Mid-class (0.8-1.2)	29.04
Low-class (<0.8)	41.84
Neighborhood income	
High-inc (>=1.2)	13.04
Mid-inc (0.8-1.2)	75.21
Low-inc (<0.8)	11.76
Sex	
Female	51.07
Male	48.93
Resides in Landskrona	
Yes	50.51
No	49.49
Mother present	
Yes	96.90
No	3.12

Table 4: Average values of the continuous variables for the individuals in the study sample (N = 7025).

Variable	Mean	Min	Max	SD
Birth year	1945.28	1935	1954	5.30
Older brothers	0.29	0	5	0.59
Household size	4.16	0	11	1.27
Neighborhood income	1.00	0	3.87	0.24
Income origin	0.98	0	19.94	0.74
Neighborhood class	0.97	0	3.81	0.49

Moreover, Figure 2 presents, as an example, the spatial distribution of the neighborhood income variable in 1960. Each symbol (square, triangle, and dot) represents the location and variable group for a child that was either 13 or 14 at the time point 1960-12-31. The high-income groups resides mostly in single family homes or in apartment blocks within the center of Landskrona (lower center of the map), whereas the low-income group mainly resides in larger apartment blocks. The children from the mid-income group are spread on both apartment blocks as well as on semi-detached/terrace houses.

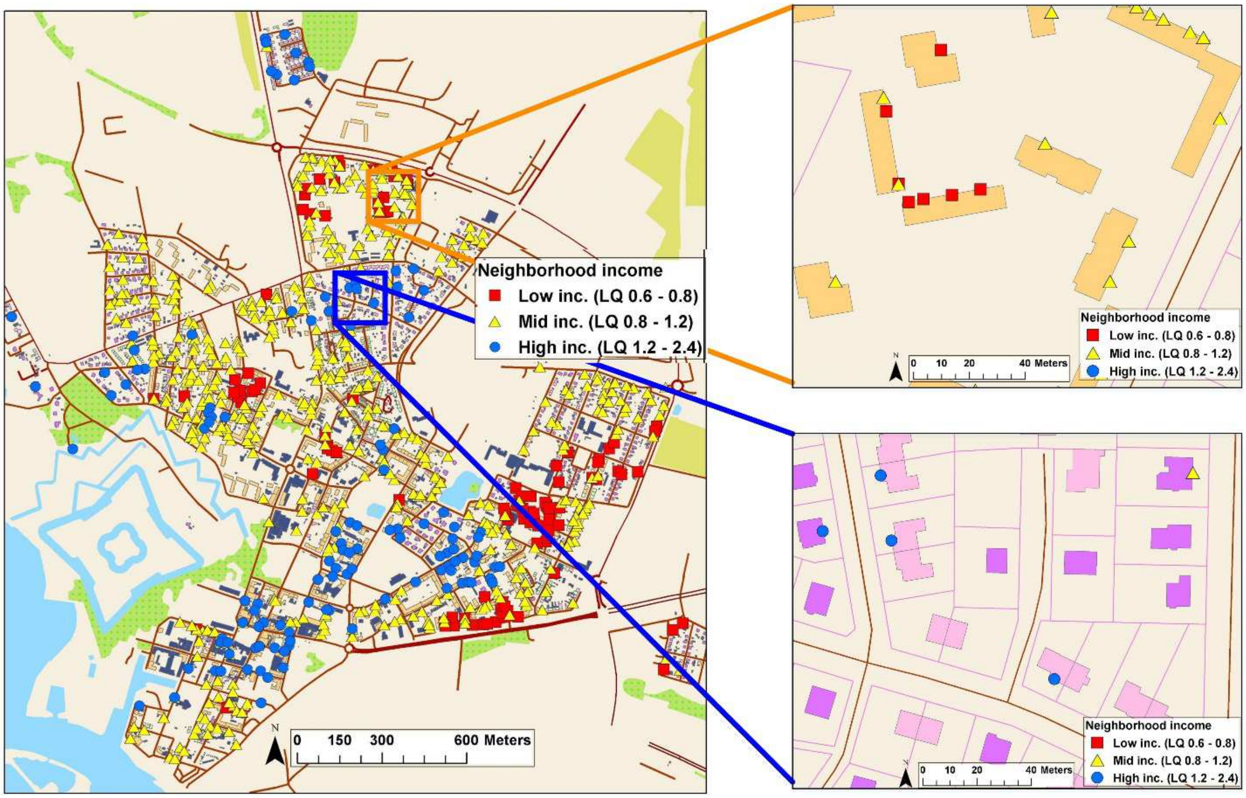

Figure 2: Spatial distribution of the neighborhood income variable in Landskrona, 1960. A) Landskrona city; B) Low- and mid-income area with apartment blocks; C) High-income area with single family homes and semi-detached houses.

4 Results

Table 5 shows the results of six logit models, which primarily estimate the association between neighborhood conditions (class and geographically weighted income of the 13 nearest childhood neighbors of same age), and having a university degree at age 40. Across all models, there is a strong relationship between both class origin and neighborhood conditions on the one hand and the likelihood of having a university degree, on the other. It is also interesting to note that the relationships are fairly linear, and that it is not only living in a low-income or low-class neighborhood that matters, but there are also differences between mid-level and high/low levels (cf. Galster 2012).

In the basic model 1, which only includes class origin, individuals from high-class origins had a 1.75 higher odds of obtaining at least a 3-year university degree compared to the individuals from low-class origins. As seen in model 2, also family income in childhood is positively associated with the likelihood of having a university degree. A comparison of the AIC and BIC values for model 1 and 2 indicated a better fit for model 1; therefore, class origin instead of income origin was used as control variable in models 3-6. Models 3 and 4 extend the basic model 1 by including class and income, respectively, of the 13 closest childhood neighbors of same age. These models show significant associations with both neighborhood class and neighborhood income; e.g., those that resided in a high-class neighborhood (i.e., individuals having at least 20% more high-class neighbors than average) had a 1.31 higher odds of obtaining a university degree compared to those that lived in a low-class neighborhood.

Models 5 and 6 show that the magnitude of the associations for both class origin and neighborhood conditions increases when extending models 3 and 4, respectively, by controlling for other family specific and adulthood variables. For example, individuals that resided in high-income neighborhoods (model 6) were more than twice as likely to have a university degree compared to those that resided in low-income neighborhoods.

Table 5: Association between neighborhood conditions and other factors at ages 13-14 and having a university degree at age 40. Landskrona and Sweden, 1948-2011.

	1 Class origin		2 Income origin		3 Neigh. class		4 Neigh. income		5 Neigh. class full		6 Neigh. income full	
Variable	**OR**	**P>z**	**OR**	**P>z**	**OR**	**P>z**	**OR**	**P>z**	**OR**	**P>z**	**OR**	**P>z**
Class origin												
High-class	1.75	0.00			1.67	0.00	1.69	0.00	2.81	0.00	2.86	0.00
Mid-class	1.14	0.05			1.13	0.06	1.14	0.05	1.39	0.00	1.41	0.00
Low-class	1.00	rc			1.00	rc	1.00	rc	1.00	rc	1.00	rc
NA	0.99	0.92			0.97	0.83	0.97	0.83	1.25	0.33	1.25	0.33
Income origin												
High-inc. (>=1.2)			1.54	0.00								
Mid-inc. (0.8-1.2)			1.14	0.02								
Low-inc. (<0.8)			1.00	rc								
Neigh. class												
High-class (>=1.2)					1.31	0.00			1.69	0.00		
Mid-class (0.8-1.2)					1.18	0.01			1.18	0.06		
Low-class (<0.8)					1.00	rc			1.00	rc		
Neigh. income												
High-inc. (>=1.2)							1.46	0.00			2.01	0.00
Mid-inc. (0.8-1.2)							1.17	0.05			1.18	0.18
Low-inc. (<0.8)							1.00	rc			1.00	rc
Birth year	1.07	0.00	1.07	rc	1.07	0.00	1.07	0.00	1.05	0.00	1.05	0.00
Sex												
Female	1.00	rc	1.00	rc	1.00	rc	1.00	rc	1.00	rc	1.00	rc
Male	1.27	0.00	1.28	0.00	1.27	0.00	1.28	0.00	1.09	0.21	1.10	0.18
Res. Landskrona												
Yes									0.29	0.00	0.28	0.00
No									1.00	rc	1.00	rc
Older brothers									0.85	0.03	0.85	0.03
Household size									0.89	0.00	0.89	0.00
Mother present												
Yes									2.75	0.00	2.72	0.00
No									1.00	rc	1.00	rc
LR chi2	316.07		263.66		334.61		328.24		707.46		705.82	
Prob>chi2	0.00		0.00		0.00		0.00		0.00		0.00	
Subjects	7025											

OR = Odds Ratio, rc = reference category.

Model 7 extends model 5 in table 4 by including an interaction between class origin and neighborhood class (Table 6). As seen in the model, the childhood neighborhood was associated with obtaining a university degree for social class groups. The individuals from high-class origin residing in a high-class neighborhood had the highest likelihood of all groups to eventually get a university degree. The childhood class had a strong association with education by itself as well: individuals from high-class origins that resided in a low-class neighborhood had still a higher chance to get a university degree than the mid-class and low-class groups, regardless of the class of their childhood neighbors. However, individuals from low-class origins residing in a high-class neighborhood were somewhat more likely to get a university degree than the mid-class groups that resided in low- and mid-class neighborhoods.

Table 6: Association between neighborhood conditions and other factors at ages 13-14 and having a university degree at age 40. Interaction between class origin and neighborhood class. Landskrona and Sweden, 1948-2011

7 Interactions		
Variable	**OR**	**P>z**
Class origin # Neigh. class		
High-class # High-class (>=1.2)	4.60	0.00
High-class # Mid-class (0.8-1.2)	3.54	0.00
High-class # Low-class (< 0.8)	2.80	0.00
Mid-class # High-class (>=1.2)	2.65	0.00
Mid-class # Mid-class (0.8-1.2)	1.37	0.09
Mid-class # Low-class (< 0.8)	1.46	0.02
Low-class # High-class (>=1.2)	1.74	0.00
Low-class # Mid-class (0.8-1.2)	1.16	0.46
Low-class # Low-class (< 0.8)	1.00	rc
NA # High-class (>=1.2)	2.11	0.05
NA # Mid-class (0.8-1.2)	2.25	0.03
NA # Low-class (< 0.8)	0.82	0.65
Birth year	1.05	0.00
Sex		
Female	1.00	rc
Male	1.09	0.20
In Landskrona		
Yes	0.28	0.00
No	1.00	rc
Older brothers	0.85	0.03
Household size	0.90	0.00
Mother present		
Yes	2.67	0.00
No	1.00	rc
LR chi2	709.47	
Prob>chi2	0.00	
Subjects	7025	

5 Discussion and conclusions

The main aim of this study was to analyze the effect of neighborhood conditions in childhood on educational choice in adulthood. We studied these issues for individuals being 13-14 years old 1948-1967 in the Swedish town Landskrona, and which were followed throughout Sweden until they reached the age of 40. We find support for our main hypothesis that the likelihood an individual gets a university degree is strongly associated with both the class and income of the close-proximity neighbors of same age in childhood, net of their father's social class and own family's income. Moreover, we found little support for our second hypothesis; i.e., that individuals from high-class origin were more robust against neighborhood effects compared to individuals with low-class origin. The results from the interaction model indicate that the individuals from the high-class origin were only slightly less affected by the neighborhoods compared to the individuals from low-class origin. In addition, the biggest within-group differences were observed between the low-class and high-class neighborhoods in the mid-class group; hence, we find no trend in robustness from neighborhood effects from low-class to high-class origin In addition, individuals from all class origins were strongly affected by their close proximity neighbors.

By using a geocoded historical database at the address and building level, we have been able to incorporate measures of childhood segregation at a very fine scale. For each 13-14 year old individual residing in Landskrona 1948-1967, we have been able to identify their closest neighbors at same age based on their residential addresses. In addition, we have followed the complete historical population of children aged 13-14 in Landskrona city, and thereafter when they have grown up through adulthood wherever they reside in Sweden. Therefore, we have avoided bias by only looking at the stayer population. By using such detailed measures that covers a long period we believe that this study can contribute with knowledge on the mechanisms behind historical and modern social recruitment biases in education. To the best of our knowledge, this study is also one of the first to analyze the association between segregation at the micro-level, using measures of individual neighborhoods, on adult educational choices using longitudinal data.

In future work, we plan to extend this study by making some additional improvements of the segregation indices and models.

First, we aim to improve the distance-measures used in the segregation indices. We estimated the neighborhood conditions by creating individually based neighborhoods, using the k-nearest neighbors approach (see e.g., Östh et al. (2015)). For the neighborhood income variable, we also modelled the influence of neighbors' income using a non-linear Gaussian distance function, in which nearby neighbors' income had a greater influence on each individual's variable compared to neighbors' father away. Such function usually better capture the dependency found in many spatial relationships compared to linear distance functions (Fotheringham, 2003). However, for all indices we used the Euclidean distance to estimate the closest neighbors. In most cities, two individuals that have a short Euclidean distance between each other may be separated by a highway or a river, meaning that they are likely not exposed to each other. A more realistic measure, therefore, is to utilize the road network of Landskrona and estimate a distance matrix based on the network distance. Then, possible spatial constraints such as highways, can be incorporated as well. A challenge with this approach is to correctly model the temporal dimension of the road network.

Second, we will in more detail analyze the neighborhood effects at various level of details, both spatial and temporal. At the spatial scale, we plan to estimate indices that are based on both fewer and more neighbors in the neighborhoods. In addition, more geographical variables can be included in the models; e.g., population density, crowding, type of residential building, and historical school districts (if available). At the temporal scale, we defined the neighborhoods by using snapshot information on yearly basis. However, it is important to consider the full residential histories in Landskrona of the individuals when estimating the neighborhoods (see Wodtke et al. 2011). By using such information, it will be possible to estimate more realistic estimates of the exposure of neighbors throughout the childhood. Such index can be defined on e.g., a monthly basis, or by updating the index at each migration event within the city (this is, however, computational intensive). As we add data even further back in time we will also be able to assess if the importance of the neighborhood has increased or not as higher education expanded.

A third limitation with the study regards the possible spatial autocorrelation of the segregation indices; i.e., it violates the model assumption that the observations are independent from each other. To overcome this problem, we may use spatial autoregressive models which takes the spatial autocorrelation into account. Additionally. Geographically Weighted Regressions can applied to investigate how well the model performs at different locations within the city, which may reveal patterns not found in the non-spatial models.

6 References

Ainsworth, J. W. (2002). Why does it take a village? The mediation of neighborhood effects on educational achievement. *Social Forces* 81(1): 117-152.

Bailey. N.. Gent. W. P.. & Musterd. S. (2017). Remaking Urban Segregation: Processes of Income Sorting and Neighbourhood Change. *Population. Space and Place*. 23(3).

Bengtsson. T.. Dribe. M.. Quaranta. L.. & Svensson. P. (2014). *The Scanian Economic Demographic Database. Version 4.0* (Machine-readable database).

Borjas, G. J. (1995). Ethnicity, neighborhoods, and human-captial externalities. *American Economic Review* 85(3): 365-390.

Blake, J. (1989). *Family Size and Achievemen*t. Berkeley: University of California Press.

Brooks-Gunn, J., Duncan, G. J., Klebanov, P. K., & Sealand, N. (1993). Do neighborhoods influence child and adolescent development? *American Journal of Sociology* 99(2): 353-395.

Burger. K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early childhood research quarterly*. *25*(2). 140-165.

Case, A. C., & Katz, L. F. (1991). The company you keep: the effects of family and neighborhood on disadvantaged youths. NBER Working Paper 3705.

Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review* 106(4): 855-902.

Connor. D. S. (2017). Poverty. Religious Differences. and Child Mortality in the Early Twentieth Century: The Case of Dublin. *Annals of the American Association of Geographers*. 107(3). 625-646.

Crane, J. (1991). The epidemic theory of ghettos and neighbourhood effects on dropping out and teenage childbearing. *American Journal of Sociology* 96(5): 1226-1259.

Cutler, D. M., & Glaeser, E. D. (1997). Are ghettos good or bad? *Quarterly Journal of Economics* 112(3): 827-872.

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). Geographically weighted regression: The analysis of spatially varying relationships. New York: Wiley.

Galster, G. S. (2012). The mechanism(s) of neighborhood effects: theory, evidence, and policy implications. In: van Ham, M. et al. (eds) *Neighborhood Effects Research: New Perspectives*. New York: Springer.

Jacob, M. (2011). Do brothers affect their sisters' chances to graduate? An analysis of sibling sex composition effects on graduation from a university or a Fachhochschule in Germany. Higher Education, 61(3), 277-291.

Jensen, B., & Seltzer, A. (2000). Neighborhood and family effects in educational progress. *Australian Economic Review* 33(1):17-31.

Kolm. S. B.. Dryler. H.. Egeltoft. T.. Elenäs. J.. Gribbe. J.. Haglund. A.. Kahlroth. M.. Karlsson. N.. Nilsson. S.. Petterson. I.. Sadurskis. A.. Severin. J.. Svensson. F.. & Viberg. A. (2017). *Higher Education in Sweden status report - Trends and Developments*.

Lee. V. E.. & Burkam. D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children begin school*. Economic Policy Institute. 1660 L Street. NW. Suite 1200. Washington. DC 20036.

Logan. J. R.. & Bellman. B. (2016). Before The Philadelphia Negro: Residential Segregation in a Nineteenth-Century Northern City. *Social Science History*. 40(4). 683-706.

Logan. J. R.. & Shin. H. J. (2016). Birds of a feather: social bases of neighborhood formation in Newark. New Jersey. 1880. *Demography*. 53(4). 1085-1108.

Mirowsky. J. (2017). *Education. social status. and health*. Routledge.

Roderick, M. (2003). What's happening to the boys? Early high school experiences and school outcomes among African American male adolescents in Chicago. Urban Education, 38(5), 538-607.

Sampson, R. J, Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing “neighbourhood effects”: social processes and new directions in research. *Annual Review of Sociology* 28:443-478.

Shertzer. A.. Walsh. R. P.. & Logan. J. R. (2016). Segregation and neighborhood change in northern cities: New historical GIS data from 1900–1930. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*. 49(4). 187-197.

Stanfors, M. (2003). *Education, Labor Force Participation and Changing Fertility Patterns. A Study of Women and Socioeconomic Change in Twentieth Century Sweden.* Stockholm: Almqvist & Wiksell International.

Szulkin. R.. & Jonsson. J. O. (2007). *Ethnic segregation and educational outcomes in Swedish comprehensive schools*. Stockholm University Linnaeus Center.

Van Leeuwen, M. H. D., & Maas, I. (2011). *HISCLASS. A Historical International Social Class Scheme.* Leuven: Leuven University Press.

Van Leeuwen, M. H. D., Maas, I., & Miles, A. (2002). *HISCO: Historical International Standard Classification of Occupations*. Leuven: Leuven University Press.

Vartanian. T. P.. & Houser. L. (2010). The effects of childhood neighborhood conditions on self-reports of adult health. *Journal of Health and Social Behavior*. 51(3). 291-306.

Wilson, W. J. (1987). *The Truly Disadvantaged. The Inner City, the Underclass, and Public Policy*. Chicago: The University of Chicago Press.

Wodtke, G. T., & Parbst, M. (2017). Neighborhoods, schools, and academic achievement: A formal mediation analysis of contextual effects on reading and mathematics abilities. *Demography* 54: 1653-1676.

Wodtke, G. T., Harding, D. J., & Elwert, F. (2011). Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high-school graduation. *American Sociological Review* 76(5): 713-736.

Östh. J.. Clark. W. A.. & Malmberg. B. (2015). Measuring the scale of segregation using k-nearest neighbor aggregates. *Geographical Analysis*. 47(1). 34-49.