UNIVERSITÀ
DEGLI STUDI
DI VERONA

UNIVERSITA' DEGLI STUDI DI VERONA

DIPARTIMENTO DI

Scienze Economiche

SCUOLA DI DOTTORATO DI

Scienze Giuridiche ed Economiche

DOTTORATO DI RICERCA IN

Economia e Management

CICLO/ANNO
29°/2013

Inequality of Opportunity and Space

SECS-P/01

Coordinatore: Prof. Roberto Ricciuti
Firma

Tutor: Prof. Eugenio Peluso
Firma

Dottorando: Dott. Umut Türk
Firma

Inequality of Opportunity and Space
Umut Türk
Doctoral Dissertation
Verona, June 2017

Acknowledgments

I would like to express my sincere gratitude to my advisor Prof. Eugenio Peluso for the continuous support throughout my PhD study and research, for his motivation and availability. His supervision helped me to develop a very raw idea to a doctoral dissertation. He encouraged me to participate international conferences and to build a professional network. Without his support, it would not be possible to conduct my research. Besides my advisor, I would like to thank all the professors and staff of the Economics department. I could not have imagined having a better environment for my PhD studies.

I would like to thank my thesis committee: Prof. Daniele Checchi and Prof. Matz Dahlberg for their insightful comments and the challenging questions which greatly improved the quality of this thesis. I also would like to acknowledge Prof. Claudio Zoli and Dr. Franceso Andreoli for taking time to read and comment on my papers.

My sincere thanks to my second supervisor, co-author, generous host of my visiting period in Sweden and most importantly a dear friend: Prof. John Östh. He taught me how important was to recognize the role of geography in any kind of scientific inquiry. John responded my every single idea with a great enthusiasm and provided necessary means to realize them. His "everything can be done" attitude helped me to find my way through difficulties I encountered in my research. His touch is easily detectable in each chapter. I also would like to thank everybody at the department of social and economic geography, Uppsala University. From the first day, they accepted me as a member of their family, I owe so much to those long chats during the lunch breaks and fika times.

Nobel laureate Orhan Pamuk's book "The new life" begins with the sentence "I read a book one day and my whole life was changed." I may just as well say: I met a group of colleagues in my department and my life was changed. My special thanks goes to Tania Masi, for her never-ending support, for long, high quality discussions not only about Economics but also about novels, poems and life in general. For dishes, she first learned herself with her academic attention and then prepared for me, and especially for generously offering me a second home with doors wide open whenever I needed. My special thanks to Luca Alfieri, for being always there with his precious ideas and support, to Giovanni Covi for his hospitality, for nights out in Padua and for his unreplaceable friendship. Last but not the least my sincere thanks to Ludovico Bullini Orlandi, Riccardo Semeraro, Alain Julio Mbebi, Yasna Cortes and Silvia Blasi. This group has been a source of friendship as well as collaboration.

Finally, I would like to thank my family. My mother and father for their unconditional support in all my pursuits and love. I stated to every class, every project with their encouragement. They took a great interest whatever I was doing. From Km's away, they wondered why my code wasn't compiling, stayed up late with me when I had to study or write my papers. They shared every single emotion of mine, lessening the burden out of my shoulders. I am especially grateful for my sisters Utku and Duygu and my lovely nephews and niece: Mert, Çağan and Duru who long thought that their uncle lived in a country called Skype. They cherished with me every great moment, supported me spiritually throughout writing this thesis.

To My Family

Contents

1 Introduction **6**
1.1 References 20

2 Socio-Economic Determinants of Student Mobility and Inequality of Access to Higher Education in Italy **25**
2.1 Introduction 27
2.2 Theoretical Framework 30
2.2.1 A Spatial Interaction Model of Student Mobility 30
2.2.2 Adopted Accessibility Index 33
2.3 Data and Variables 34
2.3.1 Types 34
2.3.2 Distance 35
2.3.3 University Attractiveness 36
2.3.4 Interactions 37
2.4 Empirical Method and Findings 37
2.5 Conclusions 42
2.6 References 44

3 Inequality of Opportunity in Sweden:A Spatial Perspective **57**
3.1 Introduction 59
3.2 Previous Work 62
3.3 Data 65
3.4 Analytic Framework 68
3.5 Findings 71
3.6 Concluding Remarks 78
3.7 References 81

4 A Gini Measure of Spatial Segregation by Income **96**
4.1 Introduction 98
4.2 The Gini indices of Segregation 100
4.3 The GSS index 103
4.4 Empirical Application 104
4.5 Conclusions 110
4.6 References 112

List of Figures

1.1 Structure of the Method 23
1.2 Data 24

2.1 Access Maps of First 5 Groups 53
2.2 Access Maps of Last 5 Groups 54

3.1 Neighbourhood Maps Whole Stockholm Area 94
3.2 Decomposition 95

4.1 GSS values in years 1994-2004-2014 for varying scales 118
4.2 Maps with GSS values per Municipality 119
4.3 Deviation Maps for GSS values (Year 2014-Year 1994) 120
4.4 Spearman Correlations 1994-2014 121

List of Tables

2.1 Variables in analysis 49
2.2 Results of Poisson Regression First 5 Groups 50
2.3 Results of Poisson Regression Last 5 Groups 51
2.4 Decomposition of Inequality in Access 52

3.1 Moran's I Tests 86
3.2 Variables 87
3.3 K-neighbours 88
3.4 Multilevel Model: Log of Grades 89
3.5 Multilevel Model:Log of Disposable Income 90
3.6 Inequality Decomposition 91
3.7 Inequality Decomposition Female Population 92
3.8 Inequality Decomposition Male Population 93

4.1 Gini Indices of Segregation 116
4.2 GSS for different k values 117
4.3 GSS for different r values 117

Chapter 1

Introduction

1 Introduction

The uneven distribution of opportunities between socio-economic groups has been a fundamental concern of governments, policy makers and of the scholars seeking proper policies and strategies to tackle it. Despite extensive research, efforts to identify and tackle inequalities are nested in two interrelated but somehow separate domains. One genre of literature focuses on inequalities in an "aspatial" nature that exists in the form of inequality of income, wealth, consumption and also in the form of social inequality: due to gender, parental background, race and ethnicity. The other body of literature studies inequalities in a "spatial" form, which are generalized as inequalities due to locations. Various examples are found typically in urban economics: inequality in access to education, health services, public goods and to a decent employment. Consequently, often there is no explicit correspondence between the spatial and aspatial measures put forth and employed to empirically implement inequality investigations. This thesis aims to bridge this gap by employing methods and theories from both approaches and ultimately present comprehensive analysis of equality.

A considerable amount of effort has been put forward to identify the way in which inequalities prevail and the notions of equality, fairness and justice. The traditional study of equality concerns with the evenness in the distribution of outcomes across populations (Katz et al., 1999). On the other hand a relatively recent theory building on the equality of opportunity (EOp) arguments puts individual responsibility forefront when investigating inequalities in life chances. As a responsibility sensitive egalitarian theory, this approach identifies the causes of disparities in opportunity distributions by decomposing inequalities into illegitimate sources, for those individuals should not be held responsible, and legitimate sources ,those for individuals are deemed responsible. This thesis aims to single out the spatial sources of inequalities observed in opportunity distributions.

The roots of the equality of opportunity theory go back to Rawls (1971) suggesting that the political decisions to be taken behind *a veil of ignorance* where decision makers know nothing about their particular talents, tastes, social class and so on. With this thought experiment the egalitarian theory first shifted from the mere concern of equality in outcomes to a philosophical view that inquired into the equality of opportunities. The change of emphasis was a consequence of the developments in political philosophy, inspired by Rawls (and Sen (1980)), systematized by Dworkin (1981) (where he suggested a market behind Rawl's the veil of ignorance) and subsequently modified by Arneson (1989) and Cohen (1990) based upon the idea of keeping people responsible for their choices and preferences (Roemer & Trannoy, 2013). The theoretical framework of EOp was fist developed by Roemer (1993, 1998) and several contributions have taken place as the concept has become popular over years (see, Dirk Van de gaer (1993), Walter Bossert (1995, 1997),Vito Peragine (2004)). The formal application of the theory analyses the conditions of *leveling the playing field*, which holds in a society when the life chances of individuals depend solely on their own *effort* purged of exogenous factors defined as *circumstances* such as gender, race, ethnicity, family background (Dardanoni et al., 2006).

The first two chapters of this thesis aim to the bridge theory and methods of EOp literature to Spatial Accessibility and Neighbourhood Effects studies respectively and the last chapter studies the Spatial Segregation by Income from equal opportunities perspective.

Paper 1: The Socio-Economic Determinants of Student Mobility and Inequality of Access to Higher Education in Italy: The first chapter of this thesis aims to contribute to the methods of both EOp and spatial accessibility studies and by an empirical application narrows an existing gap in the analyses of Italian higher education system. The paper measures the spatial inequality in access to higher education institutions in Italy with a particular attention to the socio-economic sources of inequality. Accessibility is often computed by

Hansel-like gravity indices and disparities in the distribution of access levels between locations is interpreted as spatial inequality. Since the elements in the formulation of the index are purely geographical, the idiosyncratic characteristics of individuals who reside in areas are often neglected. This is partially an expected lack in this particular literature as public goods are, by definition, non-excludable and non-rivalrous in consumption (Kolstad, 2003). However, different socio-economic groups face different costs of access (both material and social) for the same set of opportunities in space. In other words, even though the spatial availability of opportunity in question is increased, there may remain an under-representation from disadvantaged backgrounds. Accordingly, this paper asks the following questions from two opposing directions: provided that the universities are equally distributed in geographical areas, do students with different socio-economic background have the same degree of access ? and provided that the students share identical socio-economic background, do they have the same degree of access from parental locations?

Broadening access to higher education is a goal that can produce positive outcomes both for the individuals concerned and for wider society. In line with this view, with one of the lowest graduation rates in Europe, the supply of HE was expanded drastically in the period 1990-98. Large universities set up new sites, new types of faculties and 9 new universities were founded. However, the interventions were arbitrary, operated without any field examination of accessibility or demand (Bratti et al., 2008). After these reforms the Italian HE has shown an increasing participation (still below the OECD average(OECD,2011)), yet as it was in the pre-reform period, to what extent the accessibility was widened is still unknown. Therefore, measuring accessibility with the consideration of the gap between different socio-economic groups addresses a delayed exercise.

As shown in Fig.1, to address this exercise, the theory of EOp is linked to a Hansen-like accessibility index via a spatial interaction model(SIM). Derived from the gravity logic, SIMs suggest that the interaction between any two units must be directly proportional to the masses of origin and destination whereas inversely related to the distance between them (Sen & Smith, 2012). The distance function in SIMs is where the accessibility index and EOp are linked. Roemer (1998) defines social types consisting of the individuals who share exposure to the same circumstances. The set of individual circumstances observed allows the specification of these social types in the data. Using a survey data (Inserimento professionale dei laureati, 2011) containing 14,000 male and 17,400 female students graduated in 2007 and the data from MIUR (2003-2004-2005), this paper first creates types of students on the basis of Roemerian approach.

Based on the information available in the dataset, the paper assumes that the geographic variation in access to HE could be driven by two sources: gender and parental background. Accordingly, 12 such groups are generated. Then SIM has been utilized to model the flows of students from parental locations (origin) to universities (destination). SIMs are calibrated separately for 12 groups to capture heterogeneities in response to the spatial distribution of opportunities. The model controls for factors that communicate the quality of universities. It does so by including university fixed effects when calibrating the actual flows of students. This approach identifies quality effects from comparisons between the intensity of interactions, effectively asking whether the characteristics of universities attract different types of students in varying levels. The results indicate that the poor family backgrounds are insensitive to the university-quality effects employed and only when the family background becomes better, the university preferences are revealed. Moreover, in line with initial hypothesis that different types of students must be facing different costs of mobility, varying distance-decay parameters are observed for different types of students. Using this heterogeneity, respective distance-decay parameters are imported to the accessibility index computing access distinctively for each type. The index suggests that the access at location i is a sum of the number of seats offered by university at location j, and discounted by the distance function that is empirically derived by SIMs. Results from accessibility index demonstrate that the family background proxies are relevant especially for female students. From worse to the best family origin, the access increases on average 101% and heterogeneity in access across genders disappears for students with highly educated parents.

The critical identification assumption underlying the approach of this paper is that the access to HE must vary not only by the spatial distribution of universities but also by gender and parental origin. To quantify this assumption, the computed inequality in access is decomposed first by suppressing the inequality due to geography, therefore leaving inequality only due to socio-economic background and then by suppressing inequality due to socio-economic background thus the remaining inequality is only attributable to the spatial distribution of universities. Results show that 5% of the disparity in access is due to family origin and gender when using the first approach and 7% when using the second approach.

Overall, these results suggest that when studying spatial accessibility focusing only on equality in outcomes leaves local variation due to *aspatial* factors outside of the picture and efforts to decrease inequality in access remain incomplete for conducting meaningful policy analysis. In particular to Italian data, this paper finds that the physical and social distance between different types of students persist. On the other hand, the residential decisions taken by parents are clearly exogenous factors that influence the life chances of students. From the EOp perspective this paper is the first attempt to define parental location as a circumstance.

Paper 2:Inequality of Opportunity in Sweden:A Spatial Perspective: The second chapter of the thesis links theory and methods from equality of opportunity literature with the neighbourhood effects literature. All conceptions of equal opportunity draw on some distinction between fair and unfair sources of inequality. In spite of a vast quantity of empirical applications, the spatial dimensions of the issue have been overlooked. The geography has taken place in the analysis of EOp, however at very large sizes. For instance, some papers included large administrative units of residence as circumstances (Singh, 2012; Cogneau & Mesplé-Somps, 2008), some partitioned study area into few macro regions (Peragine & Serlenga, 2008; Checchi et al., 2010). This paper analyses geography at its smaller scales when addressing following two complementary questions: Do parental neighbourhoods exert effects that persist? and if so, to what extent these effects contribute to inequality of opportunity in outcomes?

A large body of literature investigates neighbourhood effects and provides solid evidence on their importance in shaping individuals' life chances those for child outcomes (Leventhal & Brooks-Gunn, 2000), educational attainment (Garner & Raudenbush, 1991), drop outs (Crane, 1991), reading, math achievements and higher education participation (Andersson & Malmberg, 2015). All these findings are relevant to study of equality of opportunities. Being exogenous factors to children, parental neighbourhoods are obvious candidates to be defined as circumstances. Therefore, for the analysis of educational EOp, we included several neighbourhood statistics as the sources of unfair inequality among students during exposure. In addition, we also investigated whether these effects persisted, remaining influential for the adult outcomes of the same population after exposure is ceased.

The literature studying the duration of neighbourhood effects offer mixed findings. Some find declining effects on earnings and educational attainment as years pass (Raaum et al., 2006, see), some finds no evidence of such an impact for economic conditions and educational attainment even for those who were exposed to a better environment in early ages (see Ludwig et al., 2013, for instance), Chetty et al. (2015) show that each additional year spent in better neighbourhoods increases the likelihood of college attendance and of attaining higher earnings in adulthood. The paper argues that an important methodological issue has to do with the boundaries of neighbourhoods when studying the durability of neighbourhood effect. So far the main approach has been to measure neighbourhood context using aggregate values for administratively defined areas such as municipalities, counties and parishes. This implies that neighbourhood effect studies, including the ones cited above, have overlooked the variability at residential areas by defining such aggregate geographies as neighbourhoods. We argue that neighbourhoods must be treated as personal experiences implying that the methods to quantify them must construct bespoke individual neighbourhoods.

As shown on Fig.2, using the Place longitudinal data keeping the track of whole Swedish residents since the year 1990, we follow the life span of whole 1985 cohort. We investigate both the inequality in educational attainment and in income by multilevel models. Multilevel models allow using municipality random effects, which in return overcomes most of the spatial autocorrelation problem found with classic OLS approach. Next, based on a parametric IOp measure, we implement inequality decompositions.

In the first step of our analysis, we define compulsory examination grades as a linear function of variables that are informative of family background and typically used in EOp studies and other inherited circumstances with variables communicating several neighbourhood statistics in 2001 the year in which this cohort takes a compulsory examination. Instead of using predetermined administrative units, we construct individual neighbourhoods based on a population count method for each individual residence. As a form of scalable egocentric neighbour, this technique departs from an individual geo-location and begins counting towards every direction until the nearest k number of population has been reached. Neighbourhood statistics are defined as the ratio of the interested population to the total counted population. With this approach, we quantify the probability of interacting with a given population group. The method does not require the use of bounded geographical units, hence provides an efficient, comparable and robust definition of a place. The results of the multilevel regression show that parental neighbourhoods matter for children's educational attainment. Decomposing educational inequality indicates that 42.62% of educational inequality is attributable to circumstances and the effects from parental neighbourhoods weight as much as 36.94% of overall circumstances.

In the second step, we follow the same cohort up to their employable age. In the year 2010 we create own-neighbourhood statistics based on the same techniques and we also bring parental neighbourhood statistics from 2001 into the analysis of earnings' distribution. As years pass the responsibility profile changes. Since in adulthood individuals are free to choose their residences we deem them responsible for own neighbourhoods. Multilevel regression demonstrates that parental neighbourhoods matter also for children's long-term outcomes even years after the exposure. Moreover, according to our estimates 8.05% of total inequality is due to circumstances where parental neighbourhoods explain 16.66% of total circumstances and own-neighbourhood explains 1.95% of total effort.

Finally, we find analogous results when analyzing inequality in outcomes for genders separately. Though IOp measures are similar for males and females, there are some differences in response to neighbourhoods by gender. For instance, during the adulthood parental neighbourhood explains a higher portion of variation in income for females than men. Whereas, an opposite situation yields as we look at the variation in educational attainment, that for the male population parental neighbourhoods are more influential while living in them.

This paper tests the idea that parental neighbourhoods play an important role in life chances of children both immediate effects on education and also long-term effects on earnings. Overall, the findings presented in this paper suggest that a multidisciplinary methodology that links neighbourhood studies to EOp literature by the use of bespoke neighbours approach would not only provide neighbourhood effect studies with a new lease of life, but would also take EOp studies to a new domain for a more comprehensive investigation of fairness in the distribution of opportunities. The neighbourhoods statistics created in this paper are proxies for the concentration of some attribute (see Fig.2) in the areas of residence. As a consequence in addition to classic EOp policy recommendations, to decrease inequality in opportunities, an effective policy must target the residential environments of less advantaged groups such as minorities and single headed families in particular to Swedish data. Therefore, this paper is an initial step towards a study of EOp implications of residential segregation. The residents of cities are normally segregated along minority status, ethnicity or along household characteristics. For instance large families may prefer to live in similar neighbourhoods as they need larger apartments, and since household income is dived among larger family members, they can afford accommodation only in certain localities. Then a study that would like to cover all these situations simultaneously must investigate the residential segregation by income.

Paper 3 A Gini Index of Spatial Segregation by Income: The third chapter of the thesis proposes a Gini index of spatial segregation by income (GSS). Residential segregation may cause disparities in access to opportunities. Segregated areas show differential levels of accessibility and spatial match between job and housing. The residents face varying costs of access to public goods/services, the range and the quality of local amenities differ among areas of residence. This is an undesirable situation for the society as a whole since the combination of poverty, adverse neighbourhood spillovers, and isolation from opportunities all make it difficult for an individual to perform well in school, and in the labor market. For this reason quantifying the degree of residential segregation is of interest to studies of equal opportunities.

Existing segregation measures found in literature are mainly developed for dichotomous cases such as race, ethnicity, the gender gap in occupations. Although residents sort also by income across areas as choices over bundles of local public goods (Tiebout, 1956) and often segregation is related to affordability of the areas of residence, the measurement of residential segregation by income has not received much attention in the literature. People may sort into neighbourhoods because they prefer to live with similar people in terms of educational level, ethnicity or for the quality of amenities provided. However, the decision of residence is constrained by the budget (also willingness to pay) of the decision maker. Therefore, a locality may show a residential mixture by dichotomous variables but it can be highly segregated in relation to the income level of residents .

From a measurement perspective, segregation by nominal groups amounts to group individuals for membership to a given category. A similar approach is widely applied for the measurement of segregation by income. Typically individuals in study area is divided into two categories for being under and above to a given level of income (Massey et al., 2003; Abramson et al., 1995). Hence, the first issue related to previous measures of economic segregation is that using this approach discards considerable amount of information in the underlying parameters of the continuous distribution (Jargowsky, 1996). Second issue is that nearly all existing indices are aspatial in nature, implying that the distribution of individuals in space is not taken into account. The spatial ones (see Reardon et al., 2009; Reardon & Bischoff, 2011; Dawkins, 2007, for example) are rather difficult to compute, and nearly all use some administratively defined area for the unit of analysis. This last point causes further issues that are associated to the modifiable areal unit problem (MAUP). The MAUP occurs with both the scale problem that the same data portrays different spatial patterns for its varying levels of aggregates and with the zoning problem that altering the grouping schemes produce different results even if the units are of the same scale (Openshaw, 1984; Wong, 2004). In this paper we aim to address issues stated above with a particular attention to the scale problem. We test the proposed index with an empirical application to Swedish data, where we show how the definition and the scale of the neighbourhood influence the measurement of economic segregation.

We define economic segregation as a ratio of two Gini indices, at the numerator we replace individuals' incomes by average incomes in neighbourhoods, and we normalize the resulting inequality between neighbourhoods with the individual level Gini at the macro area where neighbourhoods are nested as follows:

$$GSS(y,n) = \frac{\frac{1}{N^2\mu_s}\sum_i\sum_j |\mu_{is} - \mu_{js}|}{\frac{1}{N^2\mu}\sum_i\sum_j |y_i - y_j|} = \frac{I_B}{I_G} \tag{1}$$

where $\mu_{is} = \frac{\sum_{j\in s} y_j}{n_{is}}$

Population N individuals y_i is the income of individual i, μ_{is} is the average income in individual i's neighbourhood. Unlike other measures of income segregation found in literature, the index makes use of individualized-neighbourhoods, where the shape (s) of the neighbourhood varies for the definition chosen:

- radii-based neighbourhoods: average income around an individual location within a radius (r)
- population count-based neighbourhoods: average income around an individual location among k nearest neighbours (knn).

The size (n_{is}) can be set to meet various scales of geography by configuring k or r. The GSS is a measure of income homogeneity/diversity at individualized-neighbourhoods, It takes a minimum of 0 (no segregation) in two case scenarios: if the numerator is zero thus the between spatial inequality is zero or when the size of the neighbourhood is equal to that of the whole study area: $n_{is} = N$. It takes the value 1 (perfect segregation) if the distribution of individualized-neighbourhood average incomes is identical to that of individual incomes thus when all persons live only with others who have identical incomes or when the size of the neighbourhood $n_{is} = 1$, that every neighbourhood consists of one individual only.

We test the GSS with Swedish data for both approaches to individualized neighbourhoods and for varying scales of geography. Results show that the segregation has increased from 1994 to 2014. Furthermore, the findings illustrate a clear difference with radii and *knn* approaches to neighbourhood. The reason for this is that the radii-approach uses the pure geography as the basis, computing the area constituted within a radius without a direct reference to the number of people living in those areas. This is a disadvantage for comparative analyses, since population density may change over time and among different areas. Even though the GSS is weighted for the number of population covered in each radii-neighbourhood still an obvious difference from *knn* definition of neighbourhood yields. *Knn* approach instead computed the likelihood of interaction between individuals and disregards distances between them. This is again a disadvantaged for the analyses of sparsely populated areas since the *kth* neighbour may live far away. To solve this issue we found an intermediate way by employing spatial weights matrix based on the distance between neighbours. As distances between individuals increase, neighbour's contribution to mean incomes decreases. This approach communicates both the geography similar to radii and population count as *knn*.

Next, we assess the robustness of our results by conducting spearman correlation analysis between computed GSS values and several attributes of Swedish municipalities. For all scales of neighbourhood size, we find positive and highly significant correlations with the ratio of high and low educated residents to the total population of municipalities and with the election participation rates by municipality. These two attributes are informative for the sorting behavior of population and their *choices*. Additionally, we find positive and significant correlation between GSS values and employment growth, whereas negative and significant correlation with transfers among municipalities. The latter two are considered as exogenous factors to residents but are correlated to residential segregation by income.

Overall, based on the evidence from Swedish data, we argue that the GSS is able to render the patterns of economic segregation in a highly accurate degree. The index is not subject to robustness issues associated with MAUP and checkboard phenomenon and as a ratio of two Gini indices, the index has the advantage of preserving desirable properties of the Gini. It respects to the Pigou-Danton principles of transfers, less sensitive to outliers, deviations from normality and finally is suitable for the segregation measure of continuous variables.

References

Abramson, A. J., Tobin, M. S., & VanderGoot, M. R. (1995). The changing geography of metropolitan opportunity: The segregation of the poor in us metropolitan areas, 1970 to 1990. *Housing Policy Debate*, *6*(1), 45–72.

Andersson, E. K., & Malmberg, B. (2015). Contextual effects on educational attainment in individualised, scalable neighbourhoods: Differences across gender and social class. *Urban Studies*, *52*(12), 2117–2133.

Arneson, R. J. (1989). Equality and equal opportunity for welfare. *Philosophical studies*, *56*(1), 77–93.

Bratti, M., Checchi, D., & De Blasio, G. (2008). Does the expansion of higher education increase the equality of educational opportunities? evidence from italy. *Labour*, *22*(s1), 53–88.

Checchi, D., Peragine, V., & Serlenga, L. (2010). Fair and unfair income inequalities in europe.

Chetty, R., Hendren, N., & Katz, L. F. (2015). *The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment* (Tech. Rep.). National Bureau of Economic Research.

Cogneau, D., & Mesplé-Somps, S. (2008). Inequality of opportunity for income in five countries of africa. *Research on Economic Inequality*, *16*, 99–128.

Cohen, G. A. (1990). Equality of what? on welfare, goods and capabilities. *Recherches Économiques de Louvain/Louvain Economic Review*, *56*(3-4), 357–382.

Crane, J. (1991). The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American journal of Sociology*, 1226–1259.

Dardanoni, V., Fields, G. S., Roemer, J. E., & Puerta, M. L. S. (2006). How demanding should equality of opportunity be, and how much have we achieved? *Mobility and Inequality: Frontiers of research in Sociology and Economics*, 59–82.

Dawkins, C. J. (2007). Space and the measurement of income segregation. *Journal of Regional Science*, *47*(2), 255–272.

Dworkin, R. (1981). What is equality? part 1: Equality of welfare. *Philosophy & public affairs*, 185–246.

Garner, C. L., & Raudenbush, S. W. (1991). Neighborhood effects on educational attainment: A multilevel analysis. *Sociology of education*, 251–262.

Jargowsky, P. A. (1996). Take the money and run: Economic segregation in us metropolitan areas. *American sociological review*, 984–998.

Katz, L. F., et al. (1999). Changes in the wage structure and earnings inequality. *Handbook of labor economics*, *3*, 1463–1555.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes. *Psychological bulletin*, *126*(2), 309.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2013). *Long-term neighborhood effects on low-income families: Evidence from moving to opportunity* (Tech. Rep.). National Bureau of Economic Research.

Massey, D. S., Fischer, M. J., Dickens, W. T., & Levy, F. (2003). The geography of inequality in the united states, 1950-2000 [with comments]. *Brookings-Wharton papers on urban affairs*, 1–40.

Openshaw, S. (1984). The modifiable areal unit problem..

Peragine, V., & Serlenga, L. (2008). Higher education and equality of opportunity in italy. *Research on Economic inequality*, *16*, 67–97.

Raaum, O., Salvanes, K. G., & Sørensen, E. Ø. (2006). The neighbourhood is not what it used to be*. *The Economic Journal*, *116*(508), 200–222.

Rawls, J. (1971). *A theory of iustice.* Harvard University Press, Cambridge.

Reardon, S. F., & Bischoff, K. (2011). Income inequality and income segregation 1. *American Journal of Sociology*, *116*(4), 1092–1153.

Reardon, S. F., Farrell, C. R., Matthews, S. A., OSullivan, D., Bischoff, K., & Firebaugh, G. (2009). Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990–2000. *Social Science Research*, *38*(1), 55–70.

Roemer, J. E. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs*, 146–166.

Roemer, J. E. (1998). *Theories of distributive justice.* Harvard University Press.

Roemer, J. E., & Trannoy, A. (2013). Equality of opportunity.

Sen, A. (1980). *Equality of what?* (Vol. 1). na.

Sen, A., & Smith, T. (2012). *Gravity models of spatial interaction behavior.* Springer Science & Business Media.

Singh, A. (2012). Inequality of opportunity in earnings and consumption expenditure: The case of indian men. *Review of Income and Wealth*, *58*(1), 79–106.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of political economy*, *64*(5), 416–424.

Wong, D. W. (2004). The modifiable areal unit problem (maup). In *Worldminds: Geographical perspectives on 100 problems* (pp. 571–575). Springer.

Tables and Figures

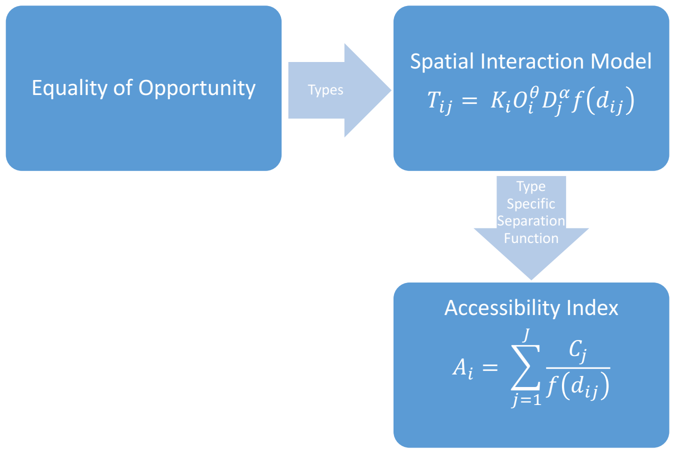

Figure: Structure of the Method

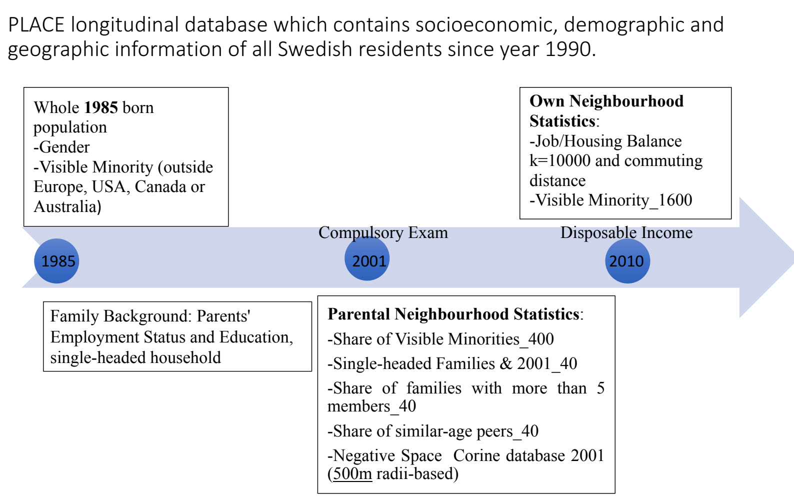

Figure 1: DATA

Chapter 2

Socio-Economic Determinants of Student Mobility and Inequality of Access to Higher Education in Italy

Socio-Economic Determinants of Student Mobility and Inequality of Access to Higher Education in Italy*

Umut TÜRK

Abstract

This paper introduces a modified version of the Hansen-gravity model as a framework to estimate the accessibility of higher education (HE) institutions in Italy from equal opportunities perspective. The key assumption underlying gravity models is that accessibility decreases with spatial distance from opportunities. The paper extends the gravity equation to allow for the inclusion of socio-economic factors influencing the access to HE. The findings reveal differences in response to quality and to other institutional characteristics by parental background and gender. Finally, decomposition of overall inequality into spatial and aspatial components reveals both the physical and social distance between groups of students seeking higher education opportunities in the country.

Keywords Spatial Interaction, Higher Education Accessibility, Gravity Model, Equality of Opportunity

*This paper has benefited from the insightful comments and suggestions of the participants of 9th GRASS Workshop in Lucca, Italy; RSAI/ERSA Workshop on Regional Science and Urban Economics Spatial Perspective of Human Capital in Barcelona, Spain; Joint meeting NECTAR Cluster 6 Accessibility/Regional Science Academy in Paris, France.

1 Introduction

An intuitive way to increase spatial accessibility is to decentralize the service in question. This was the strategy implemented by the Italian authorities in the period 1990-1998. With one of the lowest participation and graduation rates in Europe, the supply of higher education (HE) was expanded drastically. The reforms required larger universities to set up new types of faculties and 9 new higher education institutions were established as a result of the decentralization process (MIUR). However these reforms took place without any field examination of accessibility or demand (Bratti et al., 2008). More than a decade later, there is no explicit measure of spatial accessibility of the universities in the country.

In a system granting free access to HE for every potential candidate, the foremost aim of policy makers is to guarantee full accessibility to the service irrespective of the location of residence. Previous research has focused on measuring accessibility through an examination of the match between the locational distribution of facilities or services and the locational distribution of residents (Talen & Anselin, 1998). In this framework, the spatial distance between residence of origin and the location where opportunities are located is regarded as an important factor determining the spatial accessibility. The underlying idea is that people from more isolated locations face larger costs to access to opportunities, with costs growing with spatial distance.

Since the residential location of students prior to HE enrolment is detemined by parents, inequalities in access to HE across students' locations of origin should be regarded as unfair. Modern theory of inequality, building on equality of opportunity (EOp) arguments, suggests that the differences in outcomes due to factors that are beyond individual responsibility are unfair and should be compensated by society (Dardanoni et al., 2006). Reducing geographical disparities in accessibility can be seen, in fact, as a way of *leveling the playing field* (Roemer, 1998) and providing equal opportunities to benefit from HE irrespective of the place of origin. The geographical location can be an unfair source of accessibility on the large scale: a student living in an urban area where HE services are supplied faces smaller costs of transportation, lower opportunity costs in commuting and no housing costs, compared to students living in the countryside who need to commute or move to benefit from HE services. However, focusing only on geography may leave the influence of socio-economic factors, in relation to gender, experiences at home and parental background, unexplored.

The paper argues the existence of a gradient of economic circumstances of origin on distance elasticity. Although distance matters in explaining accessibility, there are other variables that determine differences in costs of movement, correlated with distance which, at the same time, might influence the distribution of accessibility. This paper tries to single out the contribution of spatial distance and economic circumstances on inequality in accessibility to HE by using a multidisciplinary approach, where the problem of disparities in spatial access is redefined on the basis of both the physical distance from universities and the social distance between student groups that generates an additional inequality in access within the same location. The paper looks at the variability in access both when focusing on comparisons of people located at different origin points from HE, but all sharing the same family background (highlighting the share of inequality due to spatial distribution of HE institutions in the country) and when comparing people located at the same origin points but with differing backgrounds of family origin, which is taken as a proxy of the ability of families to cover costs of displacement and, if possible, to compensate for distance from the location of origin. The latter shows the share of inequality in access due to the socio-economic background of students.

In an empirical application, the paper sequentially employs a model and an index to measure overall inequality in access, which is then decomposed into its geographical and socio-economic components ; first a spatial interaction model (SIM) is used to disentangle the migration dynamics of different student groups. Being flexible and simple enough, these models enable the investigation flows between origins and destinations (Sen & Smith, 2012). Actual flows of commodities, information, emails, phone calls, money and of people along with any other sort of movements are likewise applicable to SIMs (see Haynes & Fotheringham, 1984; Sen & Smith, 2012, for reviews). In the present application, student flows from parental residents to universities are defined as interactions between localities and, to account for socio-economic factors in place, the total observed flows of students are partitioned into subgroups each representing a different *type*. It is a common practice for EOp studies to partition the population according to exogenous factors, which are assumed to be beyond people's control (see for instance Checchi & Peragine, 2010; Ferreira & Gignoux, 2011) and resulting subgroups are defined as types (Roemer, 1998). For the second step, the parameters that are distinctively calibrated for each type by the SIMs are imported to a Hansen (1959)-like index to measure potential accessibility for 110 Italian provinces (NUTS3 level regions). Finally the inequality in accessibility among provinces is decomposed as follows: the access score in each province is replaced with its average access score across socio-economic groups hence only variation is allowed to be due to the geographical distribution of universities. Then the access scores computed for each socio-economic group is replaced with its average access score across provinces hence remaining variation is allowed to be due to socio-economic backgrounds. This operation enables investigating the relative contributions of spatial and aspatial factors.

The paper contributes to the literature by extending classical spatial accessibility analysis to incorporate the socio-economic circumstances of students in a spatial accessibility measure for Italian HE institutions. This practice goes beyond the mere concern of inequalities on outcomes. For the spatial accessibility analysis this means that the inquiry may shift from "spatial accessibility where?" to "spatial accessibility where and for whom?". It also contributes to the EOp literature by showing how the spatial dimensions of the theory can be incorporated into models that rely solely on geography. Finally, the findings in this paper provide highly detailed information for policy makers regarding which groups of students to target and specifically in which locations the assistance is needed most.

The remainder of the paper is organized as follows: the second section introduces the model and the accessibility index adopted, the third section sets out the data and variables, the fourth section shows empirical method for calibration and findings where inequality in access is decomposed into within and between components. Finally the conclusions and possible policy implications are given in the fifth section.

2 Theoretical Framework

This section presents the model adopted for student flows and the potential accessibility index. The link between these two builds on the distance parameter assumed to reflect both physical and social costs in migrating or commuting to destination universities. The response to distance is expected to be conditional on the socio-economic background of students.

2.1 A Spatial Interaction Model of Student Mobility

Spatial interaction models are used to predict the size of spatial flows between origins and destinations in areas of interest. They have been used mainly for transportation and environmental planning, then developed further for a variety of applications where a movement and/or interaction takes place. Recently, most applications relate to health system planning, decisions concerning hospital locations, the analysis of interaction between patients and physicians as well as labour studies such as job accessibility and an investigation of the daily commute to work (Mayhew et al., 1986; R. M. Wilson & Gibberd, 1990;

Reggiani et al., 2011).

Particularly with regard to SIMs applied to HE choice, Sa et al. (2004) studied the demand for HE in the Netherlands, given the attractiveness and accessibility of universities. Alm & Winters (2009) correlated the distance from parental residence to state HE institutions with tuition, financial aid and school quality as institution fixed characteristics and found a varying deterrence effect of distance in relation to these characteristics. Cooke & Boyle (2011) included several origin and destination attributes to SIMs, including the number of high school graduates in origins, employment growth both in origins and destinations and the relative quality of amenities. Singleton et al. (2012) integrated SIMs with geodemographic analysis, and looked at both socio-spatial conditions in the neighbourhood and the attractiveness of destinations. For the Italian data, Dotti et al. (2013) investigated the role of universities in attracting successful students to certain regions and to settle down there after graduation. These studies estimate the distance elasticity of university choice given the attributes of origins/destinations and some also include university characteristics. However they do not incorporate the socio-economic profile of students in the analysis. The present paper examines the role of socio-economic characteristics of students in distance elasticity. Furthermore, based on the distance elasticity values, the paper transforms SIMs into an explicit measure of accessibility.

SIMs can incorporate a range of origin and destination constraints and take a number of forms according to this constraint structure. The following formula is a production-constrained form of SIMs that suggests that the interaction between any two units must be directly proportional to the masses of origin and destination and inversely related to the distance between them. The basic assumption is that a positive interaction between each pair of locations exists [1].

[1](see A. S. Fotheringham & Webber, 1980; A. Fotheringham & O'Kelly, 1989; Sen & Smith, 2012, for reviews)

$$T_{ij} = K_i O_i^{\theta} D_j^{\alpha} f(d_{ij}) \tag{1}$$

T_{ij} refers to student flow from their residence i to university j

O_i origin dummies for 110 Italian provinces (NUTS3 level regions)

D_j total number of students reaching university j

$K_i = [\sum_j D_j^{\alpha} f(d_{ij})]^{-1}$ is the balancing factor ensuring that the marginal total constraint $\sum_j T_{ij} = O_i$ is satisfied. [2]

$f(d_{ij}) = d_{ij}^{-\beta}$ where d_{ij} is the Euclidean distance between city i and university j

For this application, the model is extended to include several university fixed characteristics and interactions with distance. Finally the following model is obtained:

$$T_{ij} = K_i O_i^{\theta} D_j^{\alpha} {S_j}^{\gamma} {L_j}^{\eta} \exp(-\beta ln(d_{ij}) + \mu\delta_{ij} + \sum_l \lambda_l ln(d_{ij}) U_{jl}) \tag{2}$$

where S_j and L_j are two variables accounting for the attractiveness of a destination. Following the previous studies (see for example Lowe & Sen, 1996; Gitlesen & Thorsen, 2000; McArthur et al., 2011) a Kronecker delta is added to the model as follows:

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

[2] With K_i the model becomes production-constrained. The choice of this model is justified by the fact that most of the programmes are provided in an open-access fashion in Italy. Therefore, theoretically, students are free to choose any destination desired hence the model is not constained by destination (not attraction constrained) but to make sure that the number of trips produced by an origin do not exceed the number of residents, the model is constrained from production side. For the formal development see A. G. Wilson (1971)

The common interpretation of μ is that it reflects the benefit of residing and studying in the same city, or a start-up cost in case i and j are not in the same province. Furthermore $f(d_{ij})$ interacts with several destination characteristics U_{jl} where l is the number of interaction terms and λ_l is the distance elasticities given the institutional characteristics.

2.2 Adopted Accessibility Index

In this paper, the accessibility concept is interpreted as the potential availability of HE given the spatial distribution of institutions in the country. The roots of the index go back to Hansen (1959) when he first proposed the following gravity model of accessibility:

$$A_i = \sum_{j=1}^{J} S_j d_{ij}^{-\beta}$$

where A_i is a measure of accessibility , S_j is the number of opportunities at the destination and d_{ij} is the distance between an origin and a destination. A similar accessibility index can be constructed as follows:

$$A_i = \sum_{j=1} \frac{C_j d_{ij}^{\hat{\beta}}}{\delta_{ij}} \tag{3}$$

where

$$\delta_{ij} = \begin{cases} exp(\hat{\mu}) & i \neq j \\ 1 & \text{otherwise} \end{cases}$$

$\hat{\mu}$ and $\hat{\beta}$ are the two parameters that channel (2) to (3) and are calibrated beforehand by the production-constrained SIM (2). C_j is the total number of places offered by each institution. Additionally, the index discounts accessibility when i and j are not located in the same province by δ_{ij}.

3 Data and Variables

Table 1 shows the variables used in the analyses carried out in this paper. The data is extracted from a data survey (*Inserimento professionale dei laureati*, 2011) including 14,000 male and 17,400 female graduates in 2007 and data from MIUR (Ministry of Education, 2003-2004-2005). The survey data includes the information of student residence in 110 Italian provinces (NUTS3 level regions) before enrolling to a university and the name of university enrolled. The actual flow of students between the province of residence and the exact addresses of universities is extracted and stacked into a table as a column vector as the variable of interest.

3.1 Types

This paper argues that at least three aspatial factors are particularly relevant to the study of HE accessibility. Firstly, the role of parental education is a well-explored factor that affects the educational choices and outcomes of students. Specifically in the Italian context, the educational level of parents is found to be highly influential for the academic attainment of Italian students (Checchi et al., 2003; Bratti et al., 2008). Higher HE participation rates and less drop-outs are observed for students with highly educated parents(Checchi & Flabbi, 2007; Brunori et al., 2012). Moreover, since commuting or migrating to a place involves a cost, the financial condition of families is another aspatial factor relevant to access (see Frenette, 2003; Lupi & Ordine, 2009). Finally, even though education is the primary area where women have made substantial gains and now largely out-perform men (DiPrete & Buchmann, 2006), the question whether there are systematic differences in spatial access to education among males and females remains an important one.

Observed flows are partitioned according to three sets of proxies referring to the socio-economic circumstances of students. Each subgroup forms a type, which cannot be chosen by students. Accordingly, three circumstance variables are employed as shown in Table 1: the presence of at least one highly educated parent at home where the alternative is both parents with basic education. Here “basic education” means the 8 years of compulsory schooling in Italy. ”High education” consists of parents with at least a bachelor’s degree. In the survey 40.66% of mothers and 39.69% fathers are categorized as basically educated , and 14.56% and 20.06% as highly educated, respectively. Survey data contains information on parents’ professions. This information is categorized as high and low for both fathers and mothers. Hence three groups are constructed as both-low, both-high and one-high-one-low. The gender of students is included in addition to the parental background. The combination of these three categories resulted in 12 types as shown in Table 1.

3.2 Distance

The distance from parental residence to HE institutions strongly influences the likelihood of participation and the HE outcomes of students (see for example Gossman et al., 1967; McHugh & Morgan, 1984; Tinto, 1973; Ordovensky, 1995; Gibbons & Vignoles, 2009; Suhonen, 2014). The costs of commuting or migrating may deter access or may impose a barrier when enrolling at a university. For Italy, the empirical evidence confirms that geographical proximity strongly influences the choice of university (Pigini et al., 2013). Indeed, in the survey data, 59.28% studied in their hometown, 40.64% of students was motivated by the closeness of the institution and only 9.74% by the prestige of the university. For student mobility, distance does not only represent costs but is also a predictor of how far students are allowed to live away from their families, which is very relevant in Italy as it is a country characterized by strong family ties (Alesina & Giuliano, 2010).

In this application d_{ij} is the Euclidean distance between the centroid of city i and the exact address of university j. In the QGIS environment the coordinates of city centres are matched with the coordinates of exact location of universities and the Euclidean distance is calculated for each pair. According to the interest of the investigation and the data behaviour it is possible to find the exponential form, exponential square root, the log of distance or a relevant combination of these (De Vries et al., 2009). To choose the most relevant form of deterrence function, the predicted values are examined against observed flows and the power specification of the distance proved most suitable for the data [3]

3.3 University Attractiveness

As far as institutonal attractiveness is concerned, previous studies of SIMs provide mixed findings. Sa et al. (2004) use university rankings as a quality indicator for Dutch students, but the coefficient proves to be negative. Although the authors explain this counterintuitive result as consumption behavior by students in relation to HE (Sa et al., 2004), this is not entirely convincing. Similarly Singleton et al. (2012) employ Times Good University Guide rankings but set an arbitrary power of 0.5 rather than empirical derivation. Dotti et al. (2013) construct an index identifying a province as attractive if inflows exceed outflows, neglecting institutional attractiveness. This paper employs two university fixed characteristics in order to account for attractiveness: the share of successful students in the year before our sample's enrolment and the share of limited places provided by each university. In Italy after secondary school, students take a national exam (Esame di Maturità). The share of students with the highest grades (90-100) from this test in the period 2002-2003 is included in the model (source: MIUR, 2004). Although many programmes are offered on a free-access principle, some require entrance tests, indicating excess demand for these programs. The proportion of limited places to the to-

[3] Also in previous studies the power- decay function has been found to be more suitable for long distance interactions owing to the log-cost perception (A. S. Fotheringham & Webber, 1980; Reggiani et al., 2011).

tal number of places available at University is then used as a quality indicator for the same period (source MIUR, 2004).

3.4 Interactions

Finally, several destination characteristics are interacted with distance to see how the willingness to migrate to further destinations varies among different types (see Gibbons & Vignoles, 2009, for a similar application to British students).

Table 1 shows the interacted institutional characteristics as follows: whether the university at destination is a private institution, dummies for south, central and island locations and a dummy with value 1 for polytechnic universities.

4 Empirical Method and Findings

The examination of the model is operated through related statistical log linear models which were developed alongside entropy maximizing models. There are several ways of handling spatial interaction models [4] (see A. G. Wilson, 1971; Yun & Sen, 1994; LeSage & Pace, 2008). This study makes use of the Poisson gravity models, [5] with the same statistical properties, producing identical estimations to entropy maximization models (Baxter, 1982).

The Poisson gravity model takes the following form:

$$E(N_{ij}) = T_{ij} = O_i^{\theta} D_j^{\alpha} f(d_{ij}) \tag{4}$$

where N_{ij} indicates observed flows, whereas T_{ij} is the expectation of observed

[4] Ordinary least square (OLS) estimation proves to be insufficient especially in handling zero flows (Piermartini & Yotov, 2016), which constitute important information for immobile students.

[5] (see Flowerdew & Aitkin, 1982; Smith, 1987, for theoretical development)

flows, treated as a random variable and assumed to have a Poisson distribution (Baxter, 1982) [6].

The model is calibrated by the generalized linear model (GLM) package in R [7], where flows follow a Poisson distribution with a logarithmic link between variables. The estimations are carried out separately for 12 subgroups. Finally, the Poisson regression is:

$$T_{ijk} = \exp[constant + O_i + \alpha D_{jk} + \mu\delta_{ij} + \beta ln(d_{ij}) + \gamma S_j + \eta L_j + \sum_l \lambda_l ln(d_{ij}) U_{jl}] \tag{5}$$

where $k = 1, 2, 3, ..., 12$ represents types, S_j the share of successful students and L_j the share of limited places at j, and U_{jl} a set of interactions on distance. This regression produces an exponential value of factor for origin i and is proportionally equivalent to the product $K_i O_i$, and is therefore equivalent to the production-constrained model of the entropy-maximizing system.

Table 2 and Table 3 show the results of the first set of regressions where the model has been applied to 10 groups. Groups 5 and 6 are not taken into consideration due to the lower number of observations. As expected, distance has a very strong significantly negative effect, indicating a deterrence impact for each group. For 1 km decrease in the distance the expectation number of student flows increases by factors varying form 1.499 to 1.709. The impact is higher for female students than for male students except for those with at least one highly educated parent. As a student's family background becomes more favourable in terms of the proxies specified above, the difference between male and female shrinks and ultimately female students feel less deterred. As in previous studies, δ is significant at the 0.01 level for all groups and positive in sign, capturing the benefit of residing and studying in the same city. Similar values are observed for different types but with different motivations. For socially advantaged groups the parameter μ captures the fact that these students usually live in big cities where large universities are located and

[6] The probability mass function of flows is given by $Pr(T_{ij}) = \frac{\exp^{-N_{ij}} N_{ij}^T ij}{T_{ij}!}$

[7] (see Dennett, 2012, for details)

hence they do not need to migrate. On the other hand, groups 1 to 4 it may reflect the actual startup costs where these students decide to migrate. As far as the attractiveness of universities is concerned, S_j (the share of successful students at university) is significant and positively affects flows only for students who have at least one highly educated parent. Other students seem to be unaffected by institutional quality. The effect is observed for L_j (the share of limited places offered by universities) again for socially advantaged groups. Among students with disadvantaged parental backgrounds, only female students (groups 2 and 4) are attracted to these limited positions. This is probably due to the fact that female students are interested in faculties such as medicine and nursing, requiring entrance tests. Hence, for students with a poor parental background, what seems to matter is obtaining a degree irrespective of the prestige of the University (Triventi & Trivellato, 2009).

The remaining results allow for interactions between institutional characteristics and distance. Private universites at destination increase the tendency of travelling longer distances for all groups. It is an expected result since for any type deciding to enrol a private university, distance must be becoming irrelevant. Looking at the significance levels, polytechnics do not seem to induce students to travel far except for groups 2 and 7. In contrast to Dotti et al. (2013), interacting distance with macro regions, where universities are located, shows that the central region attracts more students than the south for all students except types 1 and 3. These two types comprise male students with poor family backgrounds who may prefer Universities in the south due to the lower cost of living. Finally universities located in Sicily and Sardinia fail to attract students from all backgrounds.

Table 2-Table 3 [About Here]

After obtaining the parameter values from (5), accessibility scores are calculated through Equation 3 for each group with their respective impedance functions as follows:

$$A_{ik} = \sum_{j=1} \frac{C_j d_{ij}^{\hat{\beta}_k}}{\delta_{ijk}} \tag{6}$$

where $k = 1, 2, 3, ..., 12$. As it for the production constrained SIM, d_{ij} is constructed from city centroids to the exact addresses of universities (to the largest campuses),so there is no zero distance, which in return accounts for the self-potential (local demand) of universities within provinces. The measured scores indicate potential access in terms of the places offered to students. Higher scores indicate better accessibility to the 77 total number of universities located in 101 different provinces. Maps 1-2 illustrate the access scores of 10 groups, where darker blue indicates a higher score (The scores are also shown in Appendix Table 1-2).

Figure 1-2 [About Here]

The first thing to note from the figures is that if a student belongs to a socially advantaged group, then their access is relatively higher where ever they live, except very far south in the country. Similarly for socially disadvantaged groups, even if they live in a big city where large universities are located, access remains low, particularly in the south. The lowest access is observed for group 2 where the type comprising female students with lower-class parents with a basic education. The types constructed for this paper seem particularly relevant for female students. Access increases on average 101% from group 2 to group 12 whereas parental background does not seem to affect male student access to HE as much as female students. From the least to the highest, access increases 10% on average. Moreover a degree of gender discrimination in access is observed in the first 4 groups, but lessens as parental education and financial condition improve.

Decomposition of Access Inequality

To disentangle the relative contributions of spatial and aspatial factors to inequality, a decomposable inequality index is used [8]. As shown in Table 4, the resulting inequality is a sum of within and between inequalities. The first row of Table 4 shows the inequality decomposition where the variation within types of students is suppressed by substituting each type's accessibility score with its arithmetic mean. By this method the inequality between the types of students is computed as 0.01776 and represents the contribution of socio-economic factors to total inequality (5% of total inequality). Whereas in the second row the variation within provinces is eliminated by substituting each province's accessibility score with its arithmetic mean, in this approach the inequality between provinces is computed as 0.34637 and the contribution of socio-economic factors to total inequality is measures at 7%.

For the sake of a better understanding of the computed inequality, take a female student with a poor family background (group 2) living in Matera, in order for her to have as much access as a male student with the same family origin (group 1) living in the same city, she has to travel 151 km to the nearest city, Foggia (social distance). Moreover, in order for her to have similar access to a female student with better family origin (group12) living in Napoli, she has to travel 460 km to the nearest city (social+physical distance). Therefore, the findings indicate that despite the expansion of HE supply in the country, access to HE is strongly unequal due to the spatial distribution of opportunities with additional disparity due to socio-economic factors at the locations of origin.

Table 4 [About Here]

[8]Mean Logarithmic Deviation is a path-independent decomposable inequality measure (Foster & Shneyerov, 2000). It is defined as: $MLD(X) = \frac{1}{N}\sum_{1}^{N} \ln \frac{\mu_x}{x_i}$ where X is a distribution, N population size and μ_X is mean.

5 Conclusions

This paper provides empirical evidence for the dynamics of student mobility in Italy and measures inequality in access to HE institutions with particular attention to the socio-economic background of students. Using a spatial interaction model, the flows of students to universities are defined as interactions between provinces in Italy. The results demonstrate that poor family background students are impervious to university-quality effects and university quality becomes relevant only for students with better family backgrounds. The model allowed for interactions between institutional characteristics and distance to see how elasticity with respect to spatial distance varies given the heterogeneity of the universities. The results indicates that private universities attract students and increase their willingness to travel longer distances. Universities located in the central region attract more students than in the south and the location in Sicily or Sardinia deters flows.

As far as distance is concerned, the values of these parameters are highly significant and negative in sign, indicating a deterrence effect for each group of students. In line with previous studies, δ was significant at the 0.01 level for all groups and positive in sign, capturing the benefit of residing and studying in the same city. For the second step computed deterrence functions and δs were imported into a Hansen-like accessibility index and accessibility scores of 110 Italian provinces to 77 Italian universities were computed. The results show that socio-economic background matters especially for female student mobility. Finally, the share of aspatial factors in inequality of access between types proved to be 5% with the first approach and 7% when computed with the second approach.

This paper contributes to the accessibility literature by a multidisciplinary approach providing a spatial accessibility measure for Italian HE institutions with particular attention to socio-economic sources of inequality. It also contributes to EOp literature by showing how spatial dimensions of

EOp could be incorporated into models that rely solely on spatial elements. Furthermore, the investigation of 10 types provides a clear ranking. In other words, through this application, this study empirically shows which socio-economic group is better off and by how much. Finally, this study is the first attempt to define parental location, clearly an exogenous factor for students, as a circumstance.

The findings provide highly detailed information for policy implications. In order to increase accessibility three policy strategies can be adopted. Firstly, an effective policy may target the types with lower potential accessibility to assist them though loans, scholarships and grants. Secondly, the geographical locations where accessibility is lower can be identified and accessibility can be increased by the reduction of geographic barriers for cities such as Nuoro, Brindisi, Ragusa and Belluno where new universities and/or places may be set up. Finally a combination of these two can be used. For instance, the empirical evidence in this paper shows that female students with disadvantaged family backgrounds located in southern Italy would benefit the most from HE funding. More precisely the identification of inequality resulting from gender and geography can be extracted from the findings as follows: for a female student living in the south with low income parents both with a basic education, on average the potential accessibility is 146.15% lower than a male student living in the North with better family origin. These examples can be extended to determine a variety of policy strategies.

References

Alesina, A., & Giuliano, P. (2010). The power of the family. *Journal of Economic growth*, *15*(2), 93–125.

Alm, J., & Winters, J. V. (2009). Distance and intrastate college student migration. *Economics of Education Review*, *28*(6), 728–738.

Baxter, M. (1982). Similarities in methods of estimating spatial interaction models. *Geographical Analysis*, *14*(3), 267–272.

Bratti, M., Checchi, D., & De Blasio, G. (2008). Does the expansion of higher education increase the equality of educational opportunities? evidence from italy. *Labour*, *22*(s1), 53–88.

Brunori, P., Peragine, V., & Serlenga, L. (2012). Fairness in education: The italian university before and after the reform. *Economics of Education Review*, *31*(5), 764–777.

Checchi, D., & Flabbi, L. (2007). Intergenerational mobility and schooling decisions in germany and italy: the impact of secondary school tracks.

Checchi, D., et al. (2003). The italian educational system: family background and social stratification. In *Isae conference on monitoring italy.*

Checchi, D., & Peragine, V. (2010). Inequality of opportunity in italy. *The Journal of Economic Inequality*, *8*(4), 429–450.

Cooke, T. J., & Boyle, P. (2011). The migration of high school graduates to college. *Educational Evaluation and Policy Analysis*, *33*(2), 202–213.

Dardanoni, V., Fields, G. S., Roemer, J. E., & Puerta, M. L. S. (2006). How demanding should equality of opportunity be, and how much have we achieved? *Mobility and inequality: Frontiers of research from sociology and economics*, 59–82.

Dennett, A. (2012). *Estimating flows between geographical locations:get me started inspatial interaction modelling* (Tech. Rep.). Citeseer.

De Vries, J. J., Nijkamp, P., & Rietveld, P. (2009). Exponential or power distance-decay for commuting? an alternative specification. *Environment and Planning A*, *41*(2), 461–480.

DiPrete, T. A., & Buchmann, C. (2006). Gender-specific trends in the value of education and the emerging gender gap in college completion. *Demography*, *43*(1), 1–24.

Dotti, N. F., Fratesi, U., Lenzi, C., & Percoco, M. (2013). Local labour markets and the interregional mobility of italian university students. *Spatial Economic Analysis*, *8*(4), 443–468.

Ferreira, F. H., & Gignoux, J. (2011). The measurement of inequality of opportunity: Theory and an application to latin america. *Review of Income and Wealth*, *57*(4), 622–657.

Flowerdew, R., & Aitkin, M. (1982). A method of fitting the gravity model based on the poisson distribution*. *Journal of regional science*, *22*(2), 191–202.

Foster, J. E., & Shneyerov, A. A. (2000). Path independent inequality measures. *Journal of Economic Theory*, *91*(2), 199–222.

Fotheringham, A., & O'Kelly, M. E. (1989). *Spatial interaction models: formulations and applications* (Vol. 5). Kluwer Academic Pub.

Fotheringham, A. S., & Webber, M. J. (1980). Spatial structure and the parameters of spatial interaction models. *Geographical Analysis*, *12*(1), 33–46.

Frenette, M. (2003). Access to college and university: Does distance matter. ottawa. *Statistics Canada, Catalogue*(11F0019MIE2003201).

Gibbons, S., & Vignoles, A. (2009). Access, choice and participation in higher education.

Gitlesen, J. P., & Thorsen, I. (2000). A competing destinations approach to modeling commuting flows: a theoretical interpretation and an empirical application of the model. *Environment and Planning A*, *32*(11), 2057–2074.

Gossman, C. S., et al. (1967). Migration of college and university students in the united states.

Hansen, W. G. (1959). How accessibility shapes land use. *Journal of the American Institute of planners*, *25*(2), 73–76.

Haynes, K. E., & Fotheringham, A. S. (1984). *Gravity and spatial interaction models* (Vol. 2). Sage publications Beverly Hills.

LeSage, J. P., & Pace, R. K. (2008). Spatial econometric modeling of origin-destination flows*. *Journal of Regional Science*, *48*(5), 941–967.

Lowe, J. M., & Sen, A. (1996). Gravity model applications in health planning: Analysis of an urban hospital market*. *Journal of Regional Science*, *36*(3), 437–461.

Lupi, C., & Ordine, P. (2009). Family income and students'mobility. *Giornale degli Economisti e Annali di Economia*, 1–23.

Mayhew, L., Gibberd, R., & Hall, H. (1986). Predicting patient flows and hospital case-mix. *Environment and Planning A*, *18*(5), 619–638.

McArthur, D. P., Kleppe, G., Thorsen, I., & Ubøe, J. (2011). The spatial transferability of parameters in a gravity model of commuting flows. *Journal of Transport Geography*, *19*(4), 596–605.

McHugh, R., & Morgan, J. N. (1984). The determinants of interstate student migration: A place-to-place analysis. *Economics of Education Review*, *3*(4), 269–278.

Ordovensky, J. F. (1995). Effects of institutional attributes on enrollment choice: Implications for postsecondary vocational education. *Economics of Education Review*, *14*(4), 335–350.

Piermartini, R., & Yotov, Y. (2016). Estimating trade policy effects with structural gravity.

Pigini, C., Staffolani, S., et al. (2013). *Enrollment costs, university quality and higher education choices in italy* (Tech. Rep.). University Library of Munich, Germany.

Reggiani, A., Bucci, P., & Russo, G. (2011). Accessibility and network structures in the german commuting. *Networks and Spatial Economics*, *11*(4), 621–641.

Roemer, J. E. (1998). *Theories of distributive justice*. Harvard University Press.

Sa, C., Florax, R. J., & Rietveld, P. (2004). Determinants of the regional demand for higher education in the netherlands: A gravity model approach. *Regional Studies*, *38*(4), 375–392.

Sen, A., & Smith, T. (2012). *Gravity models of spatial interaction behavior*. Springer Science & Business Media.

Singleton, A. D., Wilson, A., & OBrien, O. (2012). Geodemographics and spatial interaction: an integrated model for higher education. *Journal of Geographical Systems*, *14*(2), 223–241.

Smith, T. E. (1987). Poisson gravity models of spatial flows*. *Journal of Regional Science*, *27*(3), 315–340.

Suhonen, T. (2014). Field-of-study choice in higher education: Does distance matter? *Spatial Economic Analysis*, *9*(4), 355–375.

Talen, E., & Anselin, L. (1998). Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds. *Environment and Planning a*, *30*(4), 595–613.

Tinto, V. (1973). College proximity and rates of college attendance. *American Educational Research Journal*, *10*(4), 277–293.

Triventi, M., & Trivellato, P. (2009). Participation, performance and inequality in italian higher education in the 20th century. *Higher Education*, *57*(6), 681–702.

Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning A*, *3*(1), 1–32.

Wilson, R. M., & Gibberd, R. (1990). Allocating resources between health care regions: Catchment populations and allocating for equity. *European Journal of Operational Research*, *49*(3), 333–347.

Yun, S., & Sen, A. (1994). Computation of maximum likelihood estimates of gravity model parameters*. *Journal of Regional Science*, *34*(2), 199–216.

Table 1: Variables in analysis, data from ISTAT and MIUR

Variables	Description
Residence	Province of residence before enrollment (ISTAT)
Destination University	Enrolled University 63 state and 14 private universities(ISTAT)
Distance	Euclidean distance between city centroids and University addresses,
	measured in QGIS based on coordinates
Sex	14,000 male and 17,400 female students graduated in 2007(ISTAT)
Parent's Education	The highest degree obtained by parents(ISTAT)
	Two categories: at least one highly educated parent, both basic educated
	"basic education" covers high school degree
	high category at least bachelor' s degree.
Financial Condition	Occupation type of parents(ISTAT)
	Three categories : both-high, one-high,both-low
	High=Managers, Directors,High/Medium Qualification
	Low=Office Worker, Lower-skilled workers
S_j	Share of students who achieved highest scores (90-100)
	from compulsory test before HE enrollment in the period 2002-2003 (MIUR)
L_j	Proportion of limited places offered by universities to the total places (MIUR)
U_{jl}	Institutional characteristics to be interacted with distance (MIUR)
	U_{j1} 1 if private 0 otherwise
	U_{j2} 1 if polytechnic 0 otherwise
	U_{j3} 1 if south U_{j4} 1 if center U_{j5} 1 if island 0 otherwise
Types	group1(both basic educated parents,male , low financial condition)
	group2(both basic educated parents,female , low financial condition)
	group3(both basic educated parents,male, medium financial condition)
	group4(both basic educated parents, female, medium financial condition)
	group5(both basic educated parents, male, high financial condition)
	group6(both basic educated parents, female, high financial condition)
	group7(at least one high educated parent, male, low financial condition)
	group8(at least one high educated parent, female, low financial condition)
	group9(at least one high educated parent, male, medium financial condition)
	group10(at least one high educated parent, female, medium financial condition)
	group11(at least one high educated parent, male, high financial condition)
	group12(at least one high educated parent, female, high financial condition)

Table 2: Results of Poisson Regression First 5 Groups

Groups	(1)	(2)	(3)	(4)	(7)
Variables	Basic-Male-Lower Class	Basic-Female-Lower Class	Basic-Male-Middle Class	Basic-Female-Middle Class	$\geq$ 1high-Male-Lower Class
S_j	0.455	0.697	0.556	-0.087	1.502***
	(0.421)	(0.373)	(0.595)	(0.504)	(0.396)
L_j	0.158	0.303***	0.214	0.277*	0.140
	(0.098)	(0.086)	(0.133)	(0.119)	(0.123)
$\hat{\mu}$	0.249***	0.293***	0.261***	0.300***	0.307***
	(0.389)	(0.032)	(0.056)	(0.045)	(0.036)
Distance	-1.535***	-1.709***	-1.530***	-1.574***	-1.625***
	(0.047)	(0.045)	(0.069)	(0.060)	(0.046)
Institutional Interactions					
x Private Univ.	0.726***	0.608***	0.509***	0.599***	0.631***
	(0.070)	(0.058)	(0.093)	(0.067)	(0.057)
x Polytechnic	0.003	0.240	0.105	0.276*	0.134*
	(0.072)	(0.088)	(0.105)	(0.105)	(0.022)
x South	0.561***	0.359***	0.549***	0.386***	432***
	(0.062)	(0.059)	(0.090)	(0.076)	(0.065)
x Center	0.527***	0.545***	0.418***	0.495***	0.467***
	(0.060)	(0.058)	(0.088)	(0.078)	(0.058)
x Island	-0.206***	-0.039	-0.130	-0.375*	-0.283*
	(0.155)	(0.122)	(0.201)	(0.173)	(0.162)
(Intercept)	6.528***	7.258***	5.442***	6.304***	6.790***
	(0.227)	(0.202)	(0.392)	(0.291)	(0.221)
Observations	1,572	1,572	1,572	1,572	1,572
R2	0.88	0.89	0.87	0.86	0.86

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Education of Parents-Gender of Student-Financial Condition of Parents

Table 3: Results of Poisson Regression Last 5 Groups

Groups	(8)	(9)	(10)	(11)	(12)
Variables	≥1high-Lower Class	≥1high-Male-Middle Class	≥ 1high-Female-Middle Class	≥ 1high-Male-Higher Class	≥1high-Female-Higher Class
S_j	1.745***	2.291***	1.444***	2.006***	1.947***
	(0.337)	(0.313)	(0.287)	(0.293)	(0.260)
L_j	0.384***	0.184***	0.187***	0.112**	0.348***
	(0.065)	(0.073)	(0.072)	(0.082)	(0.081)
$\hat{\mu}$	0.337***	0.287***	0.281***	0.305***	0.273***
	(0.032)	(0.032)	(0.029)	(0.029)	(0.027)
Distance	-1.581***	-1.551***	-1.532***	-1.522***	-1.499***
	(0.041)	(0.040)	(0.034)	(0.035)	(0.032)
Institutional Interactions					
x Private Univ.	0.590***	0.490***	0.481***	0.549***	0.573***
	(0.045)	(0.043)	(0.038)	(0.037)	(0.033)
x Polytechnic	-0.538	0.070	0.706	0.072	0.042
	(0.078)	(0.051)	(0.061)	(0.047)	(0.054)
x South	0.123*	0.436***	0.198***	0.320***	0.255***
	(0.636)	(0.609)	(0.054)	(0.054)	(0.049)
x Center	0.399***	0.499***	0.418***	0.384***	0.407***
	(0.053)	(0.516)	(0.045)	(0.045)	(0.040)
x Island	-0.150	0.816	-0.252*	-0.490***	-0.331**
	(0.125)	(0.130)	(0.111)	(0.118)	(0.103)
(Intercept)	6.654***	7.113***	7.084***	6.851***	6.939***
	(0.194)	(0.187)	(0.172)	(0.170)	(0.155)
Observations	1,572	1,572	1,572	1,572	1,572
R2	0.85	0.87	0.89	0.86	0.88

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Education of Parents-Gender of Student-Financial Condition of Parents

Table 4: Decomposition of Inequality In Access (MLD measures)

	Spatial Inequality	Inequality due to socioeconomic background	Total Inequality
Inequality in Access to HE (First Approach)	0.35444	0.01776	0.37220
Inequality in Access to HE (Second Approach)	0.34637	0.02583	0.37220
Percentage Contribution (First Approach)	%95	%5	%100
Percentage Contribution (Second Approach)	%93	%7	%100

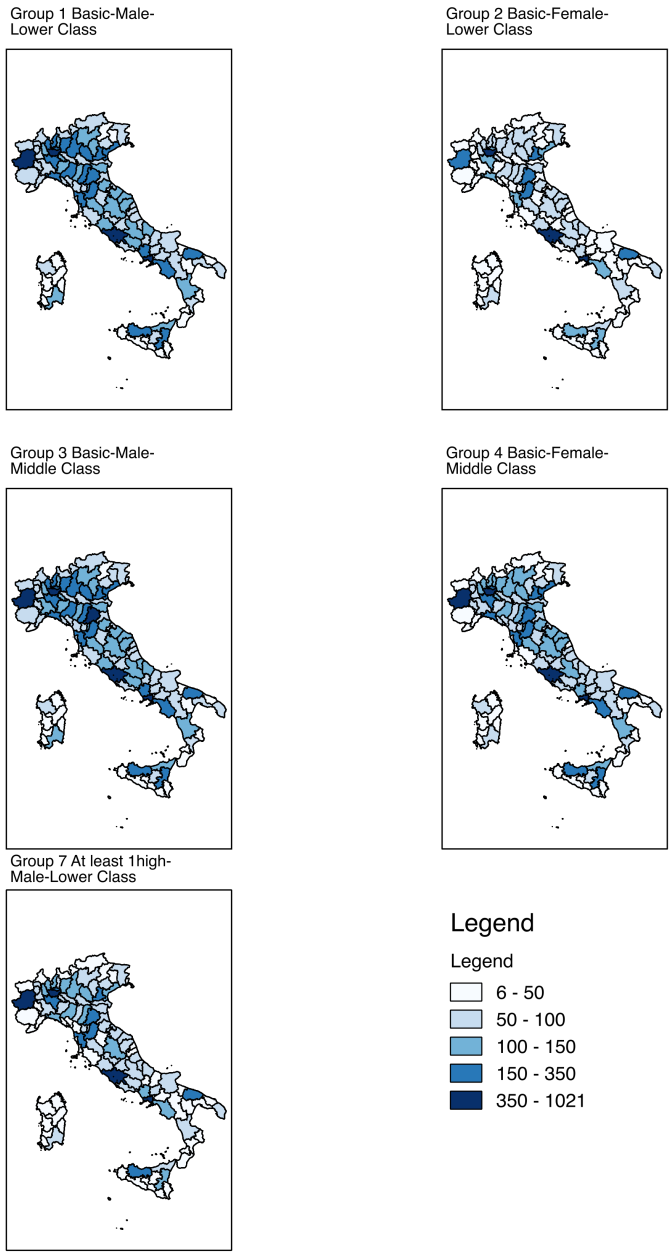

Figure 1: First 5 Groups

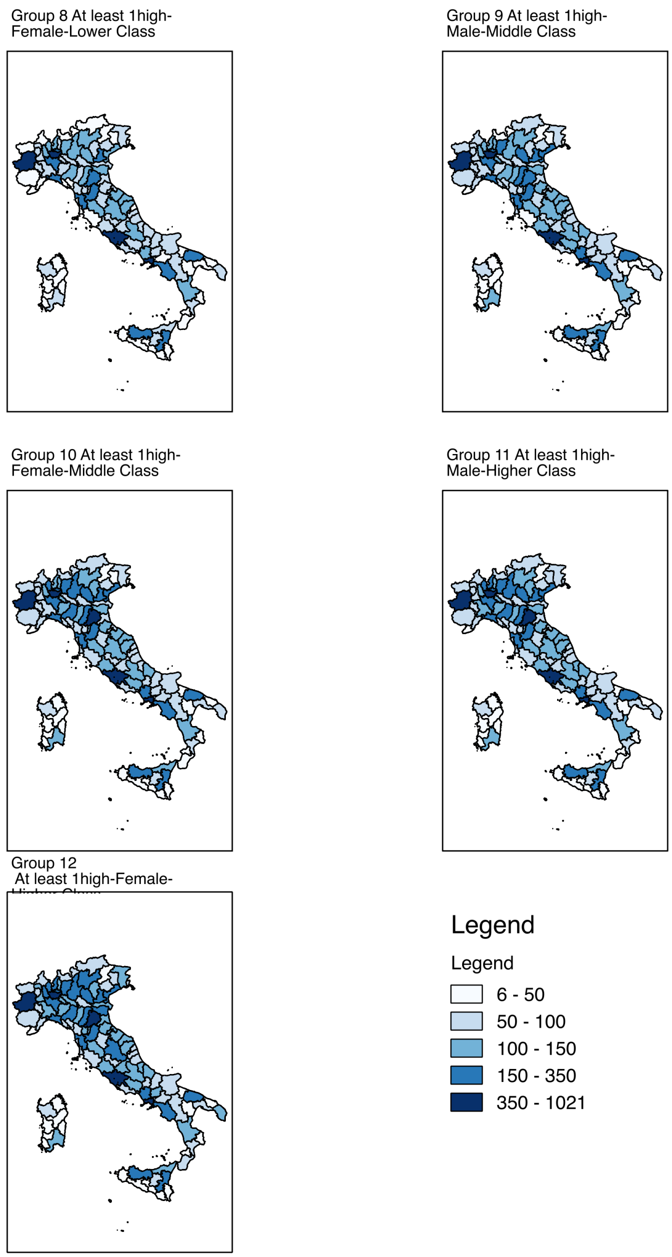

Figure 2: Last 5 Groups

Appendix 1

Table 5: Accessibility Scores

Groups	(1)	(2)	(3)	(4)	(7)	(8)	(9)	(10)	(11)	(12)
Torino	470	305	475	424	374	417	450	472	483	513
Vercelli	90	41	91	72	57	68	81	88	90	103
Novara	148	75	149	123	101	117	135	146	148	166
Cuneo	60	28	61	49	39	46	54	59	61	69
Asti	77	34	78	61	49	58	69	76	77	88
Alessandria	100	48	102	82	66	78	91	99	101	114
Valle d'Aosta	56	25	57	45	35	42	51	55	57	65
Imperia	39	16	39	30	23	28	35	38	39	45
Savona	56	25	57	45	35	42	51	56	57	65
Genova	191	113	193	167	143	162	180	191	195	212
La Spezia	67	29	68	53	42	50	60	66	68	78
Varese	157	82	158	131	109	125	144	155	158	176
Como	152	80	153	127	106	122	140	150	153	170
Sondrio	55	23	56	43	34	41	49	54	56	65
Milano	937	613	948	848	749	833	899	942	963	1021
Bergamo	174	96	176	148	125	143	161	173	176	194
Brescia	159	89	161	136	115	132	148	158	162	178
Pavia	228	127	231	195	164	187	212	226	231	254
Cremona	99	45	100	79	63	75	89	97	99	113
Mantova	82	35	83	65	51	61	74	81	82	95
Bolzano/Bozen	56	27	57	46	37	44	51	55	57	64
Trento	136	78	138	117	100	114	127	136	139	152
Verona	170	97	172	147	124	142	159	170	173	190
Vicenza	102	51	103	84	69	80	93	101	103	116
Belluno	42	17	43	33	25	31	37	41	42	49
Treviso	83	40	84	68	55	65	75	82	84	94
Venezia	174	102	176	152	130	147	163	174	177	193
Padova	304	188	307	270	234	263	288	304	311	334
Rovigo	99	47	100	80	65	76	90	98	99	113
Udine	97	56	98	84	71	81	91	97	99	108
Gorizia	51	23	51	41	32	38	46	50	51	58
Trieste	87	49	88	75	63	72	81	87	89	97
Piacenza	107	50	108	87	70	82	97	105	107	122
Parma	157	87	159	135	113	130	147	157	160	176
Reggio nell'Emilia	120	60	121	99	81	95	110	119	121	136
Modena	163	89	165	138	116	133	151	162	165	183
Bologna	350	217	354	311	270	303	332	351	358	385
Ferrara	141	76	143	120	100	115	131	140	143	159
Ravenna	93	43	94	75	60	71	84	92	94	106
Forl-Cesena	122	64	123	102	85	98	112	121	123	137
Pesaro e Urbino	130	72	132	111	93	107	121	129	132	146
Ancona	126	70	127	107	90	104	117	125	128	141
Macerata	123	67	124	105	88	101	114	122	125	138
Ascoli Piceno	70	32	71	57	45	53	63	69	71	81
Massa-Carrara	68	28	69	53	41	50	60	67	68	79
Lucca	72	30	72	56	44	53	64	70	72	83
Pistoia	90	40	91	72	57	67	81	89	90	104
Firenze	302	185	305	267	232	261	286	302	309	333
Livorno	61	26	62	48	38	45	55	60	61	71
Pisa	238	145	241	210	181	205	225	238	244	263
Arezzo	77	34	78	61	48	58	69	76	77	89
Siena	125	68	127	107	89	103	116	125	128	141
Grosseto	53	21	54	41	32	39	47	52	54	62
Perugia	146	82	148	125	106	121	136	146	149	163
Terni	85	39	86	69	55	65	77	84	86	98
Viterbo	111	58	113	93	77	89	103	111	113	126
Rieti	88	39	89	70	56	66	79	87	88	101

Table 6: Accessibility Scores Continued

Groups	(1)	(2)	(3)	(4)	(7)	(8)	(9)	(10)	(11)	(12)
Roma	914	602	924	829	733	816	878	919	940	996
Latina	88	41	89	71	57	68	80	87	88	100
Frosinone	111	58	112	93	77	89	102	110	113	126
Caserta	164	92	166	141	119	136	153	164	167	183
Benevento	94	47	95	78	64	74	86	93	95	107
Napoli	579	377	586	523	462	514	555	582	595	632
Avellino	66	29	67	53	42	49	59	65	66	76
Salerno	187	115	190	166	143	162	178	188	192	207
L'Aquila	125	67	127	106	89	102	116	125	127	141
Teramo	80	39	81	66	54	63	73	80	81	92
Pescara	99	52	101	83	69	80	92	99	101	112
Chieti	115	64	117	99	83	95	108	115	118	129
Campobasso	74	36	75	61	49	58	67	73	75	84
Foggia	79	42	80	67	56	65	74	79	81	90
Bari	253	162	256	228	200	223	242	254	260	278
Taranto	54	27	54	45	37	43	49	53	55	61
Brindisi	32	14	32	25	20	24	29	32	32	37
Lecce	93	57	94	82	71	80	88	93	95	103
Potenza	64	32	65	53	43	51	59	64	65	73
Matera	43	19	44	35	27	32	39	43	44	50
Cosenza	131	80	132	116	100	113	124	131	134	144
Catanzaro	67	39	68	58	50	57	63	67	69	75
Reggio di Calabria	45	24	46	38	32	37	42	45	46	51
Trapani	24	10	24	19	15	18	21	24	24	28
Palermo	201	129	203	180	158	177	192	202	206	220
Messina	115	70	117	102	88	99	109	115	118	127
Agrigento	33	15	34	27	21	25	30	33	34	38
Caltanissetta	31	14	32	25	20	23	28	31	31	36
Enna	65	35	66	55	46	53	60	65	66	73
Catania	176	112	178	158	138	154	168	177	181	193
Ragusa	26	12	27	21	17	20	24	26	27	30
Siracusa	30	14	30	24	19	23	27	30	30	34
Sassari	62	36	63	54	46	53	59	63	64	70
Nuoro	21	8	21	16	12	15	19	21	21	25
Cagliari	112	70	113	100	87	98	107	112	115	123
Pordenone	50	22	51	40	31	38	45	49	51	58
Isernia	71	32	72	57	45	53	64	70	71	82
Oristano	19	7	19	15	11	14	17	19	19	23
Biella	80	36	81	64	51	60	72	78	80	92
Lecco	124	60	125	101	82	95	113	122	124	140
Lodi	129	60	130	104	83	97	116	126	128	146
Rimini	95	47	96	78	63	74	86	94	96	108
Prato	125	60	126	102	82	96	113	123	124	141
Crotone	24	9	24	19	14	17	21	24	24	28
Vibo Valentia	25	10	25	19	15	18	22	25	25	30
Verbano-Cusio-Ossola	57	24	58	45	35	42	51	56	57	66
Olbia-Tempio	23	9	23	18	13	17	20	23	23	27
Ogliastra	20	7	20	15	11	14	17	20	20	24
Medio Campidano	22	9	22	17	13	16	19	21	22	26
Carbonia-Iglesias	17	6	17	13	10	12	15	17	17	20
Monza e della Brianza	270	145	272	225	188	213	247	265	267	298
Fermo	63	27	64	50	39	47	56	62	63	73
Barletta-Andria-Trani	43	18	43	33	26	31	38	42	43	50

Chapter 3

Inequality of Opportunity in Sweden:A Spatial Perspective

Inequality of Opportunity in Sweden:A Spatial Perspective *

Umut TÜRK[†]& John ÖSTH[‡]

Abstract

This paper investigates the spatial dimensions of inequality of opportunity by integrating parental neighbourhood characteristics as exogenous factors influencing the life chances of individuals. We construct egocentric neighbourhoods, where contextual variables are quantified by an approach based on k nearest neighbours. The analyses are carried out with multilevel models departing significantly from previous studies where solely OLS regressions were employed. Using Swedish longitudinal register data, we show that the parental neighbourhood is highly influential in educational inequality of opportunity and remains so for earnings inequality of opportunity even years after exposure.

Keywords neighbourhood effects, equality of opportunity

*This paper benefited from the insightful comments and suggestions of the participants of the Winter School on Space, Mobility and Opportunities in Canazei, Italy and Seminar Series in Verona, Italy

[†]Economics Department, University of Verona, Verona Italy. Email address: umut.turk@univr.it

[‡]Department of Social and Economic Geography, Uppsala University, Uppsala Sweden

1 Introduction

In recent years, the equality of opportunity (hereafter EOp) concept has often been mentioned in discussions concerning distributional disparities (Cities and Social Equity, 2009; Urban Equity of Life, 2014; Handbook of Income Distribution, 2015). It is almost a universally accepted principle developed recently by Roemer (1998) and already with numerous empirical applications (see Ramos & Van de Gaer, 2012; Roemer & Trannoy, 2013; Brunori et al., 2013, for reviews). In this framework the overall inequality observed in different spheres of social life is decomposed into ethically acceptable (fair) and offensive (unfair) components. So-called *circumstances* representing unfair sources of inequality, are predetermined and beyond people's control such as gender, race and family background (Roemer, 1998; Roemer & Trannoy, 2013). On the other hand, so-called *effort* comprises fair (acceptable) sources of inequality for which individuals are held responsible. Inequality that is due to circumstances is defined as inequality of opportunity (hereafter IOp).

The IOp is a product of several underlying inequalities, such as inequality due to differences in social treatment, inequality of access to basic opportunities, inequality due to exogenous genetic factors and inequality due to parental resources and location (de Barros, 2009). A number of empirical studies seek to disentangle these underlying inequalities in opportunities through a set of circumstance variables. To assess IOp accurately, analyses thus need to take a comprehensive approach on the circumstance variables employed. Most existing studies have been limited to accounting for inequalities due to parental resources and social treatment, with gender, race, parental income and education as typical predictors of circumstances. So far, however, there has been little discussion of the spatial sources of IOp and none for inequalities due to parental location.

It has already been suggested that residential location may generate unfair inequalities especially for children (Ross et al., 2002; de Barros et al., 2008). However, the empirical studies that have sought to evaluate IOp, include geography either as a reference to a general Urban/Rural division of birthplace (Ferreira et al., 2010), or as large administrative units, for instance regions (Peragine & Serlenga, 2008; Checchi et al., 2010; Singh, 2012). Leaving aside the problems of robustness due to spatial effects in place such as the modifiable areal unit problem (MAUP) (Openshaw, 1984), on such a scale geography does not represent the residential characteristics to which individuals are exposed. Therefore, a more specific characterization of spatial patterns that communicates the current and past residential environment of individuals and interaction among them must be included in analyses.

This paper focuses on the role of people's parental neighbourhood characteristics as a source of IOp in education and income and their own neighbourhood characteristics as a source of legitimate inequality in income distribution. The paper aims to go beyond the analysis of traditonal sources of inequality by linking the EOp literature to the literature investigating the neighbourhood effects on various outcomes of individuals. Neighbourhood studies offer empirical evidence on the link between neighbourhoods and the several life chances of residents. Most previous findings are relevant to this study, for example, neighbourhood effects on child outcomes (see Leventhal & Brooks-Gunn, 2000, for a review), on labour market and economic outcomes (see Vartanian, 1999) and spatial mismatch implications (for a review, see Kain, 1992), health outcomes such as psychological wellbeing (see Ludwig et al., 2013), behavioural outcomes such as the likelihood of committing a crime and drug-alcohol consumption (Case & Katz, 1991). However, it may be problematic to attribute causal relations to how these effects take place (Buck, 2001). Considerable effort has been put into providing a proper definition of neighbourhoods, and causal relations with their observed effects (see Ellen & Turner, 1997, for a review). Galster (2001) argues that the neighbourhood is a multidimensional phenomenon, in which four actors (households, businesses, property holders and local government) act both as consumers and producers in shaping the structural, demographic and social-interactive characteristics of neighbourhoods (Galster, 2001). As consumers, residents are exposed to institutional mechanisms, to peers and networks, to environmental aspects (i.e. polluted air) and to offered accessibility of opportunities (Sharkey & Faber, 2014). In the light of these suggestions, we treat neighbourhoods as the environment surrounding residents where its scale is based on the interaction possibilities between individuals.

The aim of the present work is twofold. The first part of the paper is devoted to the analysis of neighbourhood effects. We use a multilevel modelling strategy to disentangle the influence of circumstances in relation to parental neighbourhood on educational attainment when living with parents and additionally parental and own-neighbourhood characteristics influencing disposable income when living independently of parents. Using the longitudinal register database from Sweden, we focus on the whole population of the 1985 cohort. The database provides family background variables such as parental education, employment, marital status and national origin, and provides information on individuals disposable incomes and compulsory exam marks that are taken as dependent variables. Since residential coordinates on 100m x 100m level are available for each individual as well as over time, geographical information on residents is used to construct bespoke neighbourhoods for each individual with the aim of channeling several characteristics of residential locations based on a k-nearest neighbours approach (Östh et al., 2015). In addition, we include a measure of negative environment surrounding the parental neighbourhood derived from the Corine (coordination of information on the environment) database and a measure of job/housing balance in own-neighbourhood. In the second part of the paper, we construct a model to perform a comprehensive investigation of IOp with particular attention to the spatial sources of inequality. As in Bourguignon et al. (2007) and Ferreira & Gignoux (2011) we formulate a situation where the within-group inequalities are eliminated, so that the overall inequality both in educational attainment and disposable income is decomposed into circumstance (IOp) and effort components. As it is the foremost aim of this study, we also decompose circumstances and effort into spatial and aspatial components.

The results of the model show that parental neighbourhoods explain 36.94% of the total variation in educational attainment and 16.66% in income inequality in Sweden. Therefore, even though Sweden is characterized by lower levels of inequality in the life chances of individuals, we conclude that a larger part of IOp can be identified by quantifying spatial circumstances. Finally, we show that 1.95% of fair inequality of income is attributable to spatial effort (i.e. own-neighbourhood). This points out the percentage of additional income that can be generated by moving to better environments.

This study makes several contributions to the existing research: i) via the comprehensive register data we cover a large set of circumstance variables, often not available due to data restrictions; ii) adapting several bespoke neighbourhood characteristics provides robust, externally valid and policy-oriented identification of the spatial factors as affecting opportunities; iii) the measure of inequality of opportunity is associated with parental neighbourhood characteristics for the first time in this study, and through a multilevel modeling strategy the results are robust with respect to previous studies where problems such as spatial autocorrelation have been ignored.

2 Previous Work

Despite the increasing number of empirical papers assessing the degree and nature of inequality of opportunity in different contexts none, to the best of our knowledge, includes parental neighbourhood into typically used parental background attributes. Some associate geography-of-birthplace as a circumstance, but this is often limited to Urban/Rural classifications (see Ferreira et al., 2011) or to very large administrative units such as regions (see Cogneau &

Mesplé-Somps, 2008; Singh, 2012). Others partition the study area into fewer but bigger macro regions and evaluate inequality of opportunity separately (see Peragine & Serlenga, 2008; Checchi & Peragine, 2010). Among empirical EOp studies in which geography is taken into account, the work of Checchi et al. (2010) comes closest to what might be perceived as neighbourhoods. It provides a comparison between the inequality of opportunity levels in 25 European countries where the degree of population density in residential areas is included as circumstance variables.

The lack of spatial considerations in the current literature is probably related to a general lack of spatially coded data available to researchers. For many countries even a very limited amount of data in relation to parental background may not be available. Another reason might be the fact that a spatial approach requires the recognition of the link between the geography of residence and any opportunity distribution. As is often underlined by the scholars proposing variants of the equality of opportunity approach, children should not be held responsible for their choices in any way (de Barros, 2009; Björklund et al., 2012). Such a view is readily extendable to include the residential decisions of parents on behalf of their children. Therefore, given the context of equal opportunities literature there should be no objections to defining parental neighbourhood as a circumstance.

Several empirical studies investigate the effects of neighbourhoods on educational attainment (Garner & Raudenbush, 1991), drop-outs (Crane, 1991) and outcomes such as reading, maths achievement (see Ludwig et al., 2013) and higher education participation (Andersson & Malmberg, 2015). However, the extensive body of neighbourhood literature shows no consensus on the durability of neighbourhood effects. At least two distinctive empirical strategies seek to investigate whether neighbourhood characteristics continue to be effective years after the exposure ends.

The first strategy studies the correlation between siblings and neighbouring children in adult outcomes, a way of quantifying the variation/correlation in earnings that can be attributed to neighbourhood histories. For instance Page & Solon (2003) study the correlation between adult earnings of once neighbouring children and between brothers, where the defined neighbourhood contains 20-30 contiguous dwelling units. Their findings demonstrate a positive correlation in earnings among formerly neighbouring children (half of the correlation observed for brothers), and is interpreted as residential immobility i.e. the children who grew up in urban areas end up in urban areas where the earnings are higher and those who grew up in rural areas remain in rural areas with lower earnings. Using a similar approach Raaum et al. (2006) find declining neighbourhood effects on earnings and educational attainment as years go by.

The second empirical strategy makes use of so-called moving to opportunity (MTO) experiments, which are randomized social experiments on housing mobility conducted by the U.S. Department of Housing and Urban Development (HUD). Ludwig et al. (2013) show that moving to less disadvantaged neighbourhoods (census tracts) improves both mental and physical health conditions. However such an impact is not observed for economic conditions and educational attainment even for the children who were exposed to a better environment at an early age. In a recent study Chetty et al. (2015) conclude that for children, each additional year spent in less deprived neighbourhoods (U.S. counties) increases the likelihood of college attendance and of higher earnings in adult life [1].

Although the above papers differ in important respects in how they study neighbourhood effects, they all identify a neighbourhood as an area comprising a predefined administrative unit. A few studies make use of bespoke, individualized neighbourhood units. For example Ham et al. (2014) adopt this method for the Stockholm metropolitan area and show that the negative effects of neighbourhood history are both inherited and persistent

[1](see also Chetty & Hendren, 2015)

over time. Similarly when investigating a population of parental home-leavers in Stockholm, Hedman et al. (2015) observe negative effects of exposure to a poverty-concentration parental neighbourhood even after 17 years of living away from parents.

Throughout this study we refer to a reference year in which the statistics for neighbourhood histories are linked to individuals. The above and other papers (Quillian, 2003; Clark & Ledwith, 2006) use the same empirical technique. If there is high residential mobility where both within and between neighbourhood shifts occur, a one-year reference might lead to measurement errors. However, we believe that our reference to a single year for parental neighbourhood does not bring large measurement error bias given the high similarity in peoples neighbourhoods overtime not only in Sweden (see Ham et al., 2014) but also in many other countries (see Kunz et al., 2003; Quillian, 2003; Sharkey & Faber, 2014).

This study seeks to bridge the gap between the literature dedicated to the theory and methods of equality of opportunities and to neighbourhood effects. For the educational IOp investigation we start from previous empirical studies on neighbourhood effects, but given the lack of consensus on the durability of such effects, for the income IOp investigation we first show how the neighbourhood histories of individuals exert persisting effects on life chances, therefore their contribution to inequality should be quantified and accommodated in a matrix devoted to circumstances.

3 Data

This study uses the PLACE longitudinal database (located at the Department of Social and Economic Geography, Uppsala University) which contains socio-economic, demographic and geographical information for all Swedish residents since 1990. Following the same individuals over time, we investigate the distribution of compulsory examination marks in 2001 and the distribution of disposable incomes in 2010 for the whole 1985 cohort as our variables of interest. Two sets of independent variables are considered in the model: circumstances as measured by parental background and parental neighbourhood characteristics and effort as measured by educational level and own neighbourhood characteristics.

For each individual, we use spatial and aspatial information from the dataset (see Table 2). The aspatial set of variables includes several covariates typically used in EOp studies that are informative of the family background and other inherited circumstances. Eight such variables are used: gender, compulsory examination marks , disposable income for 16-year olds residing in the household of upbringing (measured as part of household disposable income), whether or not a visible minority (VM, here understood as all individuals born outside Europe, USA, Canada or Australia), parent's marital status (single parent or dual parent households), parental education and employment status. The parental educational level is measured as the highest educational level reached by either of the parents. Employment status is measured as each parent working or not working in 2001.

The spatial set of variables includes parental neighbourhood characteristics in 2001 and own neighbourhood in 2010. These are quantified using a k nearest neighbour (knn) algorithm. Generating a form of scalable egocentric neighbourhood, this technique departs from each residential location and begins counting in every direction until a threshold (k) is reached. It then relates the population involved to the total counted population. The method does not require the use of predetermined administrative units and thus provides an efficient, comparable and robust definition of place (Östh et al., 2015). The computations were carried out using EquiPop software (Östh, 2014), which sorts people (in this case) according to a georeferenced grid and generates contextual variables quantifying the share of a given attribute within their k nearest neighbours, including for large data sets such as ours.

Table 2 [About Here]

We channel the following parental neighbourhood characteristics from 2001: the share of similar-age peers among the nearest 40 neighbours that accounts for socialization and network patterns, the share of visible minority (VM) neighbours among the nearest 400 neighbours show the degree of segregation and deprivation, the share of single parents and families with 3 or more children (large families) among the nearest 40 neighbours account for household and housing characteristics. We use smaller k-levels for the year 2001 as the potential interaction with the neighbourhoods might be limited compared to 2010. see Table 3 for an interpretation of different k values.

In addition to these bespoke neighbourhoods, the negative environment surrounding the parental neighbourhood is constructed based on Corine (coordination of information on the environment) data, which is available as 100-meter pixel raster images. ArcGIS software is used to match the land cover data to the coordinates of individuals (both available as 100x100 geo-coordinates) and after a classification of good/bad elements of Corine, the exposure to negative surroundings within a 500m radios is imported into the data as a column vector.

The own neighbourhood in the year 2010 is defined as the share of VMs among the nearest 1600 neighbours and a measure of job/housing balance is computed for 2010 as follows: we first classify individuals according to their level of education and the jobs available to them under three categories: low, intermediate and high. Then for each residential location, the nearest 10,000 jobs and the longest distance to reach the workplace are computed by EquiPop. The assumption is that individuals seek jobs that correspond to their level of education. Thus for a lower educated person this method looks for available jobs in the low category alone. The observed Cartesian distance between home and work can be used as a crude measure of job accessibility at any location i. However, since some of the studied individuals were not in employment in 2010 and others may have travelled distances that are considerably different from others residing in close proximity, data from near neighbours need to be interpolated. In order to depict a local commuting distance that renders potential commuting distances for non-commuters and renders commuting distances that reduce outlier effects for individuals with very short or long distances we employ a Kriging strategy where the 12 nearest neighbours constitute the search radius for the commuting distance interpolation surrounding any location where a population member resided in 2010. The smoothed interpolation expresses a commuting distance used as a representation of the potential commuting distance at any location i. [2].

4 Analytic Framework

As in Ferreira & Gignoux (2011), we model compulsory examination marks as a function of circumstances (reduced form equation) as follows:

$$g_i = f(C_i, u_i,) \tag{1}$$

and the disposable income as a function of circumstances and effort as follows:

$$y_i = f(C_i, E_i(C_i, v_i), u_i,) \tag{2}$$

$$E_i = BC_i + v_i \tag{3}$$

where g_i is compulsory examination grade of i, y_i represents disposable income, C_i a vector of circumstances and E_i is of effort, finally u_i is unobserved determinants of disposable income such as luck. We recognize the correlation between effort on circumstances and other unobserved determinants with equation (3).

In general (1) and (2) are estimated by OLS regressions (see Bourguignon et al., 2007; Ferreira & Gignoux, 2011). In this study we employ a

[2]Kriging was conducted in ARCInfo using the ordinary spherical semivariogram method, k=12

multilevel model with linear specification. It is obvious that spatial factors play a key role in this study. For this reason we need to specify a model that caters for the spatial patterns of variation that may lead to erroneous inferences. By employing the Morans I test on the regression residuals we can test if there are any spatial dependencies not catered for in the specified models (Moran, 1950). Four models were tested: (1) full model OLS, (2) empty model MLM, (3) contemporary model MLM and (4) full model MLM. Model results reveal that the OLS and empty models fail to take the spatial autocorrelation into account. That the empty model fails to explain variation is expected since no parameters are included, but that the full OLS model lies comparatively close to the empty model and far from remaining models clearly indicates that using OLS does not cater for the spatial variation present in the dataset. Of the remaining two models, the contemporary model is the one with no spatial autocorrelation, whilst the full model displays a weak but significant spatial autocorrelation. The chief difference between models explains why the full models show spatial autocorrelation. In the contemporary model, individual level parameters as well as contemporary contextual variables are introduced. Variables and the multilevel approach account for the spatial variation in regression residuals. However, in the full model, contextual variables from the year 2001 are also included. The variables introduced improve the overall model fit (see Table 1) but also introduce a spatial bias related to the sorting of individuals during the years of upbringing.

Table 1 [About Here]

We specify the empirical models as:

$$g_{ij} = a_0 + a_{ij}C_{ij} + a_j x_j + t_j + q_{ij} \tag{4}$$

$$y_{ij} = \beta_0 + \beta_{ij}C_{ij} + \alpha_{ij}E_{ij} + \beta_j x_j + u_j + z_{ij} \tag{5}$$

$$E_{ij} = b_0 + b_{ij}C_{ij} + b_j x_j + v_j + e_{ij} \tag{6}$$

for individual i living in municipality j, g_{ij} represents the log of compulsory examination marks, y_{ij} is the log of disposable income, β_0 and a_0 are the intercepts, x_j represents municipality-level covariates, t_j and u_j are municipality-specific random effects. In order to measure income IOp, the correlation between circumstances C_{ij} and E_{ij} effort needs to be examined. Again we follow Roemer (1998) in treating the effort variables, because a fundamental aspect in this setting is the fact that the distribution of effort within each circumstance group is itself a characteristic of that type; since this is beyond individual control, it constitutes a circumstance. [3]. Therefore only genuine effort $\hat{e}_{ij}$ must be derived. Finally we estimate the following model:

$$y_{ij} = \beta_0 + \beta_{ij} C_{ij} + \alpha_{ij} \hat{e}_{ij} + \beta_j x_j + u_j + z_{ij} \tag{7}$$

where $\hat{e}_{ij}$ is the estimate obtained in (6).

Using the estimates from reduced form equation (4) and from the full model (7), we construct a counterfactual distribution of compulsory examination marks g_{ij} and of disposable income y_{ij} where all variation within circumstance groups is eliminated as follows:

$$g_i^c = exp[C_i \hat{a}_{ij}] \tag{8}$$

and

$$y_i^c = exp[C_i \hat{\beta}_{ij}] \tag{9}$$

Subsequently the absolute and relative inequality of opportunity measures are calculated both with a path-independent decomposable inequality index, namely the mean logarithmic deviation (MLD) and with the Gini index as $IO = I(g_i)$ and $IO = I(y_i)$. Following this procedure, we can see how much of the inequality is due to inequality in opportunities and the share attributable to effort.

$$EIOp = \frac{I(g_i^c)}{I(g_i)} \tag{10}$$

and

$$IOp = \frac{I(y_i^c)}{I(y_i)} \tag{11}$$

[3]see Jusot et al. (2013) for other approaches

Using the same techniques we further decompose the relative contributions of spatial and aspatial factors to both circumstances and effort partitions of inequality. This practice is able to pinpoint the extent to which neighbourhoods influence given outcomes.

$$EIOp_{spatial} = \frac{I(g_{ij}^{sc})}{I(g_{ij}^{c})} \tag{12}$$

In a similar manner for earnings inequality:

$$IOp_{spatial} = \frac{I(y_{ij}^{sc})}{I(y_{ij}^{c})} \quad \text{and} \quad IO_{spatial} = \frac{I(y_{ij}^{se})}{I(y_{ij}^{e})} \tag{13}$$

Therefore, $IOp_{spatial}$ quantifies the relative contribution of parental neighbourhoods to overall inequality due to circumstances and $IO_{spatial}$ indicates the relative contribution of own-neighbourhood to overall inequality due to effort.

5 Findings

The main goal of this study is to investigate the spatial sources of inequality in relation to neighbourhood characteristics to which individuals are exposed. In this section we first briefly report the regression results of two models on educational attainment and disposable income respectively, then we show the outcomes of inequality decomposition into circumstances/effort and spatial/aspatial components. Before we proceed with the inequality decomposition, we verify the temporal extent of parental neighbourhood histories.

Table 4 shows the marginal effects of circumstance variables on compulsory examination marks. We employed the following 10 circumstance variables including the spatial covariates: gender, minority status (VM or not), the highest level of education attained by parents and employment and the marital status of parents, disposable income in 2001 and the share of the following attributes in the neighbourhood (k levels in parenthesis): single headed families (k=40), same age children (k=40), families with at least 3 children (k=40) and visible minorities (k=400). All coefficients have expected signs and are statistically significant at the 0.001 level (p values=0.00), except the share of same-age peers in the neighbourhood that is also significant but at the 0.05 level with a negative coefficient sign. This result may be interpreted as the distraction impact of having same-age peers in the neighbourhood due to the longer time spent on non-school activities. However we acknowledge the importance of socialization for the development of children. Besides, as mentioned in the following paragraphs, having same-age peers in the parental neighbourhood is positively associated with the subsequent disposable income of adults.

Of the remaining variables, living in a VM-concentrated area shows the strongest effect on marks. A likely explanation for the strong effect is that the share of VMs in a neighbourhood might coincide with the poverty concentration and other possible adverse characteristics of the locality. This can be seen from the maps in Fig.1, containing all residential coordinates aggregated to 100m x 100m units for the greater Stockholm area, where on the left hand side the VM population in the 400 nearest neighbours for the whole population is shown and on the right hand side the poverty concentration(OECD criteria) among the 400 nearest neighbours is mapped for the whole Stockholm metropolitan area. These maps show how the two aspects of the locality are statistically entangled, so that almost the same pattern of segregation is observed for both attributes of neighbourhoods.

Furthermore, positive effects are observed for students with employed parents (slightly higher if the mother is employed), with the presence of at least one highly educated parent at home and for students with high disposable incomes. Negative effects are found for students with a single parent and those who belong to VMs. The estimates for the single-parent and large families in the neighbourhood as well as negative surroundings show a negative association with educational attainment although the multilevel model controls the variability both at the individual and municipality level. The negative effect of single-mother concentration in neighbourhoods is a well explored phenomenon especially in the US, the single-parent specification in our model shows a similar pattern in Sweden. Due to lower household income, these families reside in worse neighbourhoods, therefore the variable performs as a proxy of residential environment and housing conditions to which the study population is exposed. A similar interpretation can be given to large-family concentration. Since large apartments are not found in the central districts, these families reside in rural areas or areas with rural character far from amenities. Furthermore, with three or more children, mothers stop studying at an early age, therefore large-family concentrated areas might be characterized by lower human capital accumulation.

Table 4 [About Here]

To analyze disposable income, we added gender to the circumstance variables from the previous model with the following effort variables: compulsory examination marks, job/housing balance(k=10000) and observed commuting distance between job opportunities and individual residences and the share of VMs (k=1600) in 2010. The variable for the highest level of educational attainment among parents interacts with ten city classes to account for different degrees of industrialization in cities as well as functions in terms of population density, commuting and remoteness following a classification used by SALAR (Swedish Association of Local Authorities and Regions) [4]. We examine the correlation between effort and circumstances by equation (3) for all effort variables. This procedure guarantees that the effort variables reflect only pure effort, without the influence of observed circumstances. Then we substitute the resulting residuals terms in equation (2). Now, the circumstances in

[4]Multilevel regression showed a negative association with the disposable income of individuals and their parental education. To correct this we used a classification that sorted municipalities by the degree of industrialization, population density, commuting and remoteness

(5) are expected to reflect both their direct impacts on the response variable and indirect effects on 5 effort variables. An important result to note is that as we regress the VM concentration of own-neighbourhood in 2010 on the circumstance variables, most of the variation is explained by the VM share of the parental neighbourhood. It is apparent from this result that the study population ended up in similar neighbourhoods as their parents. This residential immobility, or similarity in neighbourhood characteristics over time, points to long-term exposure to whatever effects neighbourhoods produce and the likelihood that these effects are transmitted to offspring. People may sort into neighbourhoods because of hedonistic motivations (quality services etc.) or because they might prefer to live with similar people. On the other hand, some neighbourhoods might be well (or less well)-endowed in terms of public goods and services because certain income/education groups and taxpayers live in those areas. We do not attribute any causal links between the two. However, from the equal opportunities perspective, being locked-in parental neighbourhoods (or to those with similar characteristics to parental neighbourhoods) is clearly a factor influencing life chances. For this reason, even though we deem individuals responsible for their choice of neighbourhood, it seems reasonable to derive pure effort purged of the influence of parental neighbourhood and other circumstances through the procedure explained above.

Table 5 shows regression results of the multilevel model for the log of disposable income. All the variables are statistically significant and those in common with the previous model show the same association with disposable income, except the share of same-age peers, which now has a positive coefficient sign. This is an interesting result that shows how the effects of residential contexts may differ over time. Growing up in an area with the strong likelihood of interaction with similar-age peers probably increases the chances of finding a better job and of being successful when employed through childhood ties.

Of the aspatial circumstance variables, being a female, belonging to VMs and having a single parent are negatively associated and disposable income in 2001, parental education and employment positively associated with disposable income. As for the spatial circumstances (i.e. parental neighbourhood attributes), the strongest effects are found for the share of VMs in the parental neighbourhood in 2001. Furthermore, growing up in areas with a high proportion of single-parent and large families and having negative environmental surroundings are negatively associated with subsequent earnings. Turning now to the discussion of the long-term effects of neighbourhoods, we are in position to conclude that historical neighbourhood characteristics influence adult earnings even though a range of individual and household characteristics are present in the model.

The only aspatial effort variable employed was compulsory examination marks. It shows a relatively lower effect on the response variable, which is in part attributable to the fact that all effort variables were purged of any influence of circumstances, as explained in the previous section. Regarding the spatial effort variables (i.e. own-neighbourhood characteristics) living in a neighbourhood with a high proportion of VMs and the observed commuting-distance between home and job are negatively associated and the degree of job/housing balance positively associated with disposable income. Since the spatial mismatch hypothesis first advanced by Kain (1968), there has been great interest in understanding differences in unemployment and job search success rates, job accessibility and job/housing mismatch (see for example Kain, 1992; Van der Klaauw & Van Ours, 2003; Houston, 2005). We are unaware of any application of job/housing (mis)match using the k nearest neighbour algorithm, which in return accounts for both residential location-driven and skill-based job accessibility.

Table 5 [About Here]

In terms of variation, the multilevel model indicates that the variance in disposable income is largely attributable to individuals. However it is important to remember that the fixed part of the model includes not only individual-level effects but also contextual variables that are defined individually both in parental neighbourhoods and when living independently of parents. Moreover, the 1% variation is explained by the municipality level covariates. If people were forced to live in certain municipalities, we would conclude that this value is the inequality of opportunity produced by Swedish municipalities. But since individuals are less restricted in their ability to choose where to live, we consider them responsible for their choice of municipality.

Fig.2 shows the variables over which the decomposition is undertaken and the second and third columns of Table 6 illustrate the magnitude of income and educational inequality of the entire Swedish population born in 1985. Based on the estimated coefficients from equations (4) and (7), the overall opportunity share in total inequality in income is computed as 8.05% and the overall opportunity share in educational inequality is 42.63% as measured by MLD. Note that since the effort partition contains both the effort and the unexplained part of disposable income variation and only the unexplained part of variation in educational inequality, the IOp estimates are lower bound.

As far as disposable income inequality is concerned, the relative decomposition of circumstances shows that 16.66% of the total share of circumstances is attributable to spatial circumstances (parental neighbourhood) and 83.33% is to aspatial circumstances. While the fair inequality decomposition indicates that the 01.95% of the total effort is due to spatial effort (own-neighbourhood) and the remaining part is caused by aspatial effort (compulsary examination marks) and the residual of the model. The corresponding decomposition for educational inequality shows that the spatial circumstances (parental neighbourhood) represent 36.44% of total circumstances. Even though earnings and educational attainment are two different outcomes, we can conclude that the neighbourhoods are more influential while the exposure is ongoing.

Table 6 [About Here]

We also conducted separate analyses for gender. The estimates of total inequality in educational attainment and income and related decomposition results are shown in Table 7 for females and Table 8 for the male population. A higher income IOp is observed among men than women. However, the relative income IOp is almost identical for both. The latter result seems to be related to the spatial sources of IOp. For women parental neighbourhood is more influential than for men (27.27% for women compared to 16.66% for men). On the other hand, spatial effort counts more for men than for women with 5.21% and 4.06% respectively. Regarding the IOp in educational attainment, overall circumstances explain a higher degree of variation for male students. Again this result seems to be related to parental neighbourhoods. For male students 30.15 percent of the total circumstance pertains to spatial circumstances, it is only 13.75 percent for female students.

Comparing the results in Table 7 concerning the effects of spatial circumstances, parental neighbourhood is more influential for the educational attainment of male students, hence during exposure. Once in employment, male students seem to be more successful in overcoming these effects through spatial effort than the female population. That is to say the male population uses *mobility* as an instrument at their disposal to generate additional income and to decrease the gap with higher income groups. In line with the findings for the whole population, for men the influence of parental neighbourhood proportionally decreases from 2001 to 2010. However, it is interesting to note that for the female population parental neighbourhoods cause a lower variation in marks during exposure and a higher variation in subsequent earnings than males. One interpretation of this pattern is that while being exposed to characteristics explained above for parental neighbourhoods, female students might manage to focus on their studies and reflect the adverse effects of parental neighbourhoods to a relatively lower degree to their marks. Yet, as females grow up, they might be building personal identities similar to that of the residents of their parental neighbours. This is a relevant interpretation especially given the fact that neighbourhood statistics for single-parent and large families mostly relate to women. Another view relates to mobility patterns. For the female population, parental neighbourhoods potentially become own-neighbourhoods since they seem to be immobile.

Table 7 [About Here]

Table 8 [About Here]

6 Concluding Remarks

A society is said to be equal in opportunities if the life chances of individuals do not depend on the factors beyond their choice and effort and the systematic differences in any outcome that are explained by so-called circumstances is considered as inequality of opportunity. It has been shown repeatedly that parental background plays an important part in the life chances of individuals. So far, however, there has been no discussion of the role of parental neighbourhood as a source of illegitimate inequality.

In this paper investigating the inequality in educational and earnings opportunities in Sweden for the whole 1985 cohort, we included parental neighbourhood statistics in a matrix of circumstance variables and own neighbourhood characteristics in a matrix devoted to effort variables. We constructed egocentric neighbourhoods where a count of k-nearest population forms the neighbourhood and the overall share of individuals who carry given characteristics identifies the likelihood of interactions. In addition, the share of negative components surrounding parental residence and a measure of housing/job market balance and the observed commuting distance between own-neighbourhoods and job opportunities were added to the analyses. Instead of the standard OLS approach, we utilized a multilevel model, which overcame most of the spatial autocorrelation problem.

The findings indicate that as well as the conventional aspatial circumstance variables, parental neighbourhoods strongly impact educational attainment and even years after exposure they remain influential for earnings distribution. Therefore, based on the evidence from Swedish data, we can conclude that in order to obtain accurate measures of IOP, there is a definite need for a multidisciplinary approach that links individual outcomes to neighbourhoods.

We hope that our findings may influence the way in which IOp analyses are conducted both in terms of methods and techniques to quantify and include characteristics of neighbourhoods. We are aware that the latter requires detailed information on geo-locations. This is the central policy implication of our study, that data collection methods must be designed to contain necessary geographic information on residents. The recent developments in data collection methods associated with "big data" significantly facilitate the collection of contextual variables. For instance there is a vast quantity of information that is made freely available on internet through open maps and the social-media data provides a wealth of information to researchers. Therefore, what is left is the proper handling of geography. The findings of this study recommend using bespoke neighbours to define an individual's residential environment. Creating individualized neighbourhoods based on the k nearest neighbour algorithm enabled us to overcome problems associated with administratively defined areas plagued by indeterminacy. Through this approach, this paper has gone some way towards enhancing our understanding of the temporal effects of neighbourhoods.

Furthermore, our results show that the opportunity gap between individuals widens both for visible minorities and for the residents of visible minority-concentrated neighbourhoods. Therefore, another important implication specific to Swedish data is that in order to decrease inequality in opportunities, an effective policy must target the population belonging to the visible minority population and their residential environment. Observed negative effects with proxies of housing conditions suggest a need for comprehensive analysis of segregation not only by nominal categories but also by income.

References

Andersson, E. K., & Malmberg, B. (2015). Contextual effects on educational attainment in individualised, scalable neighbourhoods: Differences across gender and social class. *Urban Studies*, *52*(12), 2117–2133.

Björklund, A., Jäntti, M., & Roemer, J. E. (2012). Equality of opportunity and the distribution of long-run income in sweden. *Social choice and welfare*, *39*(2-3), 675–696.

Bourguignon, F., Ferreira, F. H., & Menendez, M. (2007). Inequality of opportunity in brazil. *Review of income and Wealth*, *53*(4), 585–618.

Brunori, P., Ferreira, F. H., & Peragine, V. (2013). *Inequality of opportunity, income inequality, and economic mobility: Some international comparisons.* Springer.

Buck, N. (2001). Identifying neighbourhood effects on social exclusion. *Urban studies*, *38*(12), 2251–2275.

Case, A. C., & Katz, L. F. (1991). *The company you keep: The effects of family and neighborhood on disadvantaged youths* (Tech. Rep.). National Bureau of Economic Research.

Checchi, D., & Peragine, V. (2010). Inequality of opportunity in italy. *The Journal of Economic Inequality*, *8*(4), 429–450.

Checchi, D., Peragine, V., & Serlenga, L. (2010). Fair and unfair income inequalities in europe.

Chetty, R., & Hendren, N. (2015). The impacts of neighborhoods on intergenerational mobility: Childhood exposure effects and county-level estimates. *Unpublished Manuscript*.

Chetty, R., Hendren, N., & Katz, L. F. (2015). *The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment* (Tech. Rep.). National Bureau of Economic Research.

Clark, W. A., & Ledwith, V. (2006). Mobility, housing stress, and neighborhood contexts: evidence from los angeles. *Environment and Planning A*, *38*(6), 1077–1093.

Cogneau, D., & Mesplé-Somps, S. (2008). Inequality of opportunity for income in five countries of africa. *Research on Economic Inequality*, *16*, 99–128.

Crane, J. (1991). The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American journal of Sociology*, 1226–1259.

de Barros, R. P. (2009). *Measuring inequality of opportunities in latin america and the caribbean*. World Bank Publications.

de Barros, R. P., Vega, J. R. M., & Saavedra, J. (2008). Measuring inequality of opportunities for children. *Washington, DC: World Bank*.

Ellen, I. G., & Turner, M. A. (1997). Does neighborhood matter? assessing recent evidence. *Housing policy debate*, *8*(4), 833–866.

Ferreira, F. H., & Gignoux, J. (2011). The measurement of inequality of opportunity: Theory and an application to latin america. *Review of Income and Wealth*, *57*(4), 622–657.

Ferreira, F. H., Gignoux, J., & Aran, M. (2010). Inequality of economic opportunity in turkey: An assessment using asset indicators and womens background variables. *State Planning Organization of the Republic of Turkey and World Bank Welfare and Social Policy Analytical Work Program Working Paper*(3).

Ferreira, F. H., Gignoux, J., & Aran, M. (2011). Measuring inequality of opportunity with imperfect data: the case of turkey. *The Journal of Economic Inequality*, *9*(4), 651–680.

Galster, G. (2001). On the nature of neighbourhood. *Urban studies*, *38*(12), 2111–2124.

Garner, C. L., & Raudenbush, S. W. (1991). Neighborhood effects on educational attainment: A multilevel analysis. *Sociology of education*, 251–262.

Ham, M., Hedman, L., Manley, D., Coulter, R., & Östh, J. (2014). Intergenerational transmission of neighbourhood poverty: an analysis of neighbourhood histories of individuals. *Transactions of the Institute of British Geographers*, *39*(3), 402–417.

Hedman, L., Manley, D., Van Ham, M., & Östh, J. (2015). Cumulative exposure to disadvantage and the intergenerational transmission of neighbourhood effects. *Journal of Economic Geography*, *15*(1), 195–215.

Houston, D. (2005). Employability, skills mismatch and spatial mismatch in metropolitan labour markets. *Urban Studies*, *42*(2), 221–243.

Jusot, F., Tubeuf, S., & Trannoy, A. (2013). Circumstances and efforts: How important is their correlation for the measurement of inequality of opportunity in health? *Health economics*, *22*(12), 1470–1495.

Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *The Quarterly Journal of Economics*, *82*(2), 175–197.

Kain, J. F. (1992). The spatial mismatch hypothesis: three decades later. *Housing policy debate*, *3*(2), 371–460.

Kunz, J., Page, M. E., & Solon, G. (2003). Are point-in-time measures of neighborhood characteristics useful proxies for children?s long-run neighborhood environment? *Economics Letters*, *79*(2), 231–237.

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes. *Psychological bulletin*, *126*(2), 309.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2013). *Long-term neighborhood effects on low-income families: Evidence from moving to opportunity* (Tech. Rep.). National Bureau of Economic Research.

Openshaw, S. (1984). The modifiable areal unit problem..

Östh, J. (2014). *Introducing the equipop software–an application for the calculation of k-nearest neighbour contexts/neighbourhoods.*

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

Page, M. E., & Solon, G. (2003). Correlations between brothers and neighboring boys in their adult earnings: The importance of being urban. *Journal of Labor Economics*, *21*(4), 831–855.

Peragine, V., & Serlenga, L. (2008). Higher education and equality of opportunity in italy. *Research on Economic inequality*, *16*, 67–97.

Quillian, L. (2003). How long are exposures to poor neighborhoods? the long-term dynamics of entry and exit from poor neighborhoods. *Population Research and Policy Review*, *22*(3), 221–249.

Raaum, O., Salvanes, K. G., & Sørensen, E. Ø. (2006). The neighbourhood is not what it used to be*. *The Economic Journal*, *116*(508), 200–222.

Ramos, X., & Van de Gaer, D. (2012). Empirical approaches to inequality of opportunity: Principles, measures, and evidence.

Roemer, J. E. (1998). *Theories of distributive justice.* Harvard University Press.

Roemer, J. E., & Trannoy, A. (2013). Equality of opportunity.

Ross, J. L., Iannotta, J. G., et al. (2002). *Equality of opportunity and the importance of place:: Summary of a workshop.* National Academies Press.

Sharkey, P., & Faber, J. W. (2014). Where, when, why, and for whom do residential contexts matter? moving away from the dichotomous understanding of neighborhood effects. *Annual Review of Sociology*, *40*, 559–579.

Singh, A. (2012). Inequality of opportunity in earnings and consumption expenditure: The case of indian men. *Review of Income and Wealth*, *58*(1), 79–106.

Van der Klaauw, B., & Van Ours, J. C. (2003). From welfare to work: does the neighborhood matter? *Journal of Public Economics*, *87*(5), 957–985.

Vartanian, T. P. (1999). Adolescent neighborhood effects on labor market and economic outcomes. *Social Service Review*, *73*(2), 142–167.

Tables and Figures

Table 1: Moran's I Tests

	Full model OLS	Empty model MLM	Contemporary model MLM	Full model MLM
Moran's I	0,005457	0,005947	0,000181	0,00104
Expect I	-0,000025	-0,000025	-0,000025	-0,000025
z-score:	13,437288	15,425446	0,534853	2,98011
p-value:	0,00000	0,00000	0,592751	0,002881

Table 2: Variables

	Variables	Description
1. Individual	Gender	1=Female, 0=Male
	Visible Minority	1=VM, 0=Not a VM
	Compulsory Exam marks	
	Family Background	Parents' Employment Status and Education, single-headed household
2. neighbourhood	Share of Visible Minorities	2001(k=400) and 2010(k=1600)
	Single-Parent Families	2001(k=40)
	Share of families with 3 or more children	2001 (k=40)
	Share of same-age peers	2001(k=40)
	Negative Space	Corine database 2001 (500m radii-based)
	Housing/Job Market Balance	2010 k=10000 and commuting distance

Table 3: K-neighbours

k-neighbours	Possible Interactions
12	Stairs in building
25	Building
50	Bicycle basement, garbage recycling bins etc.
100	Block
200	Bus stop
400	Kiosk, familiar with topology, recognize all neighbours
800	Football field
1600	Small shop
3200	Day care, school
6400	Local square, different retail stores, dentist...
12800	Upper secondary schools, Big stores, communities (sports, religion)
25600	Hospital, place-belonging, municipality

Table 4: Multilevel Model: Log of marks (Compulsory Exam)

Fixed	Coef.	Standard Error	P values
1.Individual			
Employment Father	0.2361	0.0180	0.000
Employment Mother	0.2614	0.0170	0.000
Parental Education	0.1694	0.0123	0.000
Single Parent	-0.2322	0.0143	0.000
Visible Minority	-0.1702	0.0313	0.000
Disposable Income(2001)	0.1310	0.0151	0.000
2.neighbourhood			
Single-Headed Families(2001)_20	- 0.2323	0.0143	0.000
Large Families(2001)_20	-0.2253	0.0468	0.000
Same-Age peers_20	-0.2399	0.1136	0.035
Negative Space	-0.2063	0.0411	0.000
Visible Minority(2001)_200	-1.0354	0.0937	0.000
Random Effects Parameters	Estimate	Standard Error	
Municipality Level var(_cons)	0.0094	0.0020	
Var(Residual)	2.9919	0.0139	
Number of obs = 92674			

Table 5: Multilevel Model: Log of Disposable Income

Fixed	Coef.	Standard Error	P values
1.Individual			
Gender	-0.1644	0.0030	0.000
Compulsory Exam marks	0.0188	0.0008	0.000
Employment Father	0.0714	0.0047	0.000
Employment Mother	0.0784	0.0044	0.000
Parental Education			
x CityClass2	0.0634	0.0108	0.000
x CityClass3	0.0252	0.0081	0.002
x CityClass4	0.1121	0.0223	0.000
x CityClass5	0.0808	0.0151	0.000
x CityClass6	0.0508	0.0220	0.021
x CityClass7	0.1039	0.0144	0.000
x CityClass8	0.0968	0.0303	0.001
x CityClass9	0.0850	0.0135	0.000
x CityClass10	0.0845	0.0200	0.000
Single Parent	-0.0249	0.0037	0.000
Visible Minority	-0.1152	0.0081	0.000
Disposable Income(2001)	0.1211	0.0041	0.000
2.neighbourhood			
Large Families(2001)_40	-0.0382	0.0123	0.020
Single-Headed Families(2001)_40	-0.0789	0.0113	0.000
Same-Age peers_40	0.0790	0.0296	0.002
Negative Space	-0.0291	0.0107	0.007
Visible Minority(2001)_400	-0.3598	0.0248	0.000
Job/Housing Balance(2010)_10000	0.2074	0.0129	0.000
Commuting Distance(2010)	-0.0117	0.0007	0.000
Visible Minority(2010)_1600	-0.6481	0.0205	0.000
Random Effects Parameters	Estimate	Standard Error	
Municipality Level var(_cons)	0.0024	0.0003	
Var(Residual)	0.2045	0.0009	
Number of obs = 91413			

Table 6: Inequality Decomposition

	Income Inequality				Educational Inequality		
Total Inequality (GINI)	0.2674				0.1749		
Total Inequality(MLD)	0.1315				0.2667		
Inequality of Opportunity(GINI)	0.0695				0.1528		
Inequality of Opportunity(MLD)	0.0106				0.1137		
	Effort		Circumstances		Effort	Circumstances	
Contibuiton(%) to Total inequality (MLD)	91.95%		8.05%		57.37%	42.63%	
	Aspatial (residual)	Spatial	Aspatial	Spatial		Aspatial	Spatial
Spatial/Aspatial (MLD)	98.05%	1.95%	83.33%	16.66%	residual	63.06%	36.94%

Table 7: Inequality Decomposition Female Population Only

	Income Inequality(Female)				Educational Inequality (Female)		
Total Inequality (GINI)	0.2495				0.1657		
Total Inequality(MLD)	0.1105				0.2552		
Inequality of Opportunity(GINI)	0.0548				0.1436		
Inequality of Opportunity(MLD)	0.0058				0.1091		
	Effort		Circumstances		Effort	Circumstances	
Contibuiton(%) to Total inequality (MLD)	94.76%		5.24%		57.24%	42.76%	
	Aspatial (residual)	Spatial	Aspatial	Spatial		Aspatial	Spatial
Spatial/Aspatial (MLD)	95.94%	4.06%	72.73 %	27.27%	residual	86.25%	13.75%

Table 8: Inequality Decomposition Male Population Only

	Income Inequality(Male)				Educational Inequality (Male)		
Total Inequality (GINI)	0.2729				0.1784		
Total Inequality(MLD)	0.1423				0.2747		
Inequality of Opportunity(GINI)	0.0623				0.1601		
Inequality of Opportunity(MLD)	0.0082				0.1342		
	Effort		Circumstances		Effort	Circumstances	
Contibuiton(%) to Total inequality (MLD)	94.02%		5.80%		51.12%	48.88%	
	Aspatial (residual)	Spatial	Aspatial	Spatial		Aspatial	Spatial
Spatial/Aspatial (MLD)	94.79%	5.21%	50.00 %	16.66%	residual	69.85%	30.15%

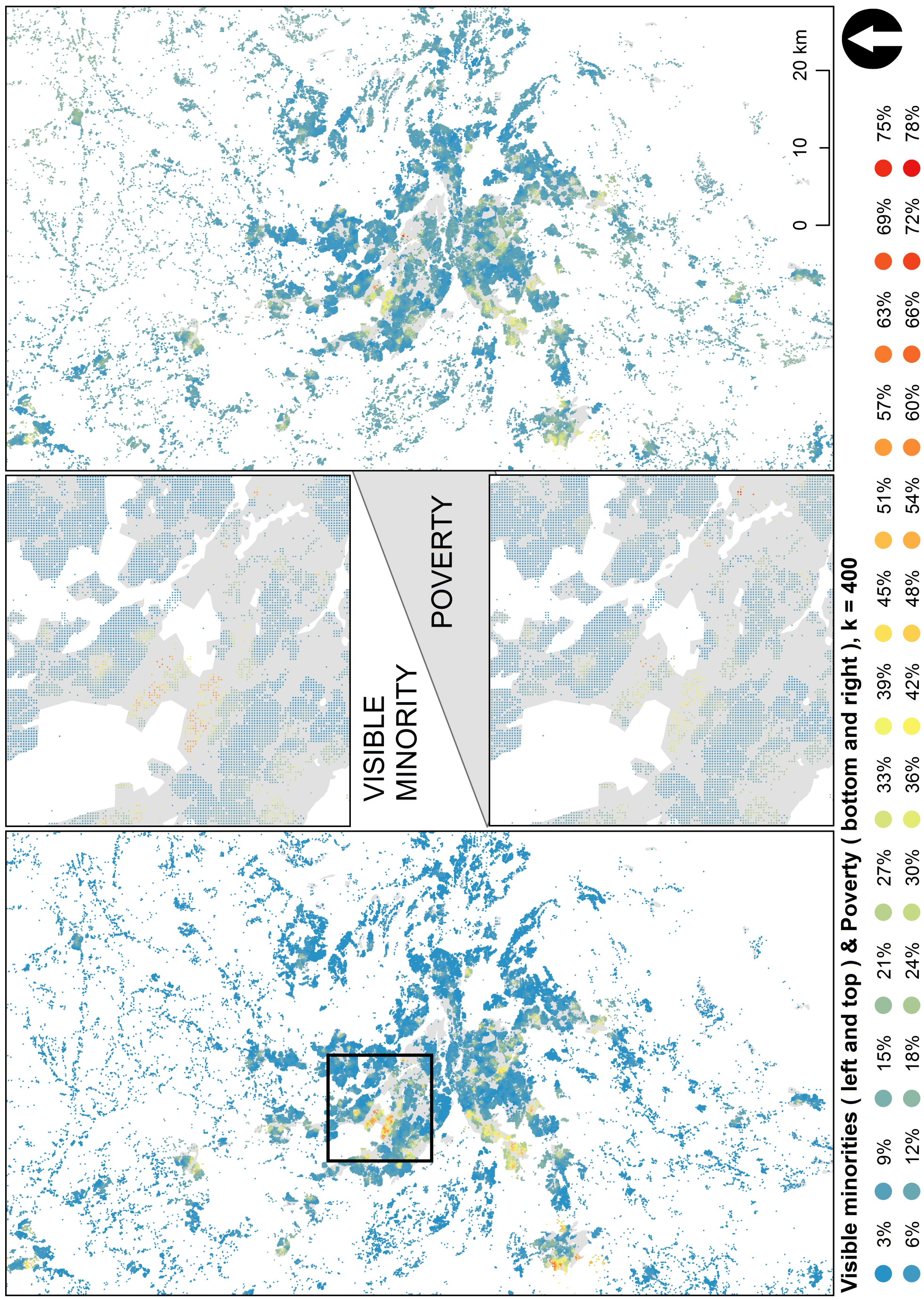
Visible minorities (left and top) & Poverty (bottom and right), k = 400
3% 9% 15% 21% 27% 33% 39% 45% 51% 57% 63% 69% 75%
6% 12% 18% 24% 30% 36% 42% 48% 54% 60% 66% 72% 78%
VISIBLE MINORITY
POVERTY
0 10 20 km

Figure 1:

	Circumstances	Effort
Educational IOp	❑ Aspatial • Visible Minority • Parents' Employment Status • Parental Income • Parental Education ❑ Spatial • **Parental Neighbourhood**	• Residual
Income IOp	❑ Aspatial • Visible Minority • Gender • Parents' Employment Status • Parental Income • Parental Education ❑ Spatial • **Parental Neighbourhood**	❑ Aspatial • Compulsory Education Grades • Residual ❑ Spatial • **Own-Neighbourhood**

Figure 2: Decomposition

Chapter 4

A Gini Measure of Spatial Segregation by Income

A Gini measure of Economic Segregation *

Umut TÜRK

Abstract

This paper proposes a new measure of spatial segregation by income that uses the Gini index as the basis of measurement. Gini Index of spatial segregation (GSS) is a ratio of two Gini indices comparing inequality between neighbourhoods to inequality between individuals at the macro area where neighbourhoods are nested. Unlike other measures of income segregation found in literature, the index uses individualized neighbourhoods. Depending on the population density and proximity between individuals, an individualized neighbourhood is defined both as an area constituted within a radius or as a population-count method around an individual geo-location. The GSS is suitable for the measurement of residential segregation by any continuous variable. It is sensitive to spatial configuration of areas, easy to compute and interpret and suitable for the comparative studies of segregation over time and across different contexts. An empirical application of the index is illustrated using data from Sweden covering the entire population in the years 1994, 2004 and 2014. We show how the definition and scale of the neighbourhood dramatically affect the measures of economic segregation.

*This chapter is based on a joint work with Eugenio Peluso (DSE, University of Verona, Italy) John ÖSTH (Department of Social and Economic Geography, Uppsala University, Sweden), Francesco Andreoli(Luxembourg Institute of Socio-Economic Research, LISER, Luxembourg)

1 Introduction

The measurement of residential segregation by income (hereinafter Economic Segregation) has attracted relatively less attention than measurement of residential segregation by race or occupation. Economic and racial segregation share many factors in common: both are distinctively spatial phenomena, may occur from similar dynamics and are often empirically entangled (Reardon, 2011). Whereas the literature studying income segregation faces the challenge of measuring segregation along a continuous dimension, hence it cannot easily borrow indices from the racial segregation literature. Racial segregation refers to the uneven distribution of people belonging to different groups across physical space, while economic segregation amounts to quantify the income homogeneity or diversity in the areas of residents.

The efforts to identify segregation in urban spaces is not new to literature. Previous studies by sociologist, economists and geographers theorized the distinctive distribution of social classes in cities. The patterns of segregation among people by ethnicity, race, social class and among institutions, roads and variety of economic activities would be observed in cities growing radially from a core as rings outwards (The Concentric Zone model by Burgess (1928)) or as star-shaped, sector base (The Second Theory Model by Hoyt (1939)) or in a multiple core fashion, where the number and size of cores (nuclei) vary highly for different cities prevailing from historical development, geography and culture (Multi-core Model by Harris & Ullman (1945)). Most residential segregation studies use these theories as a basis to understand urban structure. The similar patterns of residential segregation has been confirmed repeatedly. The effects and causes of the segregation have been shown to prevail from households sorting across neighbourhoods with differential public goods/services that are excludable for location (Tiebout, 1956; Epple et al., 1984). Similarly, by the preferences of neighbourhood racial composition (Schelling, 1969), by education and income (Jargowsky, 1996), by exogenous factors such as changes in spatial distribution of opportunities and due shifts from manufacturing activities to service-oriented economies (Morenoff & Tienda, 1997) and by demographic changes: female participation in the market, aging population thus changes in demographic composition of cities (Wyly et al., 1998). Moreover, the effects of segregation have been shown to be related to inequality in growth (Burgers & Musterd, 2002; Reardon & Bischoff, 2011) and the distribution of top 1% income in the space (Essletzbichler, 2015). However, the scale in which these factors become evident has started to attract scholarly attention only recently.

To quantify the relative clusters of people by socioeconomic characteristics, any measurement has to aggregate them into some spatial unit so called "neighbourhood". The measurements are likely to vary depending on the definition of neighbourhood chosen. Especially, when relied on some predefined administrative unit such as census tracts or municipalities findings can be erroneous. This is what has become known as the modifiable areal unit problem (MAUP) (Openshaw, 1984; Wong, 2004). The MAUP occurs both with the scale problem, where the same data portrays different spatial patterns for its varying levels of aggregates,and with the zoning problem, where altering the grouping schemes produce different results even if the units are of the same scale. In particular to racial segregation analysis, the scale problem has been recognized and addressed in several ways (Wong, 1993, 1999, 2005; Reardon et al., 2008, 2009). Since the residential segregation by definition relates to clusters of people, the way in which the geography is handled becomes not only a statistical issue but also a crucial strategy to study the effects and causes of segregation. A way to address this issue is to construct scalable egocentric local environments. Depending on the definition of neighbourhood either a set of varying radius (see Lee et al., 2008, for racial segregation) or varying population sizes around an individual location so called k nearest neighbours (knn) used (see Östh et al., 2015, for interaction among racial groups).

The aim of the present work is to construct an index to perform a comprehensive investigation of residential segregation by income with a particular attention to the scale problem. The proposed index compares the inequality between individualized-neighbourhoods to the inequality between individuals at the macro area where neighbourhoods are nested. It is flexible in deciding the definition and size of neighbourhoods.

Section 2 recalls some relevant tools used in income segregation measurement. Section 3 presents the new GSS index. Section 4 provides an empirical application to Swedish context. Neighbourhoods are considered both as the area constituted within a radius that is drawn around each individual location and as the nearest population-count approach using the information of residential coordinates. Observing the diverse patterns in segregation measurements in response to the definition of neighbourhood chosen, we propose a variant of k nearest neighbour algorithm that makes use of spatial weights matrix (see Getis, 2009). In return, this new approach communicates both the spatial distribution of individuals and the population density constituted in each neighbourhood. The paper shows how the definition and scale of the neighbourhood influence the measures of economic segregation and the use of individual neighbourhoods permits managing the related weakness of the traditional tools and to obtain robust results.

2 The Gini indices of segregation

Despite the problems associated with the dissimilarity index, it remains the most diffused index of segregation. It measures the degree to which the minority proportions of areal units differ from the minority proportion of the whole area in absolute terms (James & Taeuber, 1982). Hence, the dissimilarity index is designed to measure the unevenness in the distribution of two population groups. The fundamental issue associated with the index is that it is sensitive to the share of minority population in different spatial units and to the size of the overall areal unit (Cortese et al., 1976). On the other hand, it is insensitive to the reallocation of minority groups among areal units where minority proportion is less or higher than the overall area's minority proportion (James & Taeuber, 1982). Even so, the dissimilarity index is widely used for the measurement of economic segregation in a similar manner to racial segregation. Typically, the index computes the uneven distribution of two population groups defined under and above to a given income threshold. However, this approach discards considerable amount of information hidden underneath income distributions (Abramson et al., 1995; Massey et al., 2003, see).

There are many other alternative indices found in literature: the index of exposure, relative concentration, absolute centralisation and spatial proximity (Massey & Denton, 1988), nearly all inequality indices can measure dichotomous /categorical segregation (Kim & Jargowsky, 2005). These include entropy and Atkinson indices for analyses of evenness in distributions.

In the present paper we focus on the Gini segregation indices. In its original form, the Lorenz curve is a representation of sorted cumulative percentage of total income as a function of cumulative percentage of total households (Lorenz, 1905). Whereas, the Gini coefficient is a measure of the area between Lorenz curve and the line of perfect equality, normalized by the total areal under the 45 degree line. As for the measurement of racial segregation, the index performs similar to the dissimilarity index, where the Gini coefficient represents the area between Lorenz curve normalized by the total area under 45 degree line for the minority populations weighted across all pairs of areal units (Massey & Denton, 1988). It takes a maximum value 1 when the minority and majority members of the society are perfectly segregated and 0 for no segregation. This form of the Gini is limited to measure the segreation along two population groups only. Reardon & Firebaugh (2002) proposed extensions of six segregation indices measuring multi-group segregation, including the Gini. As a function of the disproportionality in group proportions across organizational units, the index is interpreted as the weighted sum of the weighted average absolute difference in group proportions between all possible pairs of units over multigroups. Kim & Jargowsky (2005) developed a version of the Gini segregation index that accounts for continuous variables. By this extension the Gini becomes suitable for the measurement of income segregation, where the Gini for income disproportionality among neighbourhoods is normalized by the individual level Gini.

The forms of the Gini index listed above do not account for the spatial configuration of areas, therefore, are subject to the problems associated with MAUP. Also to "checkerborad phenomenon" occuring when an index is insensitive to spatial proximity between areas(White, 1983). A spatial version of the Gini is proposed by Dawkins (2004), it measures racial segregation given the spatial proximity of neighbourhoods. The decomposition of the index produces within and between components and also a residual term that captures the correlation of neighbourhood's own position and the position of its neighbours when ranked with some proximity among neighbourhoods. Extending the standardized spatial Gini index , Dawkins (2007) proposed spatial ordering index calculated from either a nearest neighbour or a monocentric spatial ordering of neighbourhood per capita income and the Gini index of between-neighbourhood income segregation. The index represents a ratio of two covariances where numerator is the covariance between neighbourhood aggregate income and spatial reranking of neighbourhood whereas denominator is the covariance between neighbourhood aggregate income and the average ranking of neighbourhood. However, the index does not address the scale issue. Table 1 shows these Gini segregation indices with their corresponding papers.

3 The GSS Index

In this section we introduce the Gini index of spatial segregation (GSS). Given a population of N individuals, let y_i be the income of individual i and μ_{is} be the average income in individual i's neighbourhood. The latter can be either radii-based (considering people comprised within a circle of a given radius drawn around individual i) or a count of the k nearest neighbours around i's location. Therefore, the shape (s) of the neighbourhood varies for the definition chosen and the size (n_{is}) can be set to meet various scales of geography. Note that for any population size $n_{is} \neq N$, μ_s will differ from μ.

The GSS is a weighted measure of income homogeneity/diversity in the areas of residents. It is defined as the ratio of two Gini indices as follows:

$$GSS(y, n) = \frac{\frac{1}{N^2 \mu_s} \sum_i \sum_j |\mu_{is} - \mu_{js}|}{\frac{1}{N^2 \mu} \sum_i \sum_j |y_i - y_j|} \tag{1}$$

where

$$\mu_{is} = \frac{\sum_{j \in s} y_j}{n_{is}} \tag{2}$$

The GSS index is the ratio of the between neighbourhoods inequality I_B to the individual level inequality I_G. It takes a minimum of 0 (no segregation) in two case scenarios: if the numerator is zero thus the between spatial inequality is zero or when the size of the neighbourhood is equal to the size of the whole study area: $n_{is} = N$. While the index takes maximum value 1 (perfect segregation) if the distribution of individualized-neighbourhood average incomes is identical to that of individual incomes thus when $I_B = I_G$ or when the size of the neighbourhood $n_{is} = 1$ every neighbourhood consists of one individual only.

The GSS measures the extent to which neighbourhoods differ from each other in terms of income, without concerning the fairness in income distributions. In a given area, neighbourhoods might be populated solely by lower income groups. In such a case the GSS shows lower values. This implies a lower residential segregation by income for this particular area.

The index has several advantages. First, in contrast to Kim & Jargowsky (2005) the index is sensitive to the spatial configuration of neighbourhoods so that it overcomes the checkerboard problem. Second, it does not require the use of predefined administrative units, hence it is not subject to the MAUP. Finally as a ratio of two Gini indices GSS preserves several properties of the index. It respects to the Pigou-Danton principles of transfers, it is less sensitive to outliers, deviations from normality and it is suitable for the segregation measure of continuous variables. Finally normalizing the index by individual level income provides an ideal research environment for comparative studies among different contexts and over time.

4 Empirical Application

In this section we test how the GSS performs with an empirical application to Swedish register data for the years 1994, 2004 and the latest year available in Place database 2014. The database contains the disposable income and residential coordinates of entire population. A common problem shared by previous studies of segregation is the choice of predefined administrative units of analysis. This study uses residential the coordinates that are available for each individual as 100x100m grid units in the database.

We begin by creating individualized-neighbourhoods for each individual's geo-location in the country. As far as the approach to the neighbourhood is concerned, the best method that renders the spatial boundary of the neighbourhood must be chosen. The optimal method potentially varies among different context and especially with the population density at the interested area. Both the radii and knn approaches have their advantages and disadvantages. The radii-based neighbourhood depicts the geography constituted within a predetermined radius, therefore the space as the point of interest. This is a desirable way of studying geography when the analyses focus on locations, services: parks, recreational areas and when there are no changes expected in terms of population density in the space defined. But when the concern is the spatial relationship between individuals, knn approach might be more appropriate. As a population-count method knn successfully illustrates the interaction possibilities between individuals, when the areas are not populated too sparsely. In this paper we make use of both approaches and we propose an intermediate method that benefits the advantages of two.

First, we construct neighbourhoods based on a *knn* algorithm. Each neighbourhood contains the average disposable income earned by nearest 100 200 400... 51200 working-age (20-64) people for each residential location as follows: $\mu_{ik} = \sum_i \frac{y_i}{k}$. Since we work on the entire Sweden, the physical separation between neighbours is an issue that we have to handle especially for the northern parts of the country. For this reason, we introduce a distance decay model $f(\beta, d_{in})$: a function of distance between individual i and their k nearest neighbours $n = 1, 2, 3, ..., k$ with a distance decay factor β. This operation spatially weights the observations, so that as the distance between i and k nearest neighbours increases ks' relative contribution to the average income decreases. Therefore, for densely populated metropolitan areas, the neighbourhood average incomes remain similar to those produced by *knn* algorithm without a decay factor. The computations are carried out by the EquiPop software (Östh, 2014). The EquiPop permits introducing a decay factor prior to computations start and produces outputs both with and without spatial weigths in an efficient way.

Since the daily interaction behavior of residents is not feasible to extract from the data and given the spatially disaggregate nature of income distributions, β is derived mathematically by half-life models[1] as 0.0001153

[1]Common way to determine distance decay factor is to use spatial interaction models(SIMs) where observed flows of people between origins (O) and destinations(D) are regressed over distances between all possible O-D pairs. This operation reduces the overall deviation from the mean commuting distance in a given population. Mathematically derived half-life models (HLMs) are valid alternatives to SIMs when the flows of people are not observed and in the presence of spatially highly disaggregate data. HLMs use median value as departure. This is because the median commuted distance always occur at a distance where half of the population commute longer and half of the population commute shorter distance. Therefore, knowing the maximum distance from i to its kth neighbor, we can say that the probability of interacting with neighbours equals 0.5 at the observed median distance. Then for the decay function to describe this probability at various distances, the probability-value

with an exponential decay function. Then the spatial weights function becomes $\exp(-\beta d_{in})$ for each pair of neighbours.

Table 2 shows the GSS values for three years where each value corresponds to a different size of neighbourhood and every second column of a given year reports the index weighted with decaying distance. While the last row shows the overall Gini of disposable income for each year. The slight increase in the Gini from 1994 to 2004 is reflected by the GSS measured at neighbourhood size $k = 100$, for larger k values instead a similar segregation pattern yields. Therefore, from 1994 to 2004 residential segregation by income has increased only at a very small geography i.e. among 100 nearest neighbours. Moreover, the GSS values for the year 2014 show that the increase in inequality at individual level is reflected at any scale of geography.

Furthermore, using the spatial analyst tools available in ArcGIS, we construct radii-based neighbourhoods where the average disposable income in i's neighbourhood is measured as $\mu_{ir} = \frac{\sum_{j \in r} y_i}{n_{ir}}$. Computations are repeated for the radius sizes: $r = 100m, 1km, 5km, 25km$. The estimates reported on Table 3 show increasing index values parallel to increase in individual level Gini over years. There is no direct equivalence between r and k in how much geography is depicted as we move from one definition to other. Fig.1 offers a useful picture how the GSS varies between years and for different neighbourhood sizes and definitions. On the left-hand side a similar pattern is observed for the years 1994-2004, whereas the GSS in year 2014 (grey line) lies above for all k values. Therefore, what we observe from knn approach is that the residential segregation by income remained at a similar rate from 1994 to 2004 despite a slight rise in the overall inequality and it increased in 2014. The radii-based approach (below) instead shows a clear ranking among years 1994, 2004 and 2014 with a similar response to different r values. Compering pictures in the Fig.1, it is evident that the radii-approach exhibits a higher level of segregation

---

will decay from one at no distance towards almost zero at far, far away (see Östh et al., 2016, for details).

than knn. But the GSS values with decay factor display a similar pattern to radius. Both show a decreasing at a decreasing rate for increasing size of neighbourhoods. The difference between k-nearest and radii-based neighbour approaches becomes clear as we move to the analysis at the municipality level.

To show how economic segregation varies by geographic locations within the country, we compute the GSS separately for 290 Swedish municipalities. Each value represents the ratio of the inequality between average incomes earned in the bespoke neighbourhoods of people who live in the same municipality and the total inequality in the country. We use both radii and k-nearest neighbour aggregates and for the k-nearest neighbour approach we reporte values both with and without a decay factor (=0.0001153). The results for the year 2014 with different radius and k values are shown in Fig.2. The colours correspond to the fixed intervals of GSS values for all maps. This makes easier to compare the values obtained by the two approaches to neighbourhoods.

By looking at the maps for smaller to higher r and k values, a lens scans economic segregation from block level to larger units of localities such as census tracks. Smaller r and k values may communicate a residential segregation in a couple of buildings and as the scale gets larger the GSS may communicate the economic segregation in an area including schools, stores, play grounds etc. The radius and knn approaches display different patterns especially for lower values of r and k. As stated above, the reason for this is that the radii-based approach focuses on the geography only, meaning that the number of people living within a given radius varies between locations (and time) and this is not catered for by this approach. This is evident especially on the first row of Fig.2 with r=100 meters, at this scale the GSS values are very high. They vary between 0.4-0.6 for all municipalities. Even for a small k value as 200 (may be equivalent to a block in a densely populated area), a much lower segregation is observed, close to zero in some municipalities but still retaining the high GSS values for metropolitan areas.

As opposed to radii, knn approach focuses on people and neglects how far they live from each other. This becomes a relevant issue especially for the sparsely populated areas in the northern parts of the country where kth neighbour might reside kilometers away from i. The second and the third rows of the figure offer a useful comparison for this respect. For a smaller value $k = 200$ both maps display a similar pattern, while for intermediate k levels, decayed GSS values capture some of the segregation pattern similar to radii approach. Therefore, the maps on the third row lie somewhere between radii and knn maps, rendering both the number and the geographic distribution of people. Fig.3 shows the change in GSS values from 1994 to 2014. The maps are organized in the same way as in Fig.2. In the northern parts of the country economic segregation mostly remained the same and decreased in couple of northern municipalities. While in the metropolitan areas such as Stockholm, Malmö and Göteborg, it has increased, even for higher aggregates of people i.e. for larger r and k values.

In the next step we report spearman[2] rank-order correlations between the computed GSS values and several characteristics of the municipalities to explore the properties of areas signaling the current state of the residential segregation by income. The municipalities are characterized based on the variables provided by Statistics Sweden. We use the information on the changes in employment by an application of the Martin resilience-employment index (Martin, 2012) measuring the growth rate of employment overtime and expressed as the change over time $\Delta = t - t + 1$ and employment levels at local (Er) and national level (EN).

$$REI_i = \frac{(\Delta Er/Er) - (\Delta EN/EN)}{|\Delta EN/EN|}$$

We find a significant positive correlation between REIs calculated for the time intervals 1993-1994 and 2013-2014 and the GSS indices for 1994

[2]Spearmans correlation coefficient is a statistical measure of the strength of a monotonic relationship between paired data.

and 2014, respectively. This can be interpreted as a degree of job/housing balance, that the population tend to cluster both as employed/unemployed, therefore employment growth is associated with economic segregation and by the earnings generated from different types of jobs. We find negative correlation with cost-equalization grants per municipality, this suggests that the equalization grants ensure a degree of residential mixture by income. Moreover, there is a significant positive correlation with election participation rates and strong positive correlation with the number of low/high educated people in areas. These two correspond to the sorting behavior of individuals by political orientation/politicization and education.

Fig.4 displays how the spearman correlations vary for the GSS indices computed for increasing k values in 1994 and 2014. In both years a constant association is observed with REI and election participation rates for any scale of geography, whereas in both years the correlation with cost-equalization spending and number of low/high educated people decreases after $k = 6400$. This is probably because at this scale we exceed the municipality size. Interestingly, despite the increase observed in economic segregation from 1994 to 2014 (see Table 2), the exact pattern of sorting by education is observed in the two years in response to varying scales of the neighbourhood.

Sweden has three levels of government; the central government(staten), county council (landstinget) and municipality (kommunen). All three are allowed to tax personal income and the municipalities provide public services and are subject to income, cost and structural equalization grants. The first two are purely re-distributive, the income equalization grants are transferred from the high income municipalities to the ones with lower incomes and the cost equalization grants are transferred to municipalities with less favourable cost structure from the ones with better conditions. The structural equalization instead is a grant from the central government to municipalities with a small population or a high share of unemployment.

The strong local government sector in Sweden offers a useful research environment for the study of economic segregation, particularly useful to possible policy implications for other contexts. Using OLS regressions, we test the effects of all three equalization grants took place in 2013 on the GSS values computed with *knn* (decayed) approach to neighbourhoods for the year 2014 (see Appendix 1). We find that the economic segregation decrease with both the income and structural equalization grants at any scale of neighbourhood. Moreover, looking at the cost equalization grants for different structural costs, again the transfers are associated with a degree of residential mixture by income. For instance, the grants to equalize costs of heating, public transportation, upper secondary and compulsory schools decrease the economic segregation in the following year. Whereas the municipalities that receive higher grants for the costs of streets&roads and the grants devoted to children with foreign background on average experience higher degree of economic segregation. A possible interpretation as far as the grants for streets&roads are concerned is that the municipalities in need of grants due their relatively less favorable infrastructure may show higher physical separation between classes. Similarly, the municipalities that are eligible to receive grants for children with foreign background constitute higher minority population that potentially cluster in neighbourhoods due to relatively lower earnings.

5 Conclusions

So far, the segregation measures found in literature have been mainly developed to measure the extent to which individuals are clustered by groups: typically race, ethnicity, gender in occupations. The residential segregation by income instead has not received much attention in the literature. The most of the existing studies of the latter use the indices originally developed for racial segregation by dividing population into two categories; being under and above to a given level of income i.e. poor and not poor. By restricting the analysis to two groups, the indices do not make full use of the available information. Moreover, nearly all existing indices are aspatial in nature, that they do not take into account the distribution of individuals in space. Although there exits few spatial ones, they are rather difficult to compute and nearly all use some administratively defined area for the unit of analysis.

In this paper, we offer a new measure of residential segregation by income based on a individualized-neighbourhood approach and therefore makes use of the full information on the income distribution of residents and their distribution in space. The proposed index allows to handle the geography flexibly as neighbourhoods can be constructed by both radii and *knn* approaches and the scale can be set to varying levels to meet distinctive characteristics of different contexts. This last point allows the index to avoid robustness issues associated with MAUP and checkboard phenomenon, the problems that may severely distort the sensitivity of the results of spatial analyses. Additionally as a ratio of two Gini indices, the index has the advantage of preserving desirable properties of the Gini. It respects to the Pigou-Danton principles of transfers, it is less sensitive to outliers, deviations from normality and finally it is suitable for the segregation measure of continuous variables.

Moreover, using the Swedish register data we have tested the efficiency of the index. We have used both approaches to individualized neighbourhoods and by employing spatial weights matrix based on the distance between neighbours we have proposed an intermediate approach that benefits the advantages of both. In particular to Sweden, the estimates suggest that the economic segregation has remained at a similar degree from 1994 to 2004. Although it has increased from 1994 in 2014 in parallel to rise in inequality, correlation analysis has shown that the individuals sort into neighbourhoods almost identically in both year. Additionally, the analysis on the influence of transfers among municipalities has illustrated that as the smallest local governments in Sweden, the municipalities can reduce the economic segregation through equalization grants of income, and several structural costs.

References

Abramson, A. J., Tobin, M. S., & VanderGoot, M. R. (1995). The changing geography of metropolitan opportunity: The segregation of the poor in us metropolitan areas, 1970 to 1990. *Housing Policy Debate*, *6*(1), 45–72.

Burgers, J., & Musterd, S. (2002). Understanding urban inequality: a model based on existing theories and an empirical illustration. *International Journal of Urban and Regional Research*, *26*(2), 403–413.

Burgess, E. W. (1928). Residential segregation in american cities. *The Annals of the American Academy of Political and Social Science*, *140*(1), 105–115.

Cortese, C. F., Falk, R. F., & Cohen, J. K. (1976). Further considerations on the methodological analysis of segregation indices. *American sociological review*, 630–637.

Dawkins, C. J. (2004). Measuring the spatial pattern of residential segregation. *Urban Studies*, *41*(4), 833–851.

Dawkins, C. J. (2007). Space and the measurement of income segregation. *Journal of Regional Science*, *47*(2), 255–272.

Epple, D., Filimon, R., & Romer, T. (1984). Equilibrium among local jurisdictions: Toward an integrated treatment of voting and residential choice. *Journal of public economics*, *24*(3), 281–308.

Essletzbichler, J. (2015). The top 1% in us metropolitan areas. *Applied geography*, *61*, 35–46.

Getis, A. (2009). Spatial weights matrices. *Geographical Analysis*, *41*(4), 404–410.

Harris, C. D., & Ullman, E. L. (1945). The nature of cities. *The Annals of the American Academy of Political and Social Science*, *242*(1), 7–17.

Hoyt, H. (1939). The structure and growth of residential neighborhoods in american cities. *Washington, D.C.: Federal Housing Administration*.

James, D. R., & Taeuber, K. E. (1982). *Measures of segregation.* JSTOR.

Jargowsky, P. A. (1996). Take the money and run: Economic segregation in us metropolitan areas. *American sociological review*, 984–998.

Kim, J., & Jargowsky, P. A. (2005). *The gini coefficient and segregation on a continuous variable.*

Lee, B. A., Reardon, S. F., Firebaugh, G., Farrell, C. R., Matthews, S. A., & O'Sullivan, D. (2008). Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review*, *73*(5), 766–791.

Martin, R. (2012). Regional economic resilience, hysteresis and recessionary shocks. *Journal of economic geography*, *12*(1), 1–32.

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social forces*, *67*(2), 281–315.

Massey, D. S., Fischer, M. J., Dickens, W. T., & Levy, F. (2003). The geography of inequality in the united states, 1950-2000 [with comments]. *Brookings-Wharton papers on urban affairs*, 1–40.

Morenoff, J. D., & Tienda, M. (1997). Underclass neighborhoods in temporal and ecological perspective. *The Annals of the American Academy of Political and Social Science*, 59–72.

Openshaw, S. (1984). The modifiable areal unit problem. *Geo Abstracts University of East Anglia*.

Östh, J. (2014). *Introducing the equipop software–an application for the calculation of k-nearest neighbour contexts/neighbourhoods.*

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

Östh, J., Lyhagen, J., & Reggiani, A. (2016). A new way of determining distance decay parameters in spatial interaction models with application to job accessibility analysis in sweden. *EJTIR*, *16*(2), 344–363.

Reardon, S. F. (2011). Measures of income segregation. *Unpublished Working Paper. Stanford Center for Education Policy Analysis*.

Reardon, S. F., & Bischoff, K. (2011). Income inequality and income segregation 1. *American Journal of Sociology*, *116*(4), 1092–1153.

Reardon, S. F., Farrell, C. R., Matthews, S. A., OSullivan, D., Bischoff, K., & Firebaugh, G. (2009). Race and space in the 1990s: Changes in the geographic scale of racial residential segregation, 1990–2000. *Social Science Research*, *38*(1), 55–70.

Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological methodology*, *32*(1), 33–67.

Reardon, S. F., Matthews, S. A., OSullivan, D., Lee, B. A., Firebaugh, G., Farrell, C. R., & Bischoff, K. (2008). The geographic scale of metropolitan racial segregation. *Demography*, *45*(3), 489–514.

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, *59*(2), 488–493.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of political economy*, *64*(5), 416–424.

White, M. J. (1983). The measurement of spatial segregation. *American journal of sociology*, 1008–1018.

Wong, D. W. (1993). Spatial indices of segregation. *Urban studies*, *30*(3), 559–572.

Wong, D. W. (1999). Geostatistics as measures of spatial segregation. *Urban Geography*, *20*(7), 635–647.

Wong, D. W. (2004). The modifiable areal unit problem (maup). In *World-minds: Geographical perspectives on 100 problems* (pp. 571–575). Springer.

Wong, D. W. (2005). Formulating a general spatial segregation measure. *The Professional Geographer*, *57*(2), 285–294.

Wyly, E. K., Glickman, N. J., & Lahr, M. L. (1998). A top 10 list of things to know about american cities. *Cityscape*, 7–32.

Tables and Figures

Table 1: Gini indices of Segregation [a]

	Gini Segreagation	Binary	Multigroup	Continuous	Spatial
Massey & Denton (1988)	$G = \sum_i \sum_j \frac{t_i t_j \lvert p_i - p_j \rvert}{2T^2 P(1-P)}$	✓			
Reardon & Firebaugh (2002)	$G = \sum_m \sum_i \sum_j \frac{t_i t_j}{2T^2 I} \lvert \pi_{im} - \pi_{jm} \rvert$		✓		
Kim & Jargowsky (2005)	$G = \frac{\frac{1}{2N^2\mu} \sum_i \sum_j \lvert y_{ni} - y_{nj} \rvert}{\frac{1}{2N^2\mu} \sum_i \sum_j \lvert y_i - y_j \rvert}$	✓		✓	
Dawkins (2007)	$G = \frac{Cov(Y_j, \hat{R}_{j(n)})}{Cov(Y_j, \hat{R}_j)}$			✓	✓

[a] T total population, ti and t_j total populations at i and j, P total minority population, p_i minority population at i, m number of groups, π_{im} proportion in m at i, y_i income of i household, y_{ni} average income at n, μ overall average income, N total number of households, Y_j aggregate income at j, $\hat{R}_{j(n)}$ spatial re-ranking of neighbourhood, $\hat{R}_j$ average ranking of neighbourhood

Table 2: GSS for different k values

k	GSS_94	GSSdecay_94	GSS_04	GSSdecay_04	GSS_14	GSSdecay_14
100	0,315	0,319	0,323	0,327	0,328	0,333
200	0,287	0,294	0,287	0,293	0,298	0,306
400	0,264	0,272	0,260	0,269	0,274	0,286
800	0,241	0,253	0,235	0,247	0,2548	0,270
1600	0,220	0,235	0,215	0,231	0,237	0,258
3200	0,200	0,220	0,199	0,218	0,221	0,248
6400	0,179	0,204	0,180	0,207	0,204	0,237
12800	0,156	0,188	0,158	0,192	0,182	0,222
25600	0,139	0,177	0,139	0,180	0,161	0,209
51200	0,121	0,172	0,120	0,177	0,140	0,205
GINI(Individual)	0,257		0,262		0,332	

Table 3: GSS for different r values

r	GSS_96(radius)	GSS_04(radius)	GSS_14(radius)
100m	0,436	0,475	0,497
1000m	0,290	0,327	0,373
5000m	0,212	0,264	0,302
10000m	0,186	0,236	0,267
25000m	0,158	0,186	0,209
GINI(Individual)	0,257	0,262	0,332

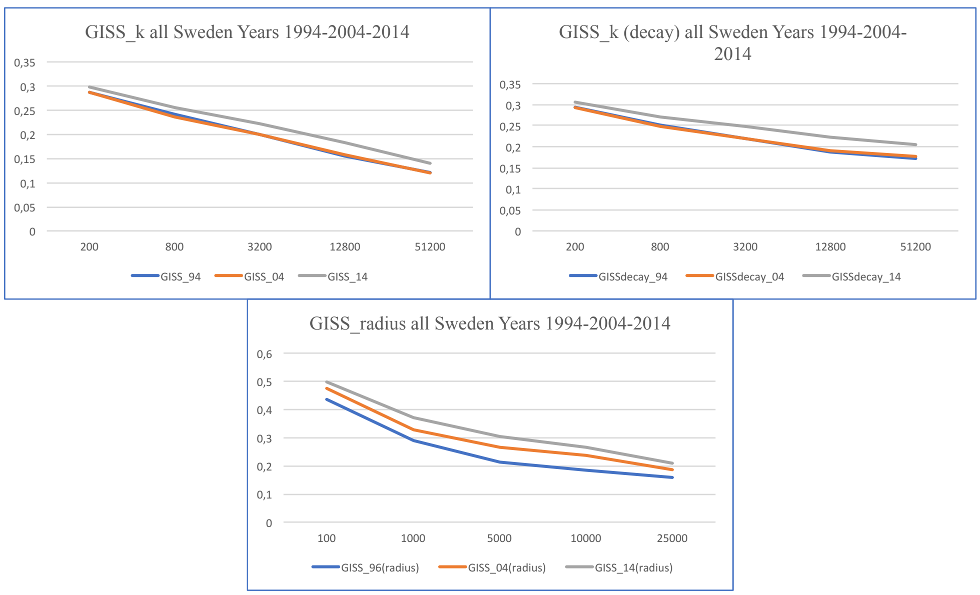
GISS_k all Sweden Years 1994-2004-2014
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
200
800
3200
12800
51200
GISS_94
GISS_04
GISS_14
GISS_k (decay) all Sweden Years 1994-2004-2014
0,35
0,3
0,25
0,2
0,15
0,1
0,05
0
200
800
3200
12800
51200
GISSdecay_94
GISSdecay_04
GISSdecay_14
GISS_radius all Sweden Years 1994-2004-2014
0,6
0,5
0,4
0,3
0,2
0,1
0
100
1000
5000
10000
25000
GISS_96(radius)
GISS_04(radius)
GISS_14(radius)

Figure 1:

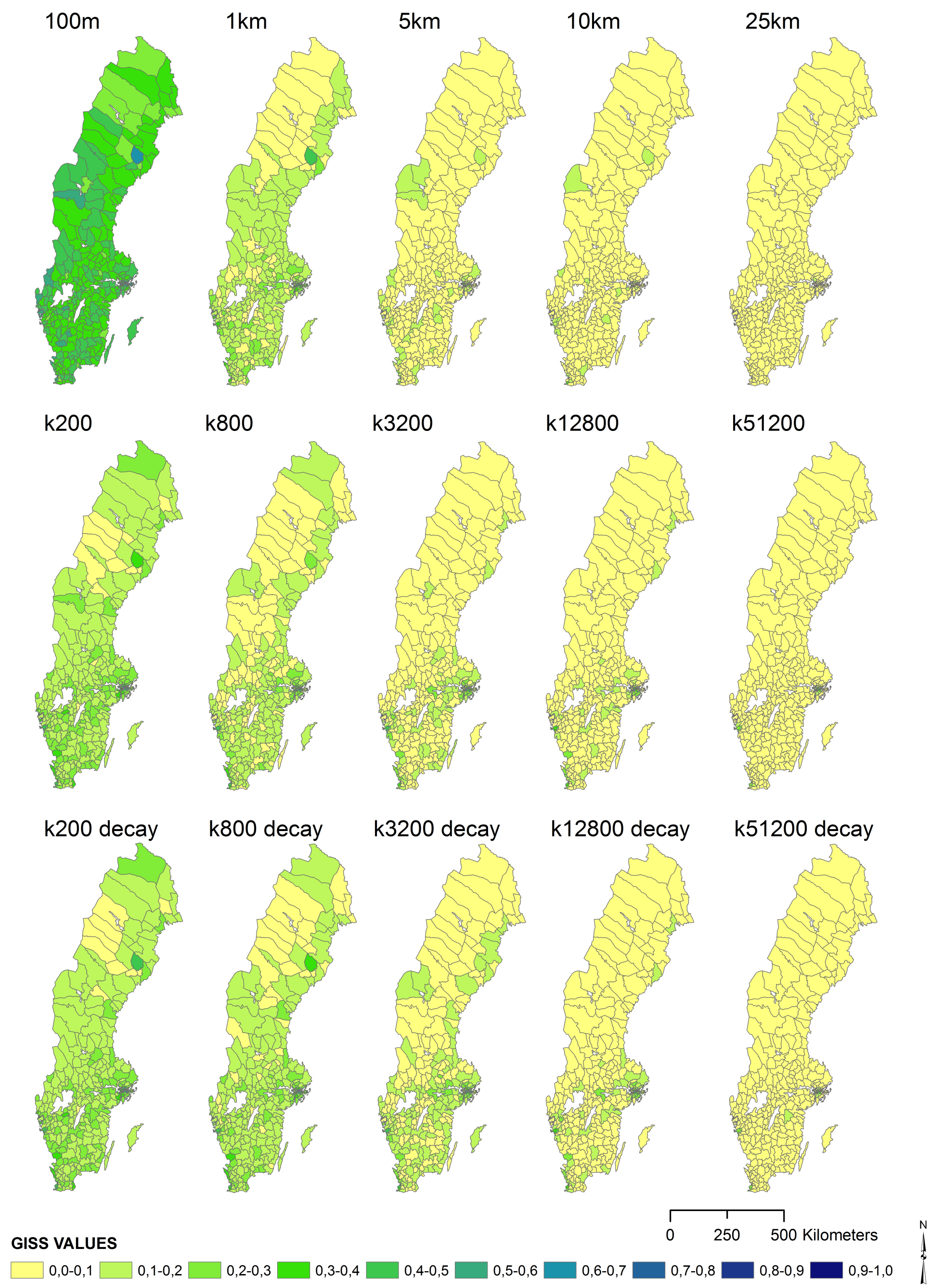

Figure 2: 2014 GSS values-first row radii-approach-second row knn-third row knn with decay

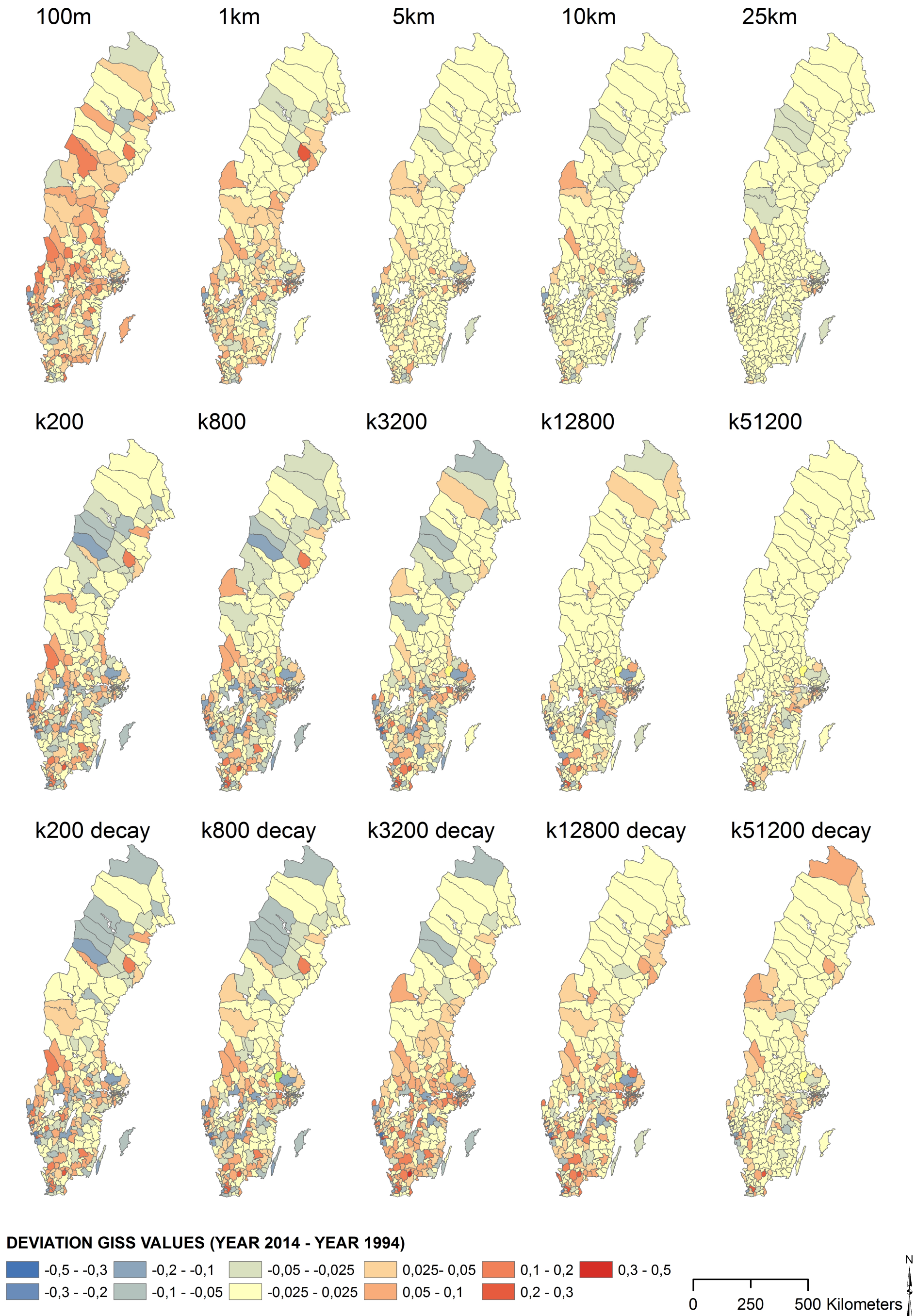

Figure 3: Deviation GSS values year 2014-year 1994. First row radii-approach-second row knn-third row knn with decay

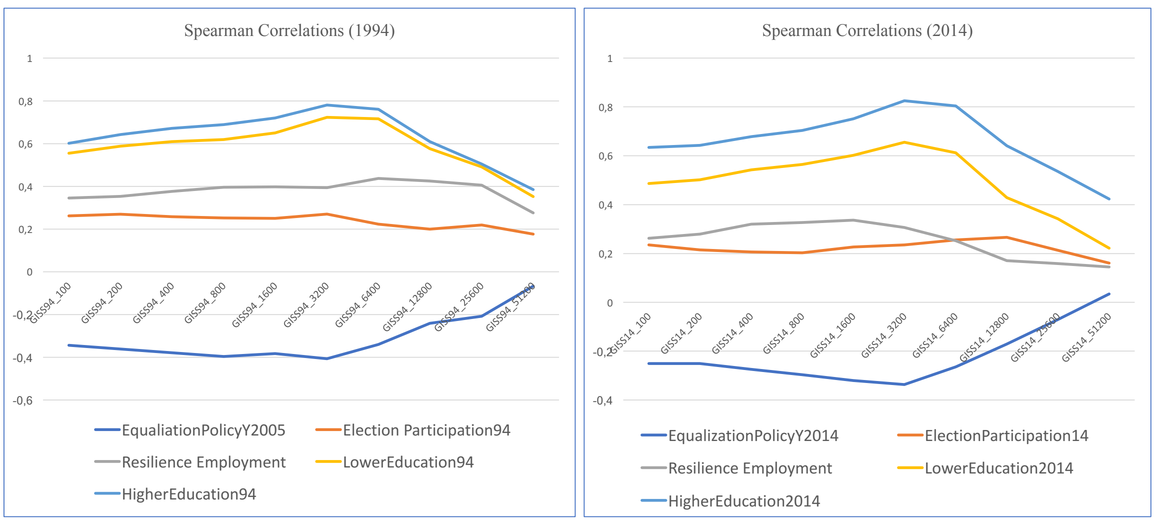
Spearman Correlations (1994)
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
GISS94_100
GISS94_200
GISS94_400
GISS94_800
GISS94_1600
GISS94_3200
GISS94_6400
GISS94_12800
GISS94_25600
GISS94_51200
EqualiationPolicyY2005
Election Participation94
Resilience Employment
LowerEducation94
HigherEducation94
Spearman Correlations (2014)
1
0,8
0,6
0,4
0,2
0
-0,2
-0,4
GISS14_100
GISS14_200
GISS14_400
GISS14_800
GISS14_1600
GISS14_3200
GISS14_6400
GISS14_12800
GISS14_25600
GISS14_51200
EqualizationPolicyY2014
ElectionParticipation14
Resilience Employment
LowerEducation2014
HigherEducation2014

Figure 4:

Appendix 1

[a]

	(1)	(2)	(3)	(4)
VARIABLES	GSS14decay_200	GSS14decay_800	GSS14decay_3200	GSS14decay_6400
Structural Eq. (2013)	-0.698***	-0.912***	-1.454***	-1.152***
	(0.212)	(0.235)	(0.240)	(0.238)
Income Eq. (2013)	-0.141***	-0.140***	-0.129***	-0.116***
	(0.0324)	(0.0337)	(0.0276)	(0.0302)
Cost Eq. (2013):				
Heating Eq. (2013)	-3.388*	-4.163**	-4.702**	-4.334**
	(2.020)	(1.950)	(1.850)	(1.884)
Streets&Roads (2004)	2.618**	3.167**	2.824*	3.085*
	(1.179)	(1.257)	(1.695)	(1.662)
Public transportation (2013)	-1.364***	-1.245***	-1.150***	-0.895***
	(0.282)	(0.289)	(0.305)	(0.330)
Elderly Care (2013)	-0.206**	-0.223***	-0.463***	-0.525***
	(0.0800)	(0.0847)	(0.0941)	(0.0954)
Foreign Children (2013)	5.739***	6.238***	7.288***	7.448***
	(1.221)	(1.320)	(1.501)	(1.571)
Upper secondary school (2013)	-0.408**	-0.534***	-0.541***	-0.697***
	(0.192)	(0.202)	(0.208)	(0.222)
Compulsory School (2013)	-0.591***	-0.617***	-0.958***	-1.016***
	(0.158)	(0.171)	(0.194)	(0.202)
Gini (2014)	2.053***	1.757***	1.210***	1.068***
	(0.150)	(0.156)	(0.134)	(0.146)
Constant	0.610***	0.910***	1.736***	1.498***
	(0.214)	(0.244)	(0.241)	(0.230)
Observations	290	290	290	290
R-squared	0.861	0.838	0.823	0.814

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

[a]Notes: i) Model controls for the Gini index computed at the municipality level. ii) Statistics Sweden provides information on the total amount of income and structural equalization grants transferred among total of 290 municipalities, for the cost equalization grants instead several structural factors are available. iii) Data source: http://www.statistikdatabasen.scb.se