An investigation of social capital in Britain using small area estimation analysis

M. K. Orteca

A thesis submitted for the degree of PhD

Department of Sociology

University of Essex

December 2015

ABSTRACT

Social Capital is considered an important asset for development, both at local and higher levels, and has been explored across the social sciences for decades. Attempts to define and measure it in increasingly precise ways continue in order to place it at the centre of policymaking. Indeed, it is considered a precious capital in times where resources are becoming scarce. This thesis investigates the dimensions of social capital and estimates them at small area level for England and Wales in 2011. The first step is the identification of three factors measuring membership, citizenship and politics and neighbourliness dimensions using survey data and a Confirmatory Factor Analysis. The second step is to test the hypothesis that other individual characteristics and geographical characteristics may influence levels of these factors. Complex Multilevel models with individual covariates and area-level covariates from the Census and administrative sources confirm the hypothesis: the factors depend on age, gender, ethnicity, religion, marital status, socio-economic class, employment, state of health and education at individual level and ethnic diversity and economic profile at area level. Lastly, Multilevel Model results have been used as a starting point for the final synthetic estimates at small area for all the Middle Super Output Areas of the average levels of the three factors. While Membership and Citizenship and Politics social capital show higher differentials, Neighbourliness seems to be more spread and, on average, higher than the other two factors across the two countries.

The phoenix hopes, can wing her way through the desert skies, and still defying fortune's spite; revive from ashes and rise.

Miguel de Cervantes Saavedra, *El ingenioso hidalgo don Quijote de la Mancha*,1605.

To my grandma Katia.
Sharing not only the same name.

A mia nonna Katia.
Condividendo non solo lo stesso nome.

ACKNOWLEDGMENTS

Soon after reading, one of my dearest friend told me she has never seen such a kind of acknowledgments. She added that that tradition wants them more formal. However, she also confirmed that these are the most honest she has ever read, and suggested to keep them. So this, here I am again. It seems that the choice (wrong or not, still to understand) to keep on investing on my education at least has the advantage that at the end of the path I can stop and think, remember, evaluate. Above all, be grateful. Many of you will not even read these acknowledgments. You may not even know that you made it to this acknowledgment. Nevertheless, it is important, for me to show my gratitude.

I will start with the official ones.

I thank Dr. Ben Anderson and Dr. Paola De Agostini. They believed in me and thank to them, this adventure started.

I thank the Department of Sociology, my supervisor Prof. Nick Allum and among the staff (academic and administrative) I especially thank Prof. Yasemin Soysal and Michele Hall. Because nothing is better than being considered like a person and not merely a student or a number or an administrative stuff. I thank ESRC and ONS for sponsoring me.

I also thank Dr. Nikos Tzavidis that accepted me as visiting student and helped me with his sincere attention. For the same reason, I thank Dr. Ian Brunton-Smith that gave me the right suggestions for a brilliant thesis.

And now, let's start with the more 'informal' ones.

The most important ones.

I cannot start differently. Thanks to my mum and my dad. For the unlimited, patient and silent support. You allowed me to leave home and come back every time, to get lost and find myself again without asking anything. I always knew that, no matter what, you would be there for me, respect my schedules, my moods and my silences. Thanks, because there is no place like home. And home is you.

Thanks to my crazy, little sister Chiara. Being far made our relation stronger and conspiratorial despite everything. We shared a lot: trips, arguments, tears and laughs. You are my sister and my friend.

I want to also thank all the other relatives, in Milan or in the South of Italy, that support me and always care about me: zia Anna, zia Lella, zio Roberto, zio Giuseppe, Luca, Marco, Marilena, Mariafilomena, zio Gianni e zia Rosanna.

I will remain in Italy and I start with my friends. Veronica and Viviana, My Divas.

I know you were waiting for my return. And here I am. I cannot wait for spending time with you again. You are my best friends, in everything.

Coming and waiting for you outside our ex-office has been always the restarting point of feeling at home again. We shared everything and I know I can count on you. I can only thank you for loving me. I also thank all the other friends from Iuppolis: Picci, Luigi, Alis, Elisa and Vale.

Then I thank the friends of a lifetime: Carla, Andrea e Ceci. Despite of our different life's stage, being with you all the times helped me a lot in making my choices. And despite of months without contacts, meeting you feels like we have never been apart.

I also thank Gloria, the other vagabond like me. Your 'amica mia' opened every times my hearth and no one can understand me better than you do.

I also thank Daniela 'F***' Briga for being so caring with my sister and, at the end, being a friend of mine. I thank you together with those crazy heads of Vins and Stefano.

I thank Cri and Albi, for sharing completely their life and happiness with me and Elena, Davidone, Paolo and Mace.

From the North to the South. My friends, my hearth: going there every summer was refreshing me, vitalizing me. I have only good memories with you of laughter and sunsets.

You are my second home and I regret that we cannot live closer: Rocco, Nzillaccio, Lele, Marco, Matteo, Cinzia, Bea, Erika Bomber, Marta, Alfredo, Serena e the other Medelli and the Romans Andrea, Francesca and Michele.

Now let's start with the UK.

I will list them according to the entrance in the Annex because honestly, I cannot give an order of importance: Ema&Bob or Bob&Ema lol! Claudio and Ludo.

I could start now listing moments: laughs, songs, food, parties, hangover, messes, fights, confidences, sdrin and sdrin, dinners, nights, coffees, movies, jokes and lunch boxes. But no, no one can have any idea of what all these meant to me for the past 5 years. I couldn't have survived without them. Despite of the regional differences (ahahahah) and the different personalities, Wimpole Road 22a, CO1 2DE has been my English home thanks to them.

Thanks forever, to my English family.

After them I thank Shilan 'DD love' and Leyre. My Whatever group. Whatever because we share everything. Super super friends ahahha I would say. When I had problems, when I was down, so down to not shine anymore, you were there, without judging my useless stubbornness. Telling me always the truth and your opinion, I felt accompanied by you and never felt being judged. Our 'circle of suggestions' saved me tons of time. I knew where to go when I had to cry and looking for a shoulder or when I had to party and laugh. Liiil!

Our coffees, brunches, shopping, gossiping, chatting, scanning, nights out and parties. Everything. My soul mates. Thanks.

After them I can only thank Stefano and Andrea. Stefano because you have been the constant friend in Essex. Always caring. I do miss our morning coffees.

Andrea, no matter the shape of our friendship, I thank you because you care. And I care about you. Especially lastly, you have been one of the few asking me what was going on and helping me on difficult choices. Thanks.

Then I have the other two Italians of the gang: Alberto and Giacomo. Thanks Tumino because your sense of hospitality have been the most pleasant in UK and thanks Giacomo, officemate! Thanks for Gran Canaria and Soton. For the gin and tonic and our eternal conversations on…everything I would say!

Thanks to all the other persons that crossed my life during this PhD, and made it special: Mysha, Pavlina, Mauro, Anna Sergi, Sylvia, Betta, Luca Panofsky, Gessica, Sandro, Marina, Federico, Paula and Daniel.

Thanks to the new, latest arrival: Kiko and Simon (for the best suggestions on jerks ahahaha) and Elisa. It's a pity that we shared such a short time but your welcome every times has been really nice for me. As well as all the times we bumped into each other and the moments we shared.

And now the Mexicans…Luis, I thank you. For the times that you told me honestly ‘Try to be happy, I wish you to be happy because you deserve it’. And thanks, especially, for this summer. Mexico was a hard moment for me but you and your amazing family made me feel completely welcomed.

Thanks Carlos Carlos Carlos. Our friendship started in a funny way and then it evolved. I will never forget your messages while I was in the hospital keeping me awake during the night.

And then the last one, only for time order…Manuel. Our friendship has been sometimes difficult and belligerent because of obvious reasons. But I always knew that in some way it was growing. Until the moment at your wedding when I realized that you were happy because I was there. As a friend. Everything was worthy then. Thank you pinche Manuel ajajajaja!

I then thank all those friends that I met during my previous experiences at Essex but that are still present in my life: Simona, Ignazio, Super Simon, Fra Bro, Vince’ Ciccetto, Patrizia, Fede la Diva, Alba.

And now I thank people from Southampton! An unexpected, unplanned experience that gave me the best six months in UK!

I start thanking you, the girls whom with I shared everything, from the fun to the serious things: Fra, Fede, Dafni and Maria!

And then I thank all the friends that shared amazing moments with me: Kate, Nicholas, Panos, Carlos, Rossella, Angela, Kristine, my flatmate ‘snorky’ Emanuela, Roger (for giving me the best and the most amazing farewell party!), Rob, Matt, Carla, Joe#Viana (and his credit card), Adriana, Paulina, Daria, Nesta. And Neil. Despite of everything, with you I felt that I breathed again and that I could have let it go.

A special thank goes to the Italians of Soton: firstly Giuseppe. I only know that it’s a pity I got to know you too late And then I thank Dario, Luigi, Pasquale, Andrea. And Markis (you are an Italian for me guapo! Ajaja!). I thank all of you for being the last ones arriving at my parties (lol, like super vips) but also the last ones leaving! And no, I won’t have the last tequila with you and the breakfast! Ahahaah!

Then I thank you. Yes, I even thank you. The last but the first one. I won’t mention your name but it is quite clear.

You have been one of the reasons to go back to the UK.

The few nice moments made me fell alive and kept me dreaming.

The difficulties and complications made me stronger. Yes. Because what doesn’t kill you makes you stronger. I thank you despite of everything.

Because of you, I discovered my limits, my fears, my weakness, even my badness. But I also discovered how big is my hearth and how intensive, truthful and unlimited my love can be.

I discovered myself better. And now, that with time, the wounds are slowly turning into scars, I can see myself like a shining, brilliant and stronger woman.

Thanks to all of you. It couldn’t have been the same intense experience without you.

You will be with me forever.

TABLE OF CONTENTS

TABLE OF TABLES AND FIGURES

LIST OF ABBREVIATIONS

SC Social Capital

BHPS British Household Panel Survey

UKHLS Understanding Society – the UK Household Longitudinal Study

SNS Social Networking Sites

FA Factor Analysis

SEM Structural Equation Modelling

CFA Confirmatory Factor Analysis

PCA Principal Component Analysis

EFA Exploratory Factor Analysis

MM Multilevel Model

FE Fixed Effects Model

RE Random Effects Model

ICC Intra Class Correlation Index

VPC Variance Partition Coefficient

RIM Random Intercept Model

RSM Random Slope Model

RCM Random Coefficient Model

CEM Contextual Effects Model

MSOA Middle Super Output Area

HH Household

SAE Small Area Estimates

SAR Sample of Anonymised Records

MAR Missing At Random

MI Modification Indices

INTRODUCTION

In the wake of significant events that defined the last decades, several implications arise on a global level for populations. Indirectly resulting from the worldwide financial crisis, they have been enlarged and have spread to different sectors of social reality.

The first remarkable consequence is the increased poverty rate among all countries in the world. While in the poorest countries it can be easily identified with the deterioration of already precarious conditions of subsistence, in the developing and developed countries we can examine it in different ways: increasing number of persons turning to be poor and the slow disappearance of the middle class. This last phenomenon involves the establishment of two major social classes: one even more rich and the second even more poor, always on the risk to fall definitively in the poverty trap.

The second consequence of this social phenomenon is the diffusion of political ideologies and trends closer to independent, racist and extreme positions. Usually, these political attitudes and beliefs drive common people to develop attitudes of mistrust and suspicion to immigrants and foreign people.
Finally, where poverty is higher, educational attainment levels are lower and minds are closed to diversity, the third indirect big consequence of the last crisis finds fertile ground: not only political extremism but also religious integralisms. This last one especially, like in a not virtuous circle, feeds again all the previous attitudes and fears.

All these consequences have taken a natural erosion of trust within people and between people and institutions, frequently incapable to deal with such changes, involving a natural decrease of social capital. Moreover, the decrease of such a necessary and important capital for development is one of the main current problems to study and deal with, given its centrality like a potential and alternative resource to money for institutions and their policies.

Indeed, the first important implications concern governmental administration at local, national and international levels. In these days, when the resources necessary to support economic, development, health, educational and social policies are scarce, political bodies search for other incentives they can work with and on which they can place their hope.

Conversely, yet at the same time, complementary, other implications emerge from the general population. Different dimensions of the daily life of an individual are continuously changing, both in terms of their meanings and consequences. The first is the structure of society with higher rates of social mobility unfortunately mirroring a higher degree of uncertainty. Other social structures that are changing are the institution of family and the social meanings of job and employment. Deeply linked with these dimensions, social networks are rapidly transforming too. Over recent years, the social sciences have sought ways to address these topics, problems and in general the complex phenomena with the concept of social capital returning to the fore and demanding renewed attention.

Intuitively, social capital is determined by the personal and cultural background, and shaped by the social network that the individual builds during his or her life. Given the presence of a strong personal dimension, it cannot be reduced to the number of ties which form the network. In contrast, many different measures have been proposed which use different individual characteristics to account for its complexity and multidimensionality, as membership to organizations, trust towards people and institutions, political involvement, voting behaviour to make some examples.

Partially abandoned due to its perceived volatility and apparent difficult empirical application (in relation to the other two capitals, physical and human), it remains an interesting concept. But even more so, it can be considered as most probably the most useful concept for contemporary times. Mohan and Mohan (2006, p.1) state indeed the renewal of the 'popularity of the concept reflects a combination of academic and political developments, notably the search for ostensibly "costless" policies of redistribution on the part of centrist governments'.

Following this trend and renewed perception, even the most important governmental bodies have started to refocus on it. In 2001, the Organisation for Economic Co-operation and Development (OECD) Statistics Directorate established, a few years after the World Bank (World Bank, 1998), a project funded by the European Commission for 'Measurement of social capital and question databank' that aimed to: '1) to assess how the notion of "social capital" has been conceptualised in the research literature; 2) to detail how it has been measured in national and international surveys; and 3) to identify priority areas for statistical development. The main outputs of the project are a report, which has been published as an OECD Working Paper, and a question “databank”’ (OECD, 2001).

Following further the international trend and according to a general collaborative idea, in 2001 the British Office for National Statics (ONS) established the ‘Social Capital Project’: recognising its importance for the general wellbeing and for future policy implications, a specific team and concrete studies were structured and implemented as well. The first works were a complex review of the concept and its literature in order to review its definitions and dimensions. A general agreement on this, indeed, would help the following steps: identification of measurements, methodologies, analysis and empirical works (ONS, 2001; Harper and Kelly, 2003)

In 2002, the ONS presented a Social Capital Question Bank. This document presented a matrix that was able to include a wide range of questions – and derived variables – about Social Capital from all the British surveys and was intended to be a reference tool for researchers interested in measuring social capital (Ruston and Akinrodoye, 2002). Characterized by a complexity and richness, social capital is composed by different dimensions that are complementary and necessitate each other but which remain perfectly identified and autonomous. Its placement at midway between society and the people, groups and the individuals, the global and the local, make it the perfect concept to address all the previous issues.

The 'Beyond 2011' ONS Project

In light of these implications and considerations, this PhD project was born four years ago.

Widely analysed and studied, I decided that a further empirical analysis of this topic would be useful not only to confirm and see if it still suits modern times but also to if new measurements of it may assist institutional bodies in their policy choices. As a consequence of these considerations, the research project was admitted to an ESRC/ONS joint funded project, named 'Beyond 2011'.

The ONS project was designed to reply to the challenge to discover statistical methods that would retain the high-quality of population and wider social statistics while falling response rates and rising costs become increasingly common in the UK. This problem is even more urgent for UK, a country based on a long tradition of important surveys and a Census approach. Therefore, following the last Census in 2011, the ONS decided to establish this coordinated programme to explore how requirements for population and other socio-demographic statistics can be met. A key work area is to assess options which may provide alternative or complementary sources for such statistics. Therefore, integrating administrative data, extended survey approaches and alternative methods of Census taking have been considered together with the need to comprehend how these different data sources together can best meet the needs of statistics users.

ONS has further specified the 'Beyond 2011' project in three fields of study supporting it with statistical supervision, access to data and funds: 1) quality measures for population and demographic statistics; 2) small area estimation and 3) statistical disclosure control for derived and administrative data. Given my sociological origins, I opted for a PhD research proposal that met both levels halfway: I chose an important and classical concept that comes from different sociological theories (from Marxist to neo-Weberian theories; Woolcock, 1998) focused on trust, social relations and development of modern industrial society named social capital. On this concept, I then worked empirically with different and challenging methodologies, attempting to estimate it at small area. The innovativeness of this PhD project comes first from the merging of these aspects, particularly from the application of this method to the concept, one of the first to my current knowledge of the current state of the discipline. The second innovative aspect of this proposed PhD work is the complex building up of the modelling work. Indeed, this work is a specific and fluid flow starting from the creation of the dependent variables of interest until the final estimates deriving by the test of different models and methodologies.

Besides, according to the description of small area estimation from 'Beyond 2011' for the small area estimation specification, social capital appears to perfectly fit this kind of methodology:

Model-based small area estimation will form a critical component of the methods used in the Beyond 2011 project to deliver data for small areas at Local Authority level and below. Our key requirement is to be able to estimate and quantify the reliability of estimates for multi-category target variables, e.g. ethnicity, to ensure consistency at different geographical levels, to explore change over time and to do this in an optimal way under resource constraints. We may also need to produce estimates across multiple target variables, e.g. ethnicity and occupation. To date, ONS has deployed model-based methods for small area estimation in univariate cases, including estimation of proportions, totals and the number of cases below key values, and there is a body of ongoing research considering multi-category and multivariate estimation. This PhD will build on previous work to investigate how model-based small area estimation procedures can be used to deliver these objectives, with particular focus on the data requirements, sample designs (including impact and appropriateness of weighting methods), reliability (in terms of spatial and temporal consistency/discriminatory power) and performance measurement of the estimators that are deployed. This work will also need to consider how due to the availability of different sources different approaches may need to be taken in the devolved administrations (ONS, 2011, pg. 2).

Social capital's multidimensionality mirrors the multivariate options for statistical studies and it is related, as I am going to show, to the target variables identified by ONS. It also fits the 'geographical' characterization necessary for small area estimation. Mohan and Mohan (2002) and Van Oorschot et al. (2006) identify the contribution that social capital may make to geography and vice versa and the related problems in deriving a spatially disaggregated measure of social capital (Van Oorschot and Gelissen, 2006 indeed develop the analysis only at country level for European countries).

Therefore, the PhD research project does not only see the contribution that social capital makes to geography but also, according to ONS project's requirements, it will attempt to estimate its different dimensions covering the second field of studies indicated - small area estimates – applied to this multi-category concept, as previously quoted.

The final research project

The overall aim of this thesis is to address these aspects and to develop an in-depth, explorative and descriptive work on social capital. More than trying to establish correlations and causal relations, I indeed assume that beyond the theoretical definitions of social capital, on which there is a general agreement by now, there remain open questions about its measurements and methods to estimate it in a more precise way, considering individual and geographical aspects together. With this work, then, I intend to address these main research questions: Can it still be defined according to the classical dimensions? What additional novel aspects can be considered? What are the characteristics that may influence it? How can we measure it? Can we predict it at small area? Are there are differences between the dimensions of social capital? If yes, how do they work? Are there are also differences in terms of which characteristics influence the different dimensions?

In order to answer to these questions, I firstly present an opening chapter with the review of the main literature on social capital. It includes the classical and older theories and definitions of the concept until the more recent and newer dimensions and types identified and added to the main body of studies.

In the second chapter using individual level data from two important English surveys I develop three social capital factors measuring three main dimensions using Factor Analysis. I hypothesis that they could depend on classical (according to the literature) variables representing membership, citizenship, trust, social networks, caring and neighbourliness but also on more recent and 'new' variables for crime and use of social media. The analysis will be carried out for 2001 and 2011.

Once these three factors have been developed and modelled, according to different hypothesis addressing which other aspects may influence this capital, in the third chapter I built and test Multilevel Models.

Indeed, as described widely in the first two chapters, social capital depends on individual and local or geographical characteristics. While the first are related to the individual person, the second affect his/her networks at micro level. To test these specific hypotheses, Multilevel Models in the third chapter are the best suitable models for testing the effects of individual-level and area-level covariates. These models use the three factors created in the second chapter as dependent variables.

At the first level, to analyse the individual effects, I use survey data to build individual covariates for personal characteristics: gender, age, ethnicity, health, marital status, religion, employment and so on. At the second level of analysis, I test geographical and local differences at Middle Super Output Areas (MSOA) geographical areas using Census information. I also add contextual variables to check the effect of area-level covariates only from the Census and other administrative sources, created with a Principal Component Analysis.

Once these models have been tested, in the fourth chapter I develop synthetic small area estimates of the three factors of SC at MSOA level using a model with individual and area-level covariates, according to the results from the Multilevel Models developed in chapter three previous and main works on small area estimation (Twigg et al. 2000). The final synthetic estimates and corresponding maps presented in this last, fourth chapter will confirm the idea that social capital and its dimensions are related to different and various characteristics both at individual and geographical levels in a complex but rich and complementary way. Because of this, it has proven to remain a possible, useful tool to leverage on for policies aiming to develop society at different levels.

CHAPTER 1 - SOCIAL CAPITAL AND ITS DIMENSIONS: A LITERATURE REVIEW

1.1 Introduction

The analysis and use of the concept of social capital (SC hereafter) within the social sciences has increased considerably in recent years. This has occurred primarily due to the recognition that it has a wide range of applications across numerous fields. This wide use in different sectors and fields of study comes from its main characteristic of being definable as a capital, remembering classical economic definitions. While the firsts types of capital identified have been more related to material assets like money, machineries or natural goods (corresponding to financial, physical and natural capitals) with the development of tertiary sector and post-industrialism jobs, intangible assets and skills have been theorised and defined. The first one was human capital, meant like the individual capital derived by the educational level attainments, talents, skills that can help the economic growth when considered in an aggregate way. After it, SC, intellectual capital and cultural capital have been identified. Focusing on SC, topic of our interest, and tracing a useful list of commonalities and difference between physical, human and social capital, Akçomak (2009) states that SC can be treated like human capital. The ways in which it is created are the basis to treat it as a capital. SC could be formed as a by-product (for a higher status or for the education's network), as an endowment on an inheritance and, finally, as result of a deliberate investments.

These are possible for four characteristics of capital:

1) *as transformative*: it converts an input to an output. Indeed, SC can allow the achievement of a certain outcome with lower cost. A typical example is the fact that high levels of SC and, then, trust, facilitate the transactions between economic actors; the effect can be empowering with the combination of other inputs (e.g. human capital);
2) it can be forgone consumption. In this sense, SC can be labelled as capital with respect to the important element of the *time* that the actors decide to spend and invest in strengthening their relationships, with expectations about future benefits;
3) deeply linked with the point above, SC involves *opportunity costs* relating to the time that could be used in other useful ways;
4) it has *durability* and can decay. Like the others two forms of capital (physical and human), SC is durable and its value depreciates over time, especially if there are no efforts made to maintain it or, as with human capital, it can increase with use.

Another important characteristic of this capital, on which there is an open debate, is the *rival* or *non-rival component*. Coleman (1988) states that some forms of SC can be defined as 'collective goods' because they are not the private property of those who benefit from them. As Adler and Known (2002) add, this is particularly true for internal capital (as we will see in the next sections, is the case with bonding capital): is non-rivalrous because use by a singular individual does not cause the diminishing of availability for other people. Moreover, it can also be traded by individuals in the form of goodwill (Putnam, 1993). But it is also true that, unlike the other public goods to which it has been compared, its use can be excludable (people can be excluded from networks or relations or can easily communicate in their networks) (Grafton *et al*, 2007).

Therefore, the former characteristic makes it a free-riding risky capital and the latter makes it a capital on the boundary of definition between private and public good.

To conclude, the most appropriate term for SC is '*collective good'*. Besides, even if strongly correlated, it differs from human capital. Human capital, indeed, resides solely in an individual and it depends on the stock of skills, knowledge and expertise accumulated.

Consequently, as shown also in different handbooks on the concept (Svendsen and Svendsen, 2009; Castiglione et al., 2008; Li, 2015), SC is a kind of capital that can leverage different phenomena in different sectors.

Theoretical and empirical studies report evidence on its impact in the fields of economics, development, health, growth, educational attainment, employment, poverty, crime, innovation, environmental behaviours, health and services, caring and community studies, ethnicity, migration, neighbourliness and so on. Moreover, it is emerging especially as a process linking actors within networks such as individuals, firms, universities, private and public institutions and government. In addition, there are studies that, in contrast to those listed above, explore not only the mechanism and results of SC in reality by several points of view but also its link with personal characteristics at character level.

The nature of these studies relates to the psychological field and the impact of SC on attitudes and behaviours: such as the openness, propensity to risk, sociality, personal satisfaction about life, relationships and work, acceptance of diversity and so on. These particular fields, in relation to those cited above, reveal the more practical and direct applications of classic SC and earlier definitions, according to the more prominent studies to be described below. With these fields of study and bodies of literature in mind, this study addresses one of the more common problems emerging from the existing literature: not only to establish a more precise definition of SC, but also to identify its multiple dimensionality and to attempt to measure it keeping in consideration its complexity and richness (OECD, 2013; ONS, 2003; The World Bank, 1998).

1.2 Social capital: An introduction

1.2.1 Concept

SC is a recent concept but one that is used increasingly to the extent that it has become diffused across different fields. This is due to the fact that the concept is multi-dimensional and therefore has wide applications; and it crucial that the very different realities it refers to in different fields are explained (Fine, 2010). In an intellectual environment in which big models or wide concepts frequently become problematic when we attempt to adapt them to a world that is simultaneously globalized yet even more local, where the strength of a system can be located in big systems or local communities and networks, the human dimension gains fundamental importance in all its features. One of these dimensions is surely the social environment created and shaped by individuals themselves. Following the assertion of the importance of human capital, over recent years SC is also becoming the central focus of studies and, more frequently, in the policies. Subsequently the character of a community is just as important as individual agents.

The term was coined by Jacobs (1961) and Loury (1977); however, typically, the authors considered the 'fathers' of the concept are Putnam *et al*. (1993) Bourdieu (1986) and Coleman (1988; 1990). Through their seminal works they promoted the concept bringing it to the forefront of sociological thought. Subsequent to these seminal works, the literature surrounding it has grown in an exponential way in recent years. Akçomak (2009) suggested that in the Social Science Citation Index, the number of articles with 'social capital' in the title increased from less than 500 citations and papers to 4000 papers in 2006 and more than 4500 citations in 2007. Similarly, the number of articles that discuss the topic more generally have increased in both categories in a similar way. These numbers do not take into consideration those articles that display a similar trend (e.g. all the papers on social capital and human capital). Despite this, there is no concrete agreement on the definitions and characterizations with even studies on the contradictions of capital and its weakness as a concept (Portes, 1995, 1998; Murray, 2005).

1.2.2 Origins, definitions and main characteristics of the social capital concept

SC is '*the sum of the actual and potential resources embedded within, available through, and derived from the network of relations possessed by an individual or* *social*' (Nahapiet and Goshal, 1998, p.243). It resides in the social structures, connections and relationships embedded at different levels of the society: families, communities and social organizations (Coleman, 1988) and the concept was originally developed in sociology in order to explain the dynamics between individuals within the communities. Enlarging its application, it has been subsequently used to explain civic engagement and associational activities in societies that lead to social and economic wellbeing. This can be identified in several ways depending on the SC accumulated: educational attainment, community development, crime reduction, economic development, democracy, governance, employment, health and caring, children's welfare and knowledge exchange (Zheng 2010). Lin (2001) divides these possible outcomes into two types: *instrumental*, such as wealth, power and reputations and *expressive*, such as health and life satisfaction.

Commencing with the early works based on a range of economic topics[1], four sources of SC were identified:

1) an individual's *social relations*, fundamental for status attainment;
2) *identification* with a group or a voluntary organization, because of the positive sense of belonging;
3) *solidarity*, that can enable an individual to consider community well-being as being as important as her own well-being;

---

[1] The works regarded a) the influences of interpersonal ties for better opportunities in the labour market and gain higher status and income; b) the role of trustworthiness between members for the ROSCAS, the rotating savings and credit associations system; c) the effects of social relationships and social support on health and well-being, both at the individual level and the community level; d) immigration and immigrant entrepreneurs. Generally speaking, SC has a well-established relationship with the outcomes pursuit by the policy makers: economic growth, social inclusion, improved health and more effective government.

4) *enforceable trust*, arising from information exchanges, social norms and monitoring capacity in social networks.

The first definition was deeply linked to the concept of human capital, which remained the main topic related to SC:

> An individual's social origin has an obvious and important effect on the amount of resources that is ultimately invested in his or her development. It may thus be useful to employ a concept of social capital to represent the consequences of social position in facilitating acquisition of the standard human capital characteristics [...] social capital refers to naturally occurring social relationships among persons which promote or assist the acquisition of skills and traits valued at the market place [...] it is an asset which may be as significant as financial bequest in accounting for the maintenance of inequality in our society...(Loury, 1977, p. 176).

Classically, the main definitions come from the 'fathers' of the concept: Bourdieu, Coleman and Putnam.

More describable as a micro approach, Bourdieu defines it as:

> The aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalised relationships of mutual acquaintance and recognition (Bourdieu, 1986, p. 210).

Bourdieu's aim was the study of the different forms of capital – cultural (Bourdieu, 1979), social, political, symbolic and economic – and to attempt an understanding of how they transform one another. His approach focused on the singular person with SC representing an individual attribute in personal networks.

Lin (2001) makes clear the concept stating that SC is the 'resources embedded in one's network or associations' (Lin, 2001, p. 29). Putnam (1995, 2001) broadens the definition by including *informal social networks* in the SC concept. After his famous case study on Italy, he specified ten years later that we are inclined to think that the higher form of social involvement includes citizenship, membership and public life. But crucial support in everyday-life is situated in the friendships and in other types of informal relationships such as those between neighbours. These can be included in the informal social networks concept.

Akçomak (2009) and ONS (2003) underline three components that are necessary conditions for the existence of SC: *social structure*, *resources* and *actions* and with Akçomak identifying the commonalities in the definitions of SC by adopting a micro approach (Akçomak, 2009, pp. 6-7):

1) SC arises from social networks;
2) the social networks themselves are necessary conditions but not sufficient for the creations of SC: they are to be utilized to produce SC;
3) individuals can invest in social relations with an expected return;
4) SC may have negative effects as well as positive outcomes, e.g. enabling enhanced information exchanges but could affect individual decision making, working as a form of social control, encouraging mutual assistance but also causing restrictions on access to the networks, closureness.

Coleman's approach (1988; 1990) is opposed to that of Bourdieu and definable as a macro approach to SC. His concept refers to a dimension of social groups, organisations and societies. He defines SC in respect to its functions:

> Social capital is defined by its function. It is not a single entity, but a variety of different entities having two characteristics in common: they all consist of some aspect of social structure and they facilitate certain actions of actors – whether persons or corporate actors – within the structure... social organization constitutes social capital facilitating the

> achievement of goals that could not be achieved in its absence or could be achieved only at a higher cost (Coleman, 1990, pp. 302-304).

Jacobs' (1961) description provides an interesting perspective as she identifies the role of neighbourhood networks in enabling self-governance:

> […] Networks are a city's irreplaceable social capital. Whenever the capital is lost [...] the income from it disappears never to return, until and unless new capital is [accumulated] (Jacobs, 1961, p. 138).

The author continues by mentioning the importance of acquaintances, the knowledge level of neighbours, the public respect and trust that arise from social relations within communities but which have influences at macro social levels.

Finally, other important specifications of the concept at macro level come with Putnam's (1993) famous work on Italian civic culture. Thanks to him, we have an introduction to the concept of trust and civic participation and the entry of SC to development and political themes. His definition posits SC as:

> features of social organisation, such as trust, norms and networks that can improve the efficiency of society by facilitating coordinated actions (Putnam *et al.*, 1993, p. 167).

Thus, following these early definitions, it is possible to add some characteristics to them regarding the macro dimensions (Akçomak, 2009, p. 9):

1) norms, values and solidarity are sources of social capital;
2) trust is an important source of social capital, both in interpersonal relationships (personalized trust that can be also generated from repeated

actions) and in relationships between actors of a community (generalized trust that can be reinforced by the system of sanctions);

3) whatever the source of SC, it is based on social networks and/or associations.

There is, nonetheless, a subtle distinction regarding the workings of norms and values. Authors who orientate towards the micro approach (Dasgupta, 2005; Knack and Keefer, 1997), lay stress on individual actions that shift resources within social networks, whereas scholars who view SC as a more communal asset highlight the role of both community and social structure in facilitating certain individual behaviour in terms of individual and/or communal well-being (Coleman, 1990; Knack and Keefer, 1997).

Before describing further, the types and classifications of SC that have been identified and proposed during the years, it is necessary to mention some important issues concerning SC: the 'adverse SC'.

SC can have, indeed, possible negative effects. Murray (2005) identifies four negative aspects of SC: the exclusion of outsiders from networks; the excessive claims on individual members; the restrictions on the individual freedom of networks' members and downward levelling norms. Nahapiet and Goshal (1998) underline that SC, despite its apparent advantages, can facilitate a certain closure of a community and restrict its sensitivity to new information and alternative and creative way of doing things, showing drawbacks. One other main problem is the fact that frequently trust, solidarity, mutual assistance can be strong within communities but not between different communities. Mohan and Mohan (2002) remember Rubio's definition of 'perverse SC', typical of criminal organizations like Mafia and gangs. They also list all the criticism to Putnam's definition of SC: real mechanism of production of SC through membership not clearly described, its choices about how to measure participations that do not consider organizations' structures and openness and the motivations of individuals, the society-centred vision of SC (while frequently States can create or destroy SC).

The wide range of these definitions and descriptions reveals the difficulties in establishing a single and universal definition of the concept. But they also fully represent the natural complexity and multidimensionality of this concept, making probably useless even the search of a unique definition. Indeed, when governmental bodies (at national and international level) set the different projects to study SC as described in the Introduction, the first step of analysis has been always a revision of the literature that, according to all of their reports, has to be related to the definitions just presented inasmuch 'classically stated and fundamental' and complementary among themselves in defining the concept: personal relationship, social network support, civic engagement and membership, trust and cooperative norms (World Bank, 1998; ONS, 2001; OECD, 2001; Ruston and Akinrodoye, 2002; Australian Bureau of Statistics, 2002; Haper and Kelly 2003; Scrivens and Smith, 2013).

1.2.3 Components and types of social capital

Having arrived at preliminary definitions it is possible to begin to outline the main aspects and components of SC.

First, we can identify possible three different levels applicable to the study of SC:

a) *individual* (Burt, 1992, 2000)

b) *organisational* (Nahapiet and Goshal, 1998)

c) *societal* (Putnam, 1993)

The common thread that runs throughout these different approaches is that the net of relationships within which a person is embedded is the fundamental resource that suggests better economic performance. At an individual (or micro) level, SC is the embedded resource in one's relationship with others. Burt (1992, 2000) underlines the fact that these networks, and their formal and informal ties, are the factors underpinning good economic performance. If these networks are developed between employees and members of an organisation they help the attainment of collective goals resulting in a new source of organisational wealth (Nahapiet and Goshal, 1998).

There is also a level regarding the macro dimension of research. These studies try to catch the impact of SC in regional or societal reality (Putnam, 1993). On this level, SC comprises a range of features. The first characterization is between the *structural* and the *cultural aspects* (also called *cognitive aspects*) (Chou, 2003). The former is related to *connections* and *social networks* (i.e. *formal* such as associations and *informal*, as in families) and can be externally observed aspects that allow the identification of how the individual gains resources that can be developed, while the latter refer to *norms*, *values*, *trust* (with further characterisations), *attitudes* and beliefs. They are more intangible and assist in the sharing of information, the collective action and decision-making process thanks to established roles, defined rules, procedures and precedents (Grootaert and van Bastelaer, 2001). As Modena (2009) describes, it is possible to open two boxes of structural and cultural aspects with different elements inside. Among the structural aspects of SC (social networks) it is further possible to distinguish between *horizontal* and *vertical relations*.

Following Coleman (1988) and Putnam (1993), horizontal relations can be further classified in *bonding* and *bridging* relations (or *bonding* and *bridging social capital*). The bonding relations are referring to the strong links between people on an everyday basis (e.g. the family) while the bridging form is based on weak connections between heterogeneous people. Granovetter (1973) pointed out that this last kind of capital allows the emergence of mobility opportunities.

The vertical relations can be thought of as *linking relations* and refer to the connections between people across different social strata. More precisely, linking social capital allows the leverage of resources, information and ideas from formal institutions beyond the community (Woolcock, 2001). Therefore, three further dimensions of SC have been identified: *bonding*, *bridging* and *linking* SC. In addition, cultural aspects include *trust* and *norms and values*. Trust, as a classical and principal element of SC, helps facilitate social exchanges and reduces the need for time consuming and costly monitoring. It is a multidimensional construct with several forms that can be described following different categorizations. Rousseau *et al*. (1998) identify three forms of trust:

- *deterrence-based*: based on utilitarian considerations and on efficient penalty mechanisms. The subjects cooperate only because the breaking of a contract or agreement is too expensive.
- *calculus-based*: based on rational choices and objective information regarding the reliability and competence of partners.
- *relational-based*: trust is the product of continuous relations in which the emotional element enters as a form of affection and intrapersonal trust.

Paldam (2000) argues that *generalized trust* is directed towards unknown people while *special trust* refers to friends and institutions. Uslaner (2002) distinguishes between *strategic trust* (also called *knowledge-based trust*) and *moralistic trust*. The former regards the trust existing in specific relationships between specific persons for a particular context (and which reflects their expectations for future behaviour of the actors engaged in this relationship). Moralistic trust is directed towards unknown people and is shaped by the expectations about how people will behave. Sabatini (2008) proposes a further classification: *horizontal trust* (in friends, relatives, etc.), *intermediate trust* (in local institutions) and *vertical trust* (in government institutions).

The second components of cultural SC are norms and values, such as obligations, tolerance, solidarity and democratic orientations (Castiglione *et al.*, 2008). Akçomak and ter Weel (2012) and Dakhli and De Clercq (2004) study two dimensions of trust: *generalized* and *institutional*. The former is directed towards other people, the latter regards the level of trust towards institutions (public or private). The second is complex. Indeed, it depends on the perception of the efficiency of institutions. If people believe or feel that institutions can be good in mediating the exchange, protecting the individuals against every breach of trust, people will be more willing to interact, assume risks, share knowledge and information. However considerable debate continues around whether it is possible to consider trust as an outcome of SC or if it is a component of the shared values which constitute social capital or both (Woolcock, 2001). Other components of SC are as follows.

Knack and Keefer (1997) add the dimensions of *being an active member of associations* and *organisations voluntary-type*. This sense of belonging increases the contacts between members and communities, mutual support and solidarity and, if there are more associations in a region, the attractiveness of resources. Putnam (1993), pointing to the differences in economic performance between North and South Italy, states that these are due to the richer associational life of northern regions. With regard to the type of organizations, we can identify two types of organizations: the *Putnam association* and the *Olson type* (Modena, 2009). The first type identifies those organizations that we can identify as being more informal and unstructured, with a real specific target goal shared by the members: sport, education, music, arts, church, charity, voluntary, ecology, human rights and peace associations. The Olson type associations, on the other hand, are more structured and formal and are linked to a more political aspect: professional associations, trade unions, political parties. The last aspect regards *the respect of norms of civic behaviour* (Knack and Keefer, 1997). This refers to informal mechanisms that limit predatory behaviour and help promote overall well-being.

Deeply linked with the issue of the sharing of knowledges, SC can be composed by *strong ties* and *weak ties*. Granovetter (1973), in attempting to link micro-level interactions and macro-level patterns with an analysis of social networks, points out that the relationships between people can exhibit frequent contacts and emotional involvement (as with close friends or with relatives and parents) or sporadic interactions with a low emotional commitment. The first kind of relationships identify *strong ties* in social networks while the second kind, *weak ties*. The latter from is especially useful and powerful for the sharing of knowledge and is an important resource for individuals. These are ties of acquaintances such as neighbours or work colleagues and can be compared to Putnam's *bridging social capital* model, while those that suggest strong ties are related to *bonding social capital*.

Another important aspect is the already proven correlation between SC, education and human capital even if it is important is to distinguish the two kinds of capitals. Human capital regards how the individual gains the resources that can be developed whereas SC includes a more sociological approach to human action in which the individual is likened to an actor shaped by social factors. Ackomak and ter Weel (2012) state that investments in human capital and education have an indirect effect on levels of trust in societies. Goldin and Katz (1999) demonstrate that high levels of educational attainment induce the dense networks in which SC is created. It generates positive externalities that increase the knowledge of other people and decrease those that behaviours that are opportunistic. Furthermore, they reduce the free-rider problems that occur in uncertain and low availability information contests.

Finally, high levels of educational attainment appear to generate a general openness of mind, an aptitude in relationships, and the creation of those networks in which SC forms (Dakhli and De Clercq 2004; Akcomac and ter Weel, 2012; Kaasa et al., 2007).

All the concepts, dimensions and theories presented up to here have been helpful in identifying how SC is built in social relations and how it can be used and spend. The specificity of this capital, indeed, is that its structural aspects like nodes of networks, types of nodes (familial, friendly, acquaintances, institutional), types and intensity of the relations with these nodes are at the same time the same channels through which and with which SC is used, spent and invested. While SC is built and nourished, it is also used and reinvested. Woolcock (1998) perfectly describe this peculiarity:

> This leaves unresolved whether social capital is the infrastructure or the content of social relations, the ``medium,'' as it were, or the ``message.'' Is it both? ``Defining social capital functionally,'' Edwards and Foley correctly point out, ``makes it impossible to separate what it is from what it does.'' (Woolcock, 1998, p.156).

This particular aspect helps to explain why theoretical dimensions identified can be considered at the same time the main measurements of this concept, as I am going to describe in the following paragraph.

1.2.4 The measurement issues

The complexity and richness of conceptualization has clearly taken to a considerable variety in the proposal and use of measures, factors and indicators. Therefore, researchers rely on broad sets of indicators to measure SC. As van Deth (in Castiglione *et al*. 2008) states, although the exact status of these indicators as conceptualizations of SC is not always unique, the common requirement is that they should account for anything that facilitates cooperation between individuals. The author also points out the four data collection methods: surveys and polling, statistical indicators and official statistics, community studies and observations and projects and experiments. He also categorizes each measure of SC based on a further two dimensions other than the data collection method: the level of analysis (individual or collective property) and its characterization (structural or cultural), as described above. His useful summary of the major measures of SC in relation to all dimensions is presented in Table 1.1 below.

As we can imagine, the first dimension about the level of the analysis is the more difficult to identify because of the natural overlapping between individual aspects and collective aspects of this type of capital. I described in the first paragraphs how even the more classical authors like Bourdieu, Coleman and Putnam identify micro, macro and intermediate levels of SC ((see par. 1.2.2). The table catches this first important dimension, focusing on aggregate and sometimes more formalized measures of SC.

The second dimension used, the characterization, fully summarized all the main components and types of SC described so far: from membership to trust to citizenship just to cite the most important and used. Chou (2003) (par. 1.2.3) identifies the two characteristics: structural versus cultural SC. This helpful definition formalizes and includes all the other definitions about formal vs. informal connections and networks, strong ties vs. weak ties, bonding vs. bridging/linking SC, generalized trust vs. institutional trust, horizontal trust vs. vertical trust, Olson organisations vs. Putnam organizations that I described above.

From the combination of the two levels to which measurements and variables correspond, we can identify main methodologies and tools to empirically study SC. Surveys and the pooling dominate this field. Asking people directly about their networks, their values, attitudes, perceptions seem to be the best way (Castiglione *et al.*, 2008). Themes such as social cohesion, engagement in networks, civic orientations, obligations, norms and reciprocity have been investigated in social sciences before the conceptualization of social capital. A large part of these studies was based on the World Values Surveys and the European Social Surveys (ESS), which provide data at the minimum regional level. Van Oorschot et al. (2006) using classical definitions and ESS data, are able to make a national comparison between European countries and their macro-regions (North, South, Western and Eastern Europe). They find that, apart for Northern Europe that shows higher levels, SC is strongly accumulated together with physical and human capital and that it is strongly related to religious beliefs and gendered and to a political left-right stance.

Then international organizations like World Bank and OECD have developed proper project of studying and monitoring SC for several purposes, mainly to enhance development policies (World Bank, 1998; OECD, 2001). But even more, the single states have begun to systematize national survey on this topic, frequently on the base of these two international surveys. International organizations like the World Bank and OECD have developed projects for studying and monitoring SC for a range of purposes, but primarily to enhance development policies. But even more, the nation states have begun to systematize national surveys on this topic, frequently on the basis of these two international surveys.

Generally speaking, mixed-method approaches using both qualitative (in-depth interviews and focus groups for example) and quantitative tools (survey and building of composed indices) are most widely used to compensate the limitations of surveys and polling approaches. It is also true that different measures can be detected more efficiently from different techniques. For example, polling methods are useful for the analysis of trust while official statistics are efficient for measuring the density of voluntary associations (based on objective information) (Castiglione *et al.*, 2008).

In the British case, the most important survey on aspects of SC is the British Household Panel Survey (relaunched as *Understanding Society* for the last two waves). Previously the British Social Attitudes Survey and Citizenship Survey were also conducted. We can also find surveys that contained additional core modules on SC for some years, such as the General Household Survey.

Table 1.1: Major measure of social capital (inverse measures in Italics)

Characteristics	Data collection	***Structural aspects***	***Cultural aspects***	
		Networks/contacts	*Trust/confidence*	*Civic norms and values*
Individual feature	*Surveys/polling*	Membership in voluntary associations Volunteerism (Ego-centred) networks and social contacts Time budgets Number of children in the household	Trust in other people Confidence in institutions Ethic and corruption	Norms of reciprocity Obligations Democratic attitudes Solidarity and identification Togetherness Subjective well-being
	Statistical indicators/ official statistics	-	-	-
	Community studies/ observations	Voluntary associations	-	-
	Projects/ experiments	-	Distribution of money	-
Collective feature	*Surveys/polling*	Aggregate membership figures Aggregate voluntarism figure Aggregate social contacts Network characteristics (density, etc.) Aggregate time budget figures Social mobility	Aggregate figures on trust in other people Aggregate figures on confidence in institutions	Aggregate figures norms of reciprocity Aggregate figures on democratic attitudes Aggregate figures on solidarity and identification
	Statistical indicators/ official statistics	Organizational activity and resources Volunteerism Mass media and use of (new) technology	Balance sheet of co-ops	*Voting turnout* *Crime rates* *Legal protection - number of lawyers* Bloodlettings Number of co-ops
	Community studies/ observations	Voluntary associations		-
	Projects/experiments		Lost wallet with money	-

Source: Elaboration on Van Deth (in Castiglione et al., 2008, p. 160)

Another important issue is concerned with the degree to which various measures indicate the existence of a single latent construct. Many studies report an attempt to construct a factor measuring SC[2]. Onyx and Bullen (2000) analyse sixty-eight items to represent all aspects of SC in order to construct a General Social Capital Factor and individuate three stronger factors than the other dimensions: local participation, social involvement and trust. Other researchers assert that the creation of overall measures makes no statistical sense but that it is more interesting to create composite measures (Hauser et al., 2007; Laursen and Masciarelli, 2008) or keep the analysis at singular indicators and variables obtained by the surveys or the original sources (Criscuolo et al, 2010). Van Deth (in Castiglione *et al*., 2008) adds a further the issue regarding the identification of meaningful sub-dimensions of measures and indicators that sometimes make the indicators appear less unrelated.

A more recent approach that is increasingly used is the incorporation of factorial analysis or principal component analysis to reduce the number of variables to be used without losing the richness deriving from the existence of many variables (Kaasa and Vadi, 2010). With this technique, it is possible to obtain 'macrovariables' incorporating those bearing similar patterns and common factors. These macrovariables can be used subsequently for models and correlations.

Other important measures used are the lack of cooperation and low levels of economic growth (interpreted as an indirect measure of a lack of SC). Modena (2009)

[2] Frequently, by an econometric point of view, happened with the use of factorial analysis or principal component analysis, like in the cited worked (see also next chapter about the empirical applications).

identifies singular measures and indicators relating to the dimensions of *formal/informal networks* for the structural aspects wherein family, neighbours and friends' networks are considered for informal networks while for the second, besides the number of associations, the author stresses two particular kind of associations already defined: the more 'informal' *Putnam associations* (sport, education, music, arts, church, charity, voluntary, ecology, human rights and peace associations) and the more 'structured' *Olson associations* (professional associations, trade unions, political parties) (Olson, 1982). For the *trust*, *norms* and *values* (cultural aspect), many indicators for the *trust* and *trustworthiness* dimensions are indicated more than those previously indicated: whether people can be trusted, corruption, number of legal proceedings for work disputes, number of protests for bank bills and cheques, number of people reported to judicial authorities by police. For the *civic and political society* measure of government inefficiency, human liberty, political stability, political rights and civil liberties are listed. Finally, for the *social integration*, social mobility, suicide rates, divorce rates and youth unemployment rates are indicated.

1.3 Social capital and its new dimensions: Evidence to date

After the presentation of the main, classical concepts and definitions of SC, it is worthwhile touching upon other recent developments in terms of the applications of this concept. Given its stated multidimensionality, the concept has become entangled in other, more recent, fields. Indeed, some of the characterizations presented in the past, have been increasing used only recently, especially in fields like health, behavioural, crime, education, organizational and psychological studies.

The first, and perhaps the central and more interesting application, derives from the famous Hofstede study on cultural dimensions (Hofstede 1980). In this seminal work, the author contemplates the ways of thinking and the social actions of people, comparing forty different countries. He conducts this comparison using like a key variable – the culture itself – stating that all previous works failed to include this factor. Hofstede uses survey data derived from questionnaires submitted to workers in multinational business organizations and in analysing them identifies four main dimensions along which all the cultures can be described and summarized. These dimensions are: uncertainty avoidance, individualism, power distance and masculinity-femininity. These dimensions are transmitted through education from a young age and they even affect the way individuals work.

The first important study to use these dimensions was Shane (1992) in which the author demonstrates that individualistic and non-hierarchical societies are more inventive than others. Shane composes an index of the values of individualism and power distance (that can identify the social hierarchy of a society) including several variables created from the Hofstede study. The index of individualism includes: the belief in the importance of freedom, the cosmopolitan orientation of a society, the preference for small organizations in respect to large organizations, the importance to people of being compensated, the importance of personal contacts who can assist in achieving one’s goals, psychological characteristics such as autonomy, independence, dominance, and non-conformity. The power distance index includes: the desirability of hierarchy, the lack of informal communication between people of different hierarchical levels in organizations, the centralization of power, the belief in the importance of detailed instructions, fatalism and weak work ethics, the control over subordination, and an unwillingness to accept change in the distribution of power. All these studies lead to the application of the SC concept and risk-aversion behaviours, also at societal level (Shane, 1995). Other more recent fields of applications of the SC concept are social capital and the use of social media, SC and crime, SC and migration (and all the related issues linked to the topic of migration and ethnicity), SC and financial crisis, SC and health and so on. Indeed, they seem to become more and more relevant. The ONS Social Capital Project, started on 2001, in one of its paper about literature review on SC highlights the arising importance of the following fields for recent policy implications: crime and economy, health, education and civic participation. New measurements or updated definitions of original and classical variables are being used for these current purposes (ONS, 2001). I choose to test some of these dimensions and include them in the first model to be used in what follows: a Factor Analysis. I will therefore focus and describe these in the following chapter.

1.4 Conclusions

This first chapter offers a critical summarization of the broad literature on SC and its definitions and dimensions, especially those regarded as classical. I will therefore base the subsequent building of SC factors on most of them and they will become the dependent variables for the following estimation work. I have presented a detailed account of the aspects necessary for my work but, at the same time, forwarded the general idea about a concept that is at once complex, rich and multidimensional. Due to this this complexity, the following work of estimation will be increasingly composite and will be innovative in terms of several aspects. First, the novel use of established datasets (i.e. the British Household Panel Survey and Understanding Society Survey). Some studies of SC have been carried out using these surveys (Li *et al*., 2005) but this is the first study that uses a Principal Component Analysis to obtain proper and complex SC's factors (see Chapter 2). The innovative contribution of the present study will not only be the use of these datasets but also the techniques used: multilevel modelling (Chapter 3) and the small area estimates approach (Chapter 4). These methods will be tested to see how individual and area characteristics are related to SC and to finally estimate average levels of SC factors for all the small areas for England and Wales. Lastly, this work will be part of a larger project aimed at discovering alternative means of achieving estimations to the Census through the use of other kinds of data, especially administrative data sources. This goal makes this work innovative in itself.

Chapter 2 – State of social capital in England and Wales at census years: A Confirmatory Factor Analysis with survey data for 2001 and 2011

2.1 Introduction

Social capital (SC hereafter) is a concept used increasingly to explain several processes across diverse fields. Related mainly to social relationships and networks between people (Coleman, 1988) (Nahapiet and Goshal, 1998), this kind of capital has become, together with physical and human capitals, an important asset (Fine, 2010). Moreover, in comparison to the other forms of capital, especially human, it is notable for its multidimensionality and intangibility. Indeed, SC implies numerous concepts and has been measured with many different variables. If one searches for academic papers on SC on IDEASRepec.org, one of the largest bibliographic databases dedicated to economics, some 280,000 articles are returned. Yet despite the increasing number of articles and studies on SC, the effective mechanisms that define its dimensions, and even more, the ways these dimensions influence each other and act together, seem to remain under intense debate. Certainly, we can recognize that there is general agreement on its components and dimensions: from the most classical and theoretical as identified in the seminal studies of Bourdieu (1979, 1986), Putnam (1995, 1993, 2001) and Coleman (1988, 1990) to those more recent and empirical such as Knack and Keefer (1997), Akçomak and ter Weel (2005, 2009). We can then surmise from these cited works that aspects like memberships, citizenships, attitudes towards institutions, trust (both generalized and institutional), voting, political attitudes and behaviours have been clearly established. On the other hand, others have been added more recently through several, important works: SC and media use and effects (or social media), local effects of SC (especially neighbourhood effects), social control and anti-crime effect, individual characteristics and their links with SC – ethnicity, religion, personal satisfactions and cognitive skills; healthcare, family networks, migration, employment opportunities, cultural improvements, schooling and educational attainments, environment and sustainability attitudes and so on. As we can see, it seems that the concept is widely linked with all aspects relative to the human being.

In this study, the dimensions taken into consideration are the classical ones: memberships, voting behaviours, attitudes and trust towards government, social and informal networks (family, friends, neighbourhoods and acquaintances). As we will see, they can be confirmed as the most important components in the creation of factors of SC, both for 2001 and for 2011 – the specific years that are the subject matter of this thesis. Something more interesting and particular occurs with the data available for 2011. In addition to these main components, new significant dimensions result: use of social media and fear of crime and crime control effect.

In this analysis, I hypothesize two different models that reflect theoretical assumptions appropriately, one for 2001 and another one for 2011. The choice of these two years is not random as they represent the two last census years in the UK but due to differences in sample and availability of variables it is not possible to make a comparison between the two models. For 2011 only, therefore, as explained in the introduction to this thesis, the factors identified in this work and confirmed by the

Factor Analysis will be used in the following multilevel modelling work, where census covariates will be required.

In section 2.2, I will describe the methodology applied in this study, the Confirmatory Factor Analysis and I will introduce the two main surveys used. I will also add a last part on specific technical issues related to the analysis. Special modules with the availability of many different variables potentially connected to SC are present in the surveys for that will be used for these two years: the British Household Panel Survey (BHPS hereafter), wave 11 – covering the year 2001, and the UK Household Longitudinal Study, later became Understanding Society (UKHLS hereafter), for which I intend to use wave 3 – covering the year 2011. In order to retain the richness of the information available in the surveys used, I carry out a Confirmatory Factor Analysis (CFA hereafter). As I am going to describe thoroughly in the next paragraphs, this method allows for the capture of the common patterns of variance and covariance between variables. These common patterns are hypothesized to underlie to the same construct (*factor*) that can be interpreted as a measure, like a 'macro-dimension' that includes them. It mainly transforms observed correlated variables into factors that are a linear combination accounting for a pre-specified proportion of the total variance.

In this way, the problem of the selection of variables to be used is avoided and the factors identified have stronger backgrounds both at the theoretical and empirical estimates levels. As I will demonstrate later, despite of a notable literature already existing around SC and CFA, this approach seems one of the more complexes, rich and innovative for the models presented; the number of variables used the use of one of a significant British survey and, finally, for the strong established theoretical background.

Later, in section 2.3 I will identify dimensions and corresponding variables and measurements that I am going to use in the models. They will be presented referring to a body of literature and studies that are mainly empirical and that found significant correlations between similar variables and SC. Some of these studies, besides, even used similar methodologies of factor analysis. They will be also introduced under the definition of common dimensions or sub-dimensions that helped the identification of singular factors.

As Di Stefano and Hesse (2005) highlight, many articles using CFA do not provide full explanations at the theoretical level concerning the development of the model. Given that CFA is mainly characterized by the assumption of an *a priori* model in contrast to other exploratory methods, it is necessary and indeed interesting to describe how the models were built, how the variables were chosen, and so on; especially in light of the wide SC literature available. All the variables included have been recoded or have generated new variables that, when ordinal, show a parallel structure of increasing SC at the increase of the items.

In section 2.4 and section 2.5, I will describe the empirical process for model 2001 and 2011. I will specify the process of recoding of the variables, the analysis strategy that ended with the proposed models (and their diagrams) and results. I will also report and explain main corresponding statistics: descriptive statistics, standardized results with correlations and cross-loadings, main goodness of fit's indices and any other particularities. I will try to offer theoretical explanations of these dimensions, especially for those that are more recent. I will indeed focus also on the aspects regarding cross-loadings of latent factors on variables and on how several variables and related factors from the models were excluded with regard to the original hypothesis. These empirical specifications about correlations of underlying dimensions cannot be identified always from the theory and they stand out from the empirical work on variables and models (according, for example, to modification indices).

Some conclusions will be drawn in the last paragraph and, last, complementary material useful to the analysis is available in the Appendix and the Supplementary Material sections.

2.2 Methodology and data used: Why a Confirmatory Factor Analysis on survey data

The complexity and multidimensional aspect of the concept of SC, as described above, is not only due to the vagueness of such dimensions but also to their deep inter-correlation and their observable/unobservable components. The variables of a study or a survey can be defined as attempts to make *directly* observable and measurable such behaviours, attitudes, beliefs and relationships that are latent and *indirectly* measurable. We can also state with a quite high degree of certainty that there are strong and common patterns between all these dimensions. These ‘macro patterns’ are not directly observed by these variables in themselves.

Over the previous twenty years, especially in research methodology and in the behavioural and social sciences, several empirical techniques have been created and employed in order to study these kinds of cases. The main ones can be brought back to Factor Analysis (FA) and Structural Equation Modelling (SEM) frameworks. They mainly aim to identify constructs, called *latent factors* (or also *endogenous variables*), that can be used as indicators of the underlying reliability of *observed variables* (called also *exogenous variables*), mainly catching common patterns of variance between these variables. These two large groups of statistical techniques are relatively similar in approach and conception. The main difference is that while FA primarily aims to investigate the structure and the patterns using two main approaches (Confirmatory Factor Analysis – CFA - and Exploratory Factor Analysis –EFA), the SEM has as its target the testing and building of appropriate models through two different steps: the measurement model (via CFA) and a structural model (Schreiber *et al*., 2006). SEM models are also more appropriate in the case of complex models or where different levels are required.

The pertinence of this methodology respect to the theories, concepts and measurement widely described in the previous chapter is well explained by Muthén & Muthén (1998):

> Confirmatory Factor Analysis is appropriate in situations where the dimensionality of a set of variables for a given population is already known because of previous research. The task is not to determine the dimensionality of a set of variables or to find the pattern of the factor loadings. Instead, CFA may be used to investigate whether the established dimensionality and factor-loading pattern fits a new sample from the same population. This is the confirmatory aspect of the analysis. CFA may also be used to investigate whether the established dimensionality and factor-loading pattern fits a sample from a new population.

I chose to carry out a Confirmatory Factor Analysis because this method, more than the EFA, is a theory driven exploratory analysis; indeed, the analysis implies an *a priori* hypothesized model to estimate a population covariance matrix that is compared with the observed covariance matrix. The main target is to minimize the difference between the estimated and the observed matrices. The estimation work is carried out in respect to unobserved variables, called *latent variables* or *latent factors*, and their patterns of variance explained as observed variables are called *factor indicators*. An important part of the CFA is to test the reliability of the observed variables in relation to the latent variables on which they are expected to load. Other steps of the analysis are also the testing of hypothesized interrelationships or covariation/correlation among the latent factors, the variables and the factors and variables.

One of the main advantages in carrying out a CFA instead of an EFA is that indices of goodness of fit are available. Despite the thresholds of these indices not being defined in a unique way, they help the assessment of the model more than an EFA is able. The following use of the factors identified for 2011 in Multilevel Models required a further, safe evaluation of their goodness (Geiser, 2013). There are still several empirical studies on SC using CFA. Ellison *et al*. (2014) apply this method to study the formation of *bridging SC* with respect to several variables relating to the use of Facebook. Narian and Cassidy (2001) use a CFA with survey data to create an inventory of SC dimensions in Ghana and Uganda according to the most classical dimensions (membership, trust, neighbourhoods' connections and so on).

Brehm and Rahn (1997) in the study cited above, analyse the correlation between variation in classical dimensions of SC (civic engagement, trust and so on) and individual characteristics such as psychological involvement with their communities, general satisfaction about his/her own life, cognitive skills and economic resources using aggregate survey data in a latent framework. Paxton (1999), using a model with three factors, investigates the real level of decline in SC in the US as stated in Putnam's (1995) study. Using survey data covering twenty years, she finds no consistent support for Putnam's claim of a general decline: while trust declines in individuals, levels of trust in institutions and associations do not decline.

In a later study, Paxton (2000) uses a structural equation modelling (a more complex model for variance analysis) to further study SC and democracy, measuring SC with respect to associations' rate of participation and membership. She finds that, despite a general and positive correlation between associations and democracy, there are isolated cases of detrimental associations. The first main aim was to clarify the ties of the associations: the higher the level, the more the positive correlation emerges. Therefore, multiple memberships are better for higher interconnections and levels of SC. Paxton finds that three associations are less connected than others are: trade unions, sports associations and religious associations. In her models, she controls also for trust and given that democracy affects trust, she confirms the idea of a reciprocal effect between SC and democracy.

Kaasa and Vadi (2010) use CFA to build factors based on cultural aspects (as in Hofstede's dimension) to test their correlations with innovative performances. Van

Oorschot *et al*.'s (2006) study referred to above of is one of the main references for this work, not only theoretically but also empirically. Using data from the European Social Survey, they theoretically sort SC factors, building a model of three factors (*Network*, *Trust* and *Civism*) confirmed by the CFA, about the geographical distribution of SC among European countries and regions (North, South, West and East) and among social categories of European citizens. They found that the distribution is relatively similar between the countries apart for some exceptions in Northern Europe and that SC is strongly gendered, related to religious beliefs and to a political left-right stance.

Several studies come from the field of management studies (i.e. organizational studies), where a significant number of them use original data collected from firms at a micro-level. An important and oft-cited work is Tsai and Goshal (1998) which uses data collected from different business units of a multinational company to hypothesize factors describing structural, relational and cognitive dimensions of SC with a proper CFA model. They examine the relation both among these dimensions and between the patterns of resource exchange and product innovation within the company, finding a positive relation. Baron and Markman (2003), using data from firms and building factors with a CFA based on the items of the created questionnaire, measure how different social skills – social competence, social adaptability and their related measurements – are more effective in achieving financial success, differentiating also with respect to the type of industrial sector.

Yli-Renko *et al*. (2001) apply a CFA to build a model focused on the relational view and knowledge based definition of SC. They show that SC facilitates external knowledge acquisition necessary as competitive advantage in key customer relationships. A last, important field of application of CFA on this topic is the psychological. Here CFA is used mainly to test the internal validity of scale-measures by created items. Luthans *et al*. (2007) based on the assumption already explained that SC is related to satisfaction, analyse how hope, resilience, optimism and efficacy can be a higher-order factor predicting work performance and satisfaction. Phongsavan *et al*. (2006) build a model for three SC constructs at the individual level to quantify the degree to which high levels of SC are associated with positive mental health in Australian adults. Feelings of trust and safety, community participation and neighbourhood connections and reciprocity were found to be related to a lower risk of mental health distress, together with control over socio-economic status conditions.

We can state finally that, despite the still open debate about the reliability of these methods of estimation based on the patterns of variance and variables reducing modelling, CFA is increasingly used in the social sciences to study SC and its links with different topics. We can even presume that the multifaceted, implicit and natural definition of this concept fits perfectly with this kind of analysis that, as described above, allows for the retaining of the richness of information derived from the several options of operationalization available. The following paragraphs will outline how the goodness of fit of the two models were reached according to the most used indices in the literature, highlighting again similarities and differences between 2001 and 2011.

Di Stefano and Hess (2005) create a useful checklist for the validation process of CFA constructs, identifying four main parts and sub-parts that a researcher should check:

1) *Background/theory*: how does theory or literature support the model?

   In this part, after the explanation on how the model can be justified according to the relevant literature, aspects that are more practical should be reported: presentation of a model or diagram, a list of alternative models tested;

2) *Data screening*: how were the data treated?

   Here, the researcher should report procedures used: outliers check, item diagnostics, assumption mentioned, sample size, level of data analysed and indices used;

3) *Results*: how does the final model fit?

   The most important section should present: estimation procedure used, input matrix used (especially covariance matrix – correlation matrices, if reported, should be standardized) and all the information about latent variables: number of items, ratio of items, estimated parameters (loadings);

4) *Discussion*: how does the model respond to the initial theories?

While the first part will be covered in the next paragraphs about how I build the two models, the other more empirical parts will be the main aspects addressed when results of the models will be explained.

2.2.1 The datasets used: British Household Panel Survey and Understanding Society Survey

The British Household Panel Survey is one of the most important longitudinal studies of individual and households in the UK. Established in 1999, it was carried out until

2009, when it was replaced by the *Understanding Society* survey that has been organized in a different way and includes important new modules. Starting with a sample of around 5,000 households with 10,000 interviews covering all aspects of daily life, its sample has been continuously updated, weighted and changed to reach the goals of both a representativeness of the British population and a focus on sensitive aspects (such as having a proper sample for Wales and rest of England with respect to London and, at the same time, ensuring that minorities are properly represented). I use the wave 11 (K) of the BHPS (corresponding to the year 2001). All the variables for SC are available in a file on individual answers that presents 1,379 variables and 18,867 records. After an important and complex work of recoding aimed at keeping as many observations as possible, I test my model for 2001 on 15,856 observations. The sub-sample will not result as equal to the totality of the sample because of the listwise deletion[3].

The BHPS has been replaced by the *Understanding Society* survey, named officially as the UK Household Longitudinal Study (UKHLS) in 2009. The UKHLS includes approximately 40,000 households in the United Kingdom. Different from the BHPS, data collection for each wave takes place over twenty-four months so periods of waves can overlap. The survey is composed of a General Population Sample (GPS) and the Ethnicity Minority Boost (EMB). From Wave 2 onward the Main Survey also includes information from BHPS panels creating a proper sub-sample (Former BHPS). The particularity of this survey is the information it has because of the precise

3 Further explanations will be given in the next paragraphs, especially about the handling of these missing data.

modules introduced: Health Assessment, with information about Health, Biomarkers and Genetics and an Innovation Panel, which includes multiple experimental studies in which households are randomly assigned to a particular instrument or survey procedure. The sampling procedure is complex and provide for stratified sampling allowing equal probabilities in all the regions, in a similar way to the BHPS.

Despite the overlapping of the waves, the wave presenting SC variables for the year 2011 is the third. Considering all the sub-samples together (General, Minorities and Former BHPS), the final count of observations include 49,739 individuals for 27,783 households as released for public access.

The recoding of the original variables has been carried out in a similar manner to the 2001 survey data, especially for the variables created in a similar way between the two surveys. Table A1 in Appendix A presents the uniform recoding of variables and their labels for the membership and active membership definitions and Table A3 the other variables used. Notwithstanding the application of listwise deletion in the CFA, the final number of observations of the model has been kept.

2.2.2 Other technical issues

With the checklist of De Stefano *et al*. (2009) in mind, I end assessing several last empirical aspects. The first technical aspect is the handling of missing data for the model 2001. As will be shown in the next paragraphs, even after the recoding of original variables, missing data are present and CFA results will show a difference in the number of observations used in the analysis.

Main methods to handle missing data are: pairwise deletion, listwise deletion and multiple imputations.

Schreiber *et al*. (2006) and Harrington (2009), generally speaking, do not recommend pairwise deletion respect to listwise deletion. Even if pairwise deletion allows retaining more observations, statistical analysis is computed on different subset of cases. Besides, this aspect can imply the case of correlation matrix not positive definite that are not desirable at all for factor analysis.

So, the choice was restricted between the other two options and I opt for listwise deletion. The choice can be justified for numerous reasons. The first main reason is that imputation in the case of categorical variables, especially dummies, has been proven to be more biased than other method of handling missing data. Allison P. D. (2000, 2005) in different studies shows how listwise deletion produced less biased results respect to multiple imputed data and that these results are even more strong when predictor are categorical and missing are MAR (Missing At Random). Therefore, I decide to do not opt for multiple imputations. The second is the sample size, large enough to allow this procedure and do not lose too many observations respect to the total sample. The third is the structure of the survey itself. The questionnaire was organized with a main filter question and following sub-sections to be filled in case the answer to the first general question was positive. In this way, the recoding of the variables in the sub-sections has been completed with the generation of dummy variables for all the 'inapplicable' responses like 'event that does not occur'. The few stated missing data in the survey common for all variables (approx. 850), were mainly attributable to *proxy respondents'* answers (few to *wild or missing*). In this case, a listwise deletion would not have affected the sample size and the distribution.

One of the main assumptions to check in the case of listwise deletion – called also *complete case analysis* – is that missing data are Missing Completely At Random

(MCAR) or at least Missing At Random (MAR). While MCAR would be a strong assumption for this case (because of the structure of the survey and the variables we cannot consider the missing data just a random subset of the data), I consider that proving that missing values are MAR is the appropriate step. The propensity for a data to be missing is related to some of the other observed data, like in this case. This is also the main reason why MAR is frequently called Missing Conditionally at Random (MCR) because the missingness is conditional on another variable (in this case, the filter question variable). To check that the presence of MAR data does not affect results, I decide to check if there is a difference in characteristics of the two samples (the initial sample and the sample after listwise deletion). At the end of the Appendix of this chapter, in the section Supplementary Materials I show results of this analysis.

The second issue is about checking for possible multicollinearity between the factors indicators. According to Byrne (2012) it can be a problem in CFA if the observed variables have been hypothetically relied on in the same construct and the construct is used to predict other variables. The problem does not arise if the construct is used as a dependent variable in another model (a later SEM for example or in this study in the next models built with the factors identified with this CFA). Incidentally, to make the pre-analysis more complete the correlation matrix of variables is presented in the Table A6 and A7 in Appendix A. Only two correlations are greater than the threshold of 0.9: between two variables bout being member (active or not) of religious groups for 2001 (*active4* and *member4*) and between two variables about closeness and supporting parents for 2011 (*clospar* and *strongsup*).

Other correlations that are worthy to look at are those ones in the range between 0.6 and 0.9. From both the tables as we can see, high correlations are present between variables belonging to three main aspects:

- Variables measuring membership and active membership;
- Variables about voting;
- Variables about informal care.

All these correlations were in some way expected. Indeed, correlating the residuals of these variables in the model allows a better fit despite the risk of reducing variance. According to Geiser (2013), the correction for measurements error leads to an attenuation of the relationships at the latent level.

Finally, with regard to the estimator, Confirmatory Factor Analysis with categorical variables can be done only with the Weighted Least Squares Means and Variance Adjusted (WLSMV) estimator available only with the software Mplus (Muthén & Muthén, 1998). In the past, categorical variables have been treated as continuous in these kinds of analyses, especially in the Monte Carlo analysis. But this procedure revealed several problems of affected X^2 statistics mainly linked to the number of categories (if less than four) and the skewness of the variables' distribution (if variables are skewed and if the skewness has opposite directions) (Green, Akey, Fleming, Hershberger and Marquis, 1997; Bentler and Chou, 1987).

According to Byrne (2012) WLSMV estimator has shown to yield more accurate test statistics, parameter statistics and standard errors under both the assumptions of normal and non-normal latent response distributions. Indeed, one of the main assumptions about observed variables in CFA is that under each categorical variable there is an unobserved latent counterpart with a scale that is continuous and normally distributed. This criterion is quite difficult to meet. The analysis with categorical variables proceeds from a frequency table comprising the number of threshold, multiplied by the number of observed variables to the estimation of the correlation matrix. The problem arises with the high occurrence of 0 cells or near-0 cases. This particular estimator not only derives parameter estimates from use of a diagonal weight matrix, robust standard errors, robust mean and variance-adjusted X^2 statistics but also fits better for categorical variables with low number of categories.

These considerations took me to use Mplus, the only software providing this type of estimator and one of the most recommended for performing Factor Analysis (both Exploratory and Confirmatory) and Structural Equation Modelling especially when outcome variables are ordinal or binary.

To help further the comprehension and the development of the models, I attach in the Supplementary Material section the Mplus syntax for both models.

2.3 Identifying observed variables and latent factors: General framework

As described in the introduction and in the previous chapter and as summarized in Table 1.1 page 53, SC has several dimensions. The main classical ones are memberships and active membership, voting behaviours, attitudes and trust towards the government, and social and informal networks.

To these ones, we can add more recent dimensions according to studies about health and caring, neighbourliness, crime control, use of social media and mobility.

Strictly related to these dimensions, several measurements have been developed and applied in different and many studies. Before describing how I built my models for 2001 and 2011 using these variables, in the following paragraphs I am going to describe how these dimensions have been turned into measurements and variables and how they have been used in empirical studies, sometimes even using different methods of Factor Analysis.

2.3.1 First dimension: Membership and active membership and religious aspects

Coleman (1988; 1990) was one of the first authors to emphasize how SC can be characterized by an initial dimension regarding the belonging of individuals to groups, organisations, associations and societies. He defines SC in terms of its functions:

> Social capital is defined by its function. It is not a single entity, but a variety of different entities having two characteristics in common: they all consist of some aspect of social structure and they facilitate certain actions of actors – whether persons or corporate actors – within the structure [...] social organization constitutes social capital facilitating the achievement of goals that could not be achieved in its absence or could be achieved only at a higher cost. (Coleman, 1990, pp. 302-304).

Putnam (2001), in expanding the SC concept, introduces the concept of *informal social networks*. He specifies that we are inclined to think that the higher form of social involvement includes citizenships, memberships and the public life. However, the crucial support in everyday-life is found in friendships and other types of informal relationships such as positive neighbourliness. These can be included in the informal social networks concept. Deeply linked with this concept are the idea of *bridging SC* (Putnam 2001 and Coleman 1988) and *weak ties* (Granovetter 1973), typical of this kind of SC related to associations and organizations, where relationships should be open and serve to connect individuals to others outside of the family and friends' networks – the so called *weak ties* – and which, due to this aspect, reveal high3er levels of openness, creativity and innovativeness.

Another important difference is between *membership* and ***active*** *membership*.

The differences between these two aspects and the impact on individuals' social networks are discussed in an interesting study on participation and SC formation in Norway (Wollerbæk and Selle, 2003). Trying to understand if face-to-face interaction and direct participation have fundamental roles in the formation of SC, the authors investigate three dynamics deeply linked with this interpersonal means of interaction: how active participation can have a higher impact than passive membership (intensity), how affiliations and number of affiliation work in relation to the level of the activity; and finally, if the political/non-political aspects of organisation makes a difference.

Wollerbæk and Selle refer to Putnam's main works wherein he asserts that civic engagement is related to participation in voluntary organizations (1993; 1995; 2000). The authors anticipate a positive response to these questions but in their conclusions, remain unclear even if they are nonetheless interesting. Indeed, while their results suggest that those affiliated display higher levels of SC than outsiders did, the difference between active and passive is not strong, and perhaps even negligible. This result changes only if they add a cumulative effect of participation, defined as belonging to several associations at the same time (and this effect is even stronger if the associations have different purposes).

The importance of this characterization is also confirmed in another interesting study incorporating CFA using data from the European Social Survey sorts SC dimensions in order to study their geographical distribution across European regions and countries. Van Oorschot *et al*. (2006) find that a first-order Factor of *Networks* accounts significantly for these variables: passive membership, active membership, socializing with friends and socializing with family. The *Trust* Factor accounts for the variables measuring trust to others and trust in State and institutions and the last Factor, *Civism*, accounts for the variance of the two variables representing civic commitment and morality and discussing and following politics.

It is quite acknowledged that, in any case, this particular difference arises clearer at measurement level with the frequent identification of specific variables for both the types of memberships than being formally theorized a priori.

Another important distinction is found in the *informal/formal character* of organisations. Putnam's associations are more *informal* and are usually related to more entertainment or personal activities like sport, education, music, arts, church, charity, voluntary, ecology, human rights and peace associations. For Olson (1982), organisations are more structured and refer to associations linked more to political/structured characteristics such as professional associations, trade unions and political parties. Olson also states that the involvement in these associations is stronger than is the case for others (and probably directly more active) and has more effects on the formation of social networks and SC. Besides, it is more formal and more linked to the institutional level (Olson, 1982).

Beugelsdijk and van Schaik (2005) in a regional study of correlation between SC and economic growth in the European regions use both definitions as measurements: Putnam vs Olson groups and active versus passive membership. They found positive relations, confirming and extending the Putnam studies on Italian regions, underlining also that active involvement makes the difference respect to the mere existence of the associations.

2.3.1.2 The religious aspect

An important sub-dimension, and the subject of many different studies, deserves an apart description: the religious aspect. The hypothesis is that it can be included under the membership dimension, both passive and active, formal and informal. Several studies have focused on the relationship between religious belonging and SC attempting to understand how religion is able to establish strong ties between members. Most studies agree on the fact that it is more related to a *bridging SC* as well as other membership variables. Ebstyne King and Furrow (2004), using a Structural Equation Model, find that religiously active youths report higher levels of SC and that the influence of adolescent religiousness on moral outcomes is mediated by other classical SC resources. Wuthnow (2002) attempts to distinguish between the ‘bonding’ and the ‘bridging’ SC and finds that religious involvement is more related to bridging SC. Using American survey data, he finds that membership of a religious congregation and holding a congregational leadership position are associated positively with a higher likelihood of having *bridging* friendships. He also states that, controlling mainly for the different ethnicity groups, data do not show a correlation between frequencies of attendance but that other variables like the religious traditions and the size of congregation are significant.

2.3.2 Second dimension: Sense of citizenship, attitudes towards the institutions and voting behaviours

The aspects of SC more linked to political definitions are revealed as important and continue to be as relevant as many of the more classical aspects. La Due Lake and Huckfeldt (1998), beginning with the Coleman's structural and functional definition of SC, test the hypothesis that the politically relevant SC, defined as SC facilitating political engagement, is generated in social networks where there are social interactions between individuals and proper citizen discussants. Moreover, these interactions, characterize the amount of SC through the level of political expertise present in social networks, the frequency of political interactions within the network and the size of this network, showing that human capital cannot be the only explanations of active behaviours at the political level. Higher levels of SC involve both higher political active participation and higher rates of voting. Duke *et al.* (2009) show that strong connections in family and community contexts, especially during adolescence, predict a greater likelihood of voting and involvement at the social level, endorsing civic trust in young adulthood.

Finally, yet importantly, all these considerations and studies can be summarized under the wider and classical definition of *institutional trust* that can be defined as the vertical trust of people towards the institutions in terms of the degree to which it is believed that they truly serve people through their work and actions. Derived from the important work of Putnam *et al*. (1993) together with the concept of *generalized trust* (the horizontal trust between individuals at more or less the same level) (Ackomak and ter Weel 2005, 2009; Dakhli and De Clercq 2004), they effectively justify the structure and use of these variables. Both for 2001 and for 2011, questions are raised about the trust towards the Government and its main institutions, the conviction that people can contribute effectively to the wellbeing of their country exerting their civic duties and rights in an active way and about their voting behaviours and parties' closeness.

2.3.3 Third dimension: Health and caring and informal networks

The third dimension identified is about health and caring provided inside and outside the household. Under this definition, we can include both a normal activity of temporary caring or supporting and more structured and permanent activities of caring, like for example, long-term illnesses and disabilities.

This dimension is therefore naturally linked with the informal networks that an individual owns. Usually these links are stronger connected to the household's network and closer acquaintances (like closer friends and neighbours).

Discussions surrounding links between SC and health and caring have increased definitively in recent years. The emerging problem of the sustainability of public health systems on a global scale and the fairness of private or semi-private systems has come to focus more on possible alternatives to tackle the problems deriving from the process of *informal care*.

With this term, the literature identifies the amount of caring for old people, disabled or seriously ill people carried out by family or close networks. In reality, this aspect becomes central in discussions on the reduction of costs for whole health systems and many studies have confirmed this perspective. Just to cite some, Rose (2000), using Russian self-assessed individual data, shows that human capital and SC account for a notable amount of variance in health. When both forms of capital are combined in a composite model, they have even more influence, demonstrating that SC does make an independent contribution to health. She also shows how the influences of SC include involvement or exclusion from formal and informal networks or friends to rely on when ill – the *informal care networks* – and also on more personal dimensions such as control over one's own life and trust. Finally, focusing on the joint effect, she finds that SC then increases physical and emotional health more than human capital but that together they can easily raise an individual's self-reported health from just below average to approaching good health.

Hendryx *et al*. (2002), using American data and trying to test the main hypothesis that variation in reported access to healthcare is positively related to the level of SC present in a community, confirm this positive correlation and identify a strong 'metropolitan' effect, whereby this means that the relation is even stronger for individuals living in the metropolitan areas of the sample. Cattell (2001) using a more qualitative approach through deep interviews conducted in two similar deprived areas, looks at the dynamics between poverty, exclusion, neighbourliness, health and wellbeing. Considering several dimensions of SC – networks' typologies with respect to structural and cultural aspects, density and level of heterogeneity – she develops the interviews to address how this complexity is related to the role of another three factors that affect an individual's health: neighbourhood characteristics and perceptions, poverty and social exclusion, and social consciousness. She finds that heterogeneity is a better predictor – as much as participation and perceptions of inequality. However, she finds mainly a partial correlation: despite the capacity of SC to buffer its harsher effects, the concept is not wholly adequate for explaining the deleterious effects of poverty on health and wellbeing.

Barrett *et al*. (2014) discuss the theme of family care and SC and the related transitions in informal care, linking all the classical characteristics of SC (bridges, bonding, micro approach, macro approach, ties, networks) and the different aspects of informal care: subjects that are expected to carry it out (family, community, professionals and policies) and subjects that receive it (children with disabilities, individuals with lifelong disabilities, old people). Mohnen *et al*. (2011), using a multilevel analysis and Dutch survey data, confirm that in neighbourhoods where SC is higher, people are more active and more likely to have healthier habits and behaviours (e.g. being non-smokers). Therefore, the direct effect of neighbourhood SC on health is reduced significantly and strongly by physical activity.

All of these studies support the idea that SC is strongly connected, mostly in a positive relation, with the familial and local provision of care, where local chiefly refers to the neighbourhood level.

2.3.4 Fourth dimensions: Neighbourliness and local networks, crime control effect, use of social media and mobility

In this last paragraph, I am going to identify variables and related uses of more recent dimensions respect to the more classical ones. These dimensions have been studied increasingly over the last years. One main consequence, as we will see, is that the availability of variables and measurements about these dimensions only for 2011 model.

2.3.4.1 Neighbourhoods' relations and networks

As hypothesized and outlined above, neighbourhood SC, family networks and individual health are strongly correlated and interdependent. Like a fundamental resource embedded within the families and between them, reaching the neighbourhood local level, all these aspects can be easily loaded on one or more factors. These kinds of 'informal' networks are frequently identified as *Bridging SC* (Putnam, 2000), characterised by those *weak ties* described by Granovetter (1973) that can be useful in increasing SC due to more heterogeneous social ties. These kinds of social ties usually imply more novel information and broadened worldviews (with respect, for example, to the *strong ties* of the family networks – called also *Bonding SC*). This kind of SC, as Granovetter describes, can sometimes limit the openness of individuals outside the parental and relative network despite its usefulness in other cases such as health and care internal to the household. Carpiano and Kimbro (2012) find evidence of a 'negative SC' when testing the hypothesis that, especially for female primary caregivers of children, neighbourhood SC moderates the influence of parenting strain on mastery, interpreted on the basis of individuals' understanding of their ability to control personal life circumstances. Using American survey data and controlling for different types of SC, they find a negative relationship between parenting strain and mastery that become even worse as informal social control increases. Social support and informal social control, however, buffer this parenting strain-mastery relationship when caregivers have stronger ties to neighbours. Thus, they define this mechanism in terms of ‘negative SC’ and justify more nuanced considerations of neighbourhood SC’s health-promoting potential.

Portes (1998), identifies four negative consequences of *Bonding SC* as can be derived from the studies of Bourdieu, Putnam and Coleman: exclusion/barring of outsiders of groups (appropriate examples are the various ethnic groups or categories and professional associations), excess claims on group members, restrictions on individual freedoms and downward levelling norms. Recent studies have further described how these relations are built and maintained between neighbours and within local communities. *Generalized trust*, the type of trust we regard as being typical of this dimension, is consistent over time so ‘…it is not simply and ephemeral notion with unknown consequences’ (Li, 2015b, pg. 55). According to this study, it is not dependent upon reciprocity or group membership but generalized trust matters where inequality matters and leads to charity and volunteering. Uslaner (2015) finds that trust in any case increases with membership even if it is not directly correlate in a causal way (by twenty-five percent for each additional membership). Sturgis et al. (2012) using BHPS in a longitudinal study find really little evidences to support the view that generalized trust results from integration within social networks, either of either a formal or an informal nature.

In a famous study, Li *et al*. (2005) state that many studies focus more on SC defined as civic engagement and membership whereas lack of data and means for investigation of informal social networks remain an actual problem in studies on SC. They build three latent variable scores defined as: neighbourhood attachment, social network and civic participation. The first two are more informal with the last considered more formal. Analysing the socio-cultural determinants of the three variables and their impact on social trust, they show that these determinants affect SC generation. Indeed, people in disadvantaged positions build up SC from weak ties whereas those in advantaged positions do so from formal civic engagement. They also find that SC has an effect over and above people's own socio-cultural positions. Informal social networks, especially having good neighbourly relations, tend to foster greater trust than does formal civic engagement.

Without lingering on this aspect, it seems from the majority of studies that the effect of SC is derived from good relations in neighbourhoods and informal networks and is, more frequently, positive. Furthermore, sometimes it works more than formal membership and civic engagement in fostering SC.

2.3.4.2 Crime control effect of SC

In recent years, many studies have begun to focus on the effect that SC has on crime. Akçomak and ter Weel (2008), using Dutch data, show that higher levels of SC are associated with lower crime rates. Using several variables derived by the literature (for example blood donations) and using an IV OLS estimation, they also test which dimensions of SC are more related with this problem and they find out that population heterogeneity of the local system analysed (in this case municipalities), religiosity and level of educational attainment are the most important.

Buonanno *et al*. (2009) confirm the idea that higher levels of SC are correlated with lower levels of crime. Besides, their paper attempts to also deal with all the empirical problems derived using data and variables for this topic: endogeneity, omitted variables, spatial correlation and measurement errors. They partially solve this problem using Italian data. Italy, indeed, shows a high variance in social and economic characteristics and, at the same time, homogeneity in policies and institutions. Using variables measuring associational networks and civic norms, they show that there is a significant negative effect of both on property crimes.

Lorenc *et al*. (2013) show that several factors in the physical environment are perceived as impacting on the fear of crime (e.g. visibility, signs of neglect, built environment) but that factors relating to local social environment such as social networks and familiarity seem to be more important drivers of the fear of crime, leading even to limitations on physical activities like going out. They also find evidence of the significance of factors at national level (i.e. national policies). Brunton-Smith and Sturgis (2011) analyse how neighbourhoods influence the fear of crime. Using the British Crime Survey in a complex, empirical multilevel model, they find evidence that individual-level differences in the fear of crime are negatively associated with cohesive neighbourhood social and organizational structures and collected signs of neighbourhood disorder (even visual, like declining areas).

These studies, like many others, present two interesting aspects: SC has a deterrent effect on crime rates and this effect can work in two ways. The first is a kind of 'stigmatizing effect': in neighbourhoods, community or local networks where SC and mutual trust are high, the incentive to commit crime and be excluded by the supporting community is lower. The second aspect is even more practical: as Buonanno *et al*. (2009) state, associational networks may offer official cover to criminal activities but more frequently they increase returns to noncriminal activities, raising detection probabilities.

Because of this, many recent surveys, including the British Crime Survey, inspect not only the proper, 'numerical' aspects of individuals' and households' experiences of victimisation in relation to different types of crime [4] but also inquire into the 'neighbourhood effect' on feeling safe when going out or walking in the dark and so on (Brunton-Smith and Sturgis, 2001), enriching, to some extent, the mere statistical treatment of this problem (classical studies using crime's records from Police forces for example).

2.3.4.3 Use of social media

The development and global diffusion of the use of social media is a recent event. Several recent studies reveal different aspects of this topic and we can notice that many of them are focused on the most famous social network: Facebook. In this work, incidentally, I am going to use variables from the UKHLS 2011 survey representing

[4] The types are usually related to crimes against the person: attack, insult, harassment, physical violence, burglaries, robbery and so on.

more general ‘social websites and networks’. Burke *et al*. (2011) and Ellison *et al*. (2007) document a relationship between the use of the Social Networking Sites (SNS), Facebook, and increased levels of SC, particularly *bridging SC*. Tong and Walther (2011) find evidence that SNS reduce the cost of maintenance of the relationships. Ellison *et al*. (2014) in attempting to explore the relationship between *bridging SC* and Facebook users using American survey data, deepen the study identifying the differences in levels of friendships with respect to Facebook-enabled communication behaviours. It is then plausible relating this aspect to the informal networks. But there is no agreement on the sign of this relation.

Many authors state that even if we can consider social media as an alternative means of interaction, their intensive use subtracts real time spent outside with people so we can find negative correlations between use of media and interactions with people (family, friends or acquaintances). References for these aspects are numerous and related further to psychological problems of interpersonal relations. Rubin *et al* (1985) set out to establish a link between a lesser degree of interpersonal communication, a higher degree of loneliness and parasocial interaction resulting from increasing number of hours spent in front of a television. Some may argue that the use of television differs from the potential ‘social’ dimension of SNS. Nie *et al*. (2002) find that Internet use at home has a strong negative impact on face-to-face interactions and time spent with family and friends, while Internet use at work is strongly related to decreasing time spent with colleagues. Even more so, time spent on the Internet during the weekend decreases further the time spent with family and friends. Caplan (2003) further confirms this hypothesis also revealing the more problematic psychological implications of this modality that lead to personal state of depression and, at the same time, problematic use of the Internet.

The last interesting aspect is that, according to more recent works, the use of social media also helps the crime control effect, especially at the level of personal perceptions relating to the fear of crime. This aspect further strengthens the hypothesis on the third latent factor. Uysal (2014) examines the predictive roles of social safeness and flourishing in connection to the problematic use of Facebook, finding a negative correlation. The author even finds that the two aspects can be predictors of problematic use. Kohm *et al*. (2012) find partial evidence that the amount of exposure to specific news media – newspapers, television, radio and Internet – can affect the fear of crime, increasing particularly the fear of violent crimes.

2.3.4.4 Mobility

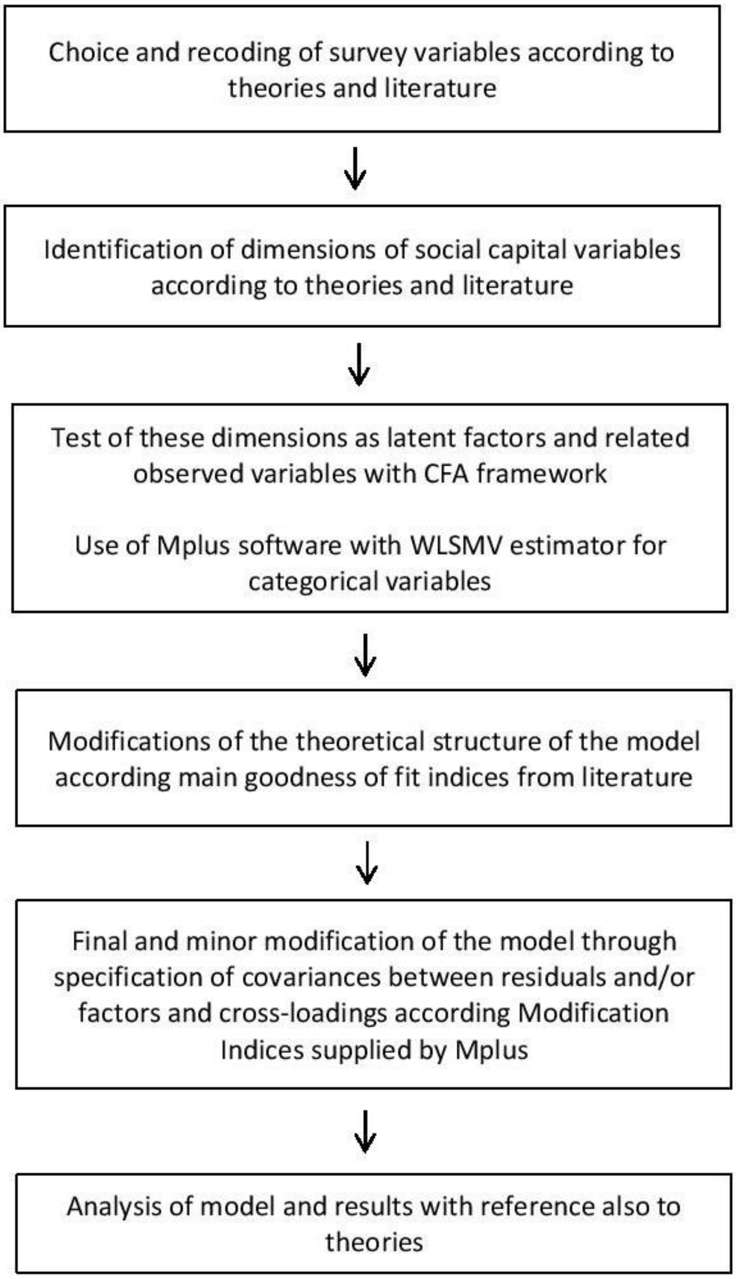
The topic of SC and mobility has been widely explored, both at international migration level and at more localised geographical levels. In a qualitative study, Kaufmann *et al*. (2004) find that spatial mobility not only refers to physical movements of goods or people but can be considered as a structuring dimension of social life that shapes important societal changes. It can be thought of as a proper asset and, consequently, differs in access opportunities, competences and appropriation. Kaufmann and colleagues define it in terms of a proper capital than can be exchanged for other forms of capital, such as economic, cultural and social. But it differs from all of these because it refers both to vertical and horizontal dimensions of social positioning. They underline how, more so than SC (which can also be referred to in terms of horizontal and vertical dimensions), it represents a new form of social inequality, whereby local and geopolitical context is even more emphasized.

Using European data, David *et al*. (2010) discover interesting connections between membership, informal networks and mobility. While in the North of Europe the club membership is higher than in the South, the frequencies of contacts with relatives, friends and neighbours are lower. Identifying the cause in the lower rates of geographical mobility of Southern people, they build a proper model of SC and mobility. They find that workers invest time and efforts in building networks and make them wider by adding ties when they do not expect to move to another region. They find also that employment protection, even in case of high rates of unemployment, nonetheless reinforces the accumulation of local SC and reduces mobility. Generally, they demonstrate how SC is more associated with lower mobility.

Teachman *et al*. (1996), examining which dimensions of SC affect the likelihood of US children dropping out early from school, find one causal element to be the decrease of SC deriving from the number of times that child changes schools, normally depending on the history of the family's moving. In a famous and interesting qualitative study, De Souza Briggs (1998) examines how SC can be linked to geographical areas or even urban opportunities. In addition, particular housing policies can have positive or negative effects. Observing and interviewing African-American and Latin-American adolescents in an area of New York subjected to a peculiar housing program, he finds that the 'movers' appeared to be more cut off in terms of SC than ‘stayers’. One of the main reasons found is that participants reported fewer chances to access good sources of information on jobs or school advice, decreasing their opportunities to enhance their situation. They also show how the perception of lack of access to such leverage changes completely when an employed adult is added to their circle of significant ties.

After the depiction of this wide framework, it is now possible describe how I build my models and relative results. In the following Figure, I outline the general strategy and then the two following paragraphs will describe empirical applications of this strategy for 2001 and 2011.

Figure 2.1: Strategy for models' derivation



2.4 Model for 2001: strategy, analysis and results

After the description of the methodology – CFA –, the dataset – BHPS – and the variables and measurements for each SC dimensions used in the literature, it is now possible to focus on the empirical building up of the model, beginning with 2001 data. All the variables used, with labels and recoded names are reported in Appendix A (Table A1 and A2).

In accordance with the empirical studies listed in the previous paragraphs, I build my theoretical latent factors based on several variables available from survey and I confirm them with the current CFA.

Starting from the variables about membership and active membership dimension, BHPS survey presents sixteen variables representing membership and sixteen for active membership to these organizations or groups: political parties, trade unions, environmental groups, parents associations, tenants or residents groups, religious groups, voluntary groups, other community groups, social groups, sports clubs, women's institute, women's groups, professional organisations, pensioners' organisations, Scout/Guides organisations and other general organisations.

In order to retain the richness of the available information, I first summarise the original variables by types of organisations, according to the following groupings (see Table A1 in Appendix A):

1) *Political and professional organisations*: parties, trade unions, professional and pensioners' organisations (*member1* and *active1*);
2) *Social organisations*: environmental, parents, voluntary, communities, women groups (*member2* and *active2*);

3) *Local organisations*: tenants or residents and social groups (*member3* and *active3*);
4) *Religious organisations* (*member4* and *active4*);
5) *Sport organisations*: sports club and Scout organisations (*member5* and *active5*).

The five types were created both for the membership and for the active membership[5]. As we can see from the Table A4 in Appendix A, the univariate proportions and the counts for these categorical variables show ordinal variables reaching even five items. Where variables are ordinal and not dummies, the recoding has been done in order to ensure uniformity between all the variables: increasing levels of SC for each variable are expressed in the increase of the item from 0 to higher level of SC. Subsequently, for example, variables measuring expected negative levels of SC have been recoded in the opposite way with respect to the other variables[6].

This is because I decide to sum the original variables up: so, for example, *member1* assumes value of 1 if the individual is a member of only one of the organizations considered for this variable (in this case a political party, trade union, professional organisation or pensioners organisation), 2 if he or she is a member of at least two organizations out of the four listed and so on. The same approach has been also used for the sum of the *active* membership variables. This mechanism allows me to keep

[5] Table A1 about the creation of these variables from the original in the survey is available in Appendix A.

[6] For example, *govern2*, a variable that asks for agreement or non-agreement to this statement 'ordinary people cannot influence government' has been recoded in terms that 1 – that is 'No' in the survey – would have had the highest score of SC. Indeed, replying 'No' or expressing disagreement in this case means demonstrating trust in the institutions.

even the intensity of membership for each singular individual. Paxton (2002) finds evidence of the importance of multiple memberships to measure ties and connections between associations. Where associations are closer than others, there is even a negative impact on democracy level.

Focusing further on the religious aspect of the membership dimension, despite the presence of a variable representing membership and active membership to religious groups or church organisations, I choose to retain also the variable related to a more subjective way of living in this personal dimension but still being able to create fundamental and strong networks (including them in the same factor). Indeed, the *relig* variable is used to describe the attendance or non-attendance at religious services.

Regarding the second dimension about sense of citizenship, attitudes towards the institutions and voting behaviours, we have three dummy variables (*govern1*, *govern2* and *govern3*) for being in agreement with general statements on government actions and principles and four variables (*vote1*, *vote2*, *vote3* and *vote4*) on voting behaviour, closeness to a party and level of interest in politics (see Table A2 in the Appendix A).

With regard to the third dimension, nine variables have been recoded and tested. Most of them are related to the help and support that the individual can receive from someone outside the household because the recoding on variables about health and caring conditions internal to the family did not allow their retention in the model. One variable, *carenr*, is about providing care to someone outside the family and only another variable is about informal networks and the frequency of meeting people.

Finally, on the last dimension, only two variables about liking and talking to the neighbours and one variable about the preference to move to a new house have been included in the final model to confirm.

2.4.1 Strategy for building up the model

With regard to the first two classical and important dimensions, initially I hypothesized a model with four factors: a first factor about membership and active membership (with religious variables included), a second one about citizenship (institutional trust, voting behaviours and similar), a third one about health and caring related issues and a fourth one about local networks and relations with neighbours, mirroring the dimensions (classical and recent) described in the previous paragraph.

On the basis of Mplus first results, I immediately found that in this way the model did not converge in first instance. Mplus provides Modification Indices (MI hereafter), as proposed by Sörbom (Byrne, 2012). Their essential function is to identify parameter constraints that are badly chosen, causing a model misspecification. As such, all parameters are assessed to identify which parameters, if freely estimated, would contribute to a significant drop in the X^2 statistics. Usually, MIs higher than 100 indicate the first variables needed to be modify or dropped out. The second highest range is between 70 and 100. Values of MIs below 70 indicate not necessary changes.

According to the modification indices available in Mplus, residuals covariances have been added especially between the variables of the same latent factor and between all the variables of membership and active memberships (even if in different factors).

In this case, adding up further specifications such as these helped the final fit of the model according to the highest thresholds of the indices used given that they are all significant with a P-value of 0.001. Besides, this can be justified by the structure of the survey used and a high, natural correlation between these two groups of variables. Modification indices were also useful to identify the weakest relationships between variables or between factors at the start of the model building. One initial important step was the change of the model from one with four factors (membership, active membership, citizenship and social network) to a model just with three factors (Wang and Staver, 2001) where the last two dimensions would have better fit as a one factor. Indeed, the second version of the model proposed with three factors immediately converged with the main goodness of fit indices for CFA showing these values: RMSEA index around 0.078 and CFI and TLI around 0.80 (description of these indices will be reported in the Results paragraph).

The first major leap in the assessment of the model was derived from the correlation of the latent variables between themselves and the correlation of the residuals of the variables about membership and active membership: RMSEA decreased further around the critical value of 0.05 and CFI and TLI around the other critical value of 0.90. As will be described in more detail in the next paragraphs, these are the cut-off values identified in the literature such as goodness of fit threshold for these indices.

To reach even better values, I then reformulated the model looking more at the *formal/informal* character of membership. And indeed, the variables regarding political and professional organizations and local organisations - *member1* and *member3*, with the equivalent *active3* and *active3* -, fit in the same factor better than the model with the first factor loading onto all the variables about membership together. Indeed, all variables are explained appropriately by the same factor of being a member (or active member) of political parties, trade unions, professional organizations, pensioners' organisations, tenants or residents' groups or social groups.

All these organizations and associations are, indeed, more structured, formal and related to political interests and general community affairs, as previously described as in Olson typology of organisations (1982). According to this hypothesis, I further checked MIs. They confirmed this idea showing how formal membership variables would better fit in one factor together with all the variables representing voting behaviours and citizenship towards institutions. This final solution saw therefore the fit of a first latent factor loading on formal membership (passive and active) and related variables of institutional trust and citizenship.

According to these studies and dealing with the empirical results of the analysis, I then decide to keep the two aspects of passive and active membership (quite critical given the high possible linearity between them) crossed with the types of organisations – Putnam organisations and Olson organisations – for the building of the model. It is possible see the underlying link between membership and attitudes towards formal institutions.

The other variables regarding more informal associations (*member2*, *member4, member5*, *active2*, *active4* and *active5)* (even personal such as religious ones) fit better in a second factor with a further variable for attendance to religious services (*relig*), probably linked with membership (active or not) of religious associations.

Parallel to this step by step work of building up the model, analytical strategy's description focuses also on the third factor. As previously sketched, a unique factor loads better on health and caring variables together with local networks and relations with neighbours. Consequently, the third and last factor fits in expressing the common pattern of variance of nine variables for providing caring to someone else and receiving help or support from outside when in a difficult situation (both at a psychological and practical level): *carenr*, *help1ext*, *help2ext*, *help3ext*, *help1net*, *help2net*, *help3net*, *help4net* and *help5net* (see Table A2 in the Appendix A) and four variables representing good relations with neighbours load on a factor in a significant way. Among these four variables, there is one variable about the willing to move from the current residence inversely related to the others.

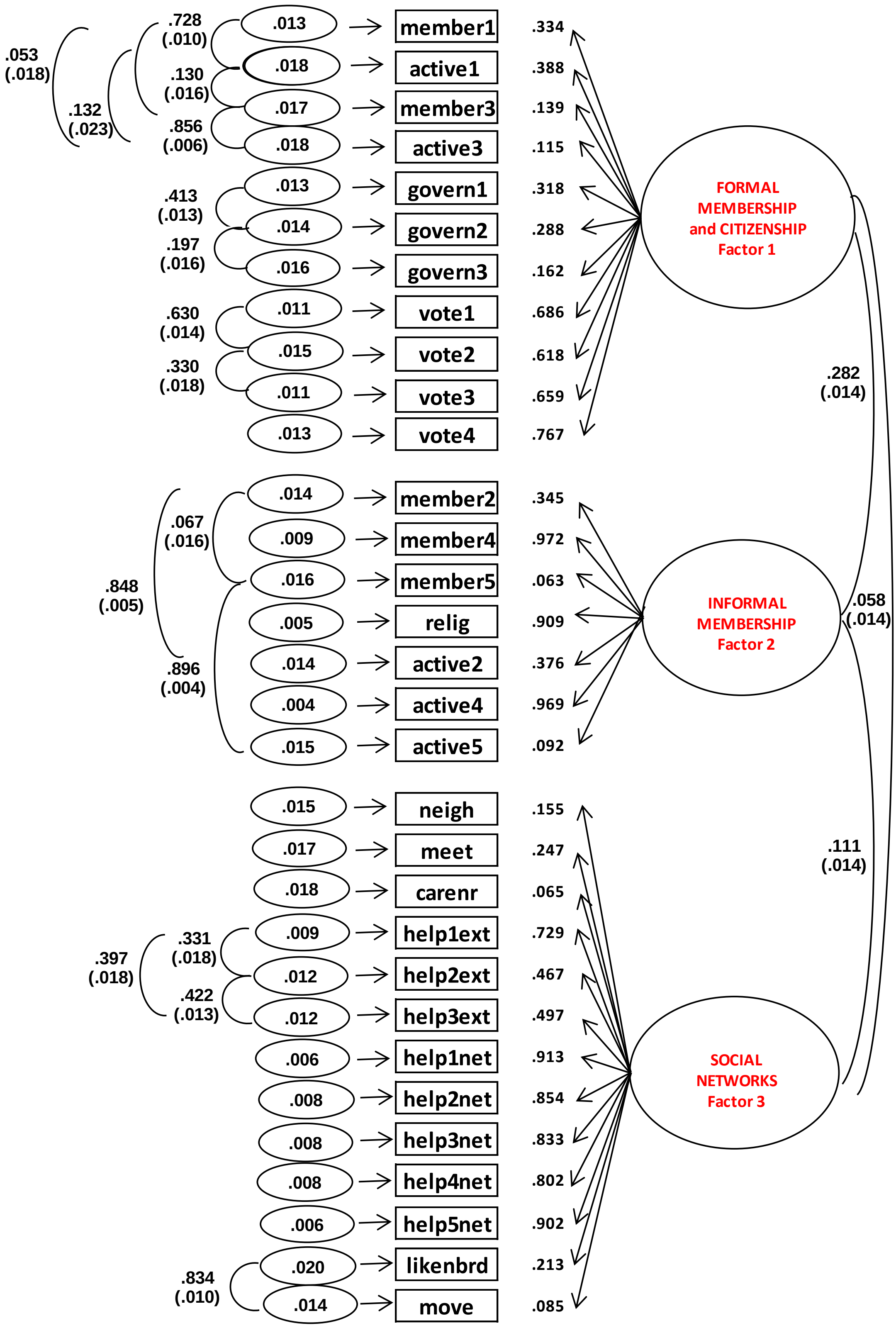
The two levels show, therefore, to be deeply linked even if overlapped in the daily life of an individual. This hypothesis has been also confirmed by the already cited study of Van Oorschot *et al*. 2006.

Finally, it is now possible to show the final fitted model: three main latent factors loading on thirty-one variables.

I name the three factors according to the characteristics of the variables on which they load:

- Factor 1: *Formal membership and citizenship*;
- Factor 2: *Informal membership*;
- Factor 3: *Social networks*.

Figure 2.2: Final model for 2001



Values shown between variables and paths from Factors represent standardized β coefficients. Other values are error terms. Significance at 0.001

Description of the strategy of building up the model implies also identification of empirical aspects like correlated covariances and cross-loadings. As previously outlined, these types of specifications cannot be theorized a priori and confirmed by CFA but they emerge from analysis of MIs.

As we can see from Figure 2.2, the final model proposed is complex and it is composed by thirty-one variances (equal to the number of variables), two covariances (between Factor 1 and Factor 2 and Factor 1 and Factor 3) and thirty-four residual covariances for a total of ninety-eight parameters. The last step to add different residual covariances was due to the structure of the survey and they have been identified using MIs. In the Mplus syntax (as reported in Supplementary Material section) they can be identified by the statement WITH. With this statement, general covariances are suggested by MIs and they can set as covariances between residuals of the observed variables or covariances between latent factors.

After the model fitted with the structure of three factors loading on the observed variables described, the first important jump towards good values of main goodness of fit indices was reached with the set of covariances between the latent factors and covariances between residuals of membership variables with active membership variables internal to the first and second factors.

For these cases, MIs showed quite clearly the importance to set such correlations (highest values of MIs reported among all the other, even greater than 120).

Following, the second important change was due to the identification of covariances between the variables about attitudes towards Government and voting behaviours in Factor 1 (MIs between 100 and 120).

Last definitive improvement to the values of indices recommended in literature for a really goodness of fit was reached thanks to the correlation of residuals between variables in the third factor (MIs between 70 and 100).

After these three steps, further slight improvements up to the threshold suggested for fit indices were due to the specifications of covariances between residuals of observed variables loading onto different factors that I am going to describe in the next paragraph (see Table 2.1). Most part of them is related again to membership and active membership variables, supporting the idea that even if the character of informal versus formal membership results primary in this analysis, this characterization is still strong.

Given the complexity of the model and the important number of residuals correlations, I report the residual covariances and their standard errors of variables loading on different factors in the following table, in order to make the diagram above more readable.

Table 2.1: Residuals' correlations – 2001

Variable	Estimate	S.E.
MEMBER1	with	
MEMBER2	0.265	0.015
MEMBER5	0.253	0.015
ACTIVE2	0.172	0.016
ACTIVE5	0.172	0.015
MEMBER2	with	
MEMBER3	0.137	0.019
ACTIVE1	with	
MEMBER2	0.243	0.021
ACTIVE2	0.258	0.021
VOTE3	with	
MEMBER2	0.263	0.020
VOTE4	with	

MEMBER2	0.329	0.026
ACTIVE2	0.287	0.026
HELP2EXT with		
MEMBER5	0.187	0.016
ACTIVE5	0.191	0.016
MOVE with		
RELIG	0.458	0.036
VOTE4	0.266	0.022

All the correlations are significant at 0.001

According to Schreiber *et al*. (2006) and Pohlmann (2004) for one sample analysis, the ratio between the number of observations and the number of parameters to be estimated should be higher than ten. Here the ratio is 161.8, definitively higher than the threshold suggested. In Appendix A, I report table for: univariate proportions and counts that are more useful in this case of categorical variables (Table A4).

The high goodness of fit, besides, made me stop to add further specifications despite of high values among the BY statements. These statements usually suggest cross-loadings of factors on observed variables from other factors and they are considered as changes slightly more structural than correlating residuals (Byrne, 2012). Indeed, adding many significant correlations make the model progressively dependent on the specific sample.

2.4.2 Results: analysis of three factors characteristics and main indices of goodness of fit

In this paragraph, I am finally going to describe and interpret the model obtained.

With regard to the latent factors, from the Table below we can see the correlation and covariances between factors. The Table 2.2 show that the factors are positively correlated, as I expected, but the correlation matrix highlights how the correlation is stronger between Factor 1 - *formal membership and citizenship* - and Factor 2 – *informal membership*. Factor 3 – *social networks* - seems slightly correlated with Factor 2 and weakly with Factor 1. This result can be explained by all the previous descriptions of variables and the reasons behind their selection. As expected, Factor 1 and Factor 2 share the strong 'membership' aspect of SC, whereby the first loads on variables measuring more formal types of membership, the second factor measures the more informal one. At the same time, Factor 2 loads on variables representing more 'informal networks', *bridging SC* and *weak ties* and I explained how, theoretically, they may be deeply linked with all the variables for *health and caring* and *neighbourhoods' relations*, that are the variables loading significantly on Factor 3.

Table 2.2: Correlation matrix for the latent variables – 2001

	Factor 1	Factor 2	Factor 3
Factor 1	1.000		
Factor 2	0.282	1.000	
Factor 3	0.058	0.111	1.000

Table 2.3: Covariance matrix for the latent variables – 2001

	Factor 1	Factor 2	Factor 3
Factor 1	0.112		
Factor 2	0.033	0.119	
Factor 3	0.003	0.006	0.024

The covariance matrix shows that Factor 1 and Factor 2 have high variance with respect to the third factor (0.112 and 0.110 respectively) and that they also co-vary together. Standard deviations confirm this higher internal dispersion of data (almost the double, even more, respect to 0.11 of Factor 3). A second important aspect to take account of is descriptive statistics and distributions. Graphics and box plots below are then reported.

Table 2.4: Descriptive statistics for Factor1, Factor 2 and Factor 3 – 2001

	Obs.	Mean	Standard Deviation	Min.	Max.
Factor 1 *Formal membership and citizenship*	10,155	0.0093	0.26	-0.424	0.858
Factor 2 *Informal membership*	10,155	0.0173	0.20	-0.259	0.832
Factor 3 *Social networks*	10,155	-0.0144	0.11	-0.378	0.141

Figure 2.3: Frequencies for Formal and citizenship factor (Factor 1) - normal distribution, 2001

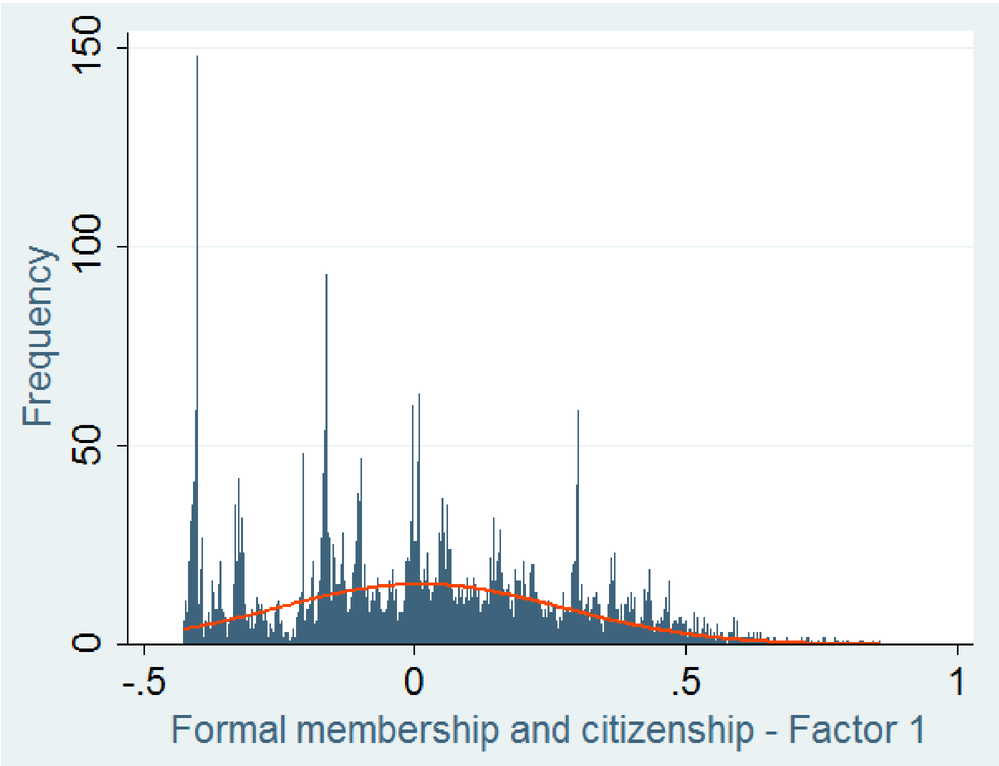


Figure 2.4: Frequencies for Informal membership factor (Factor 2) - normal distribution, 2001

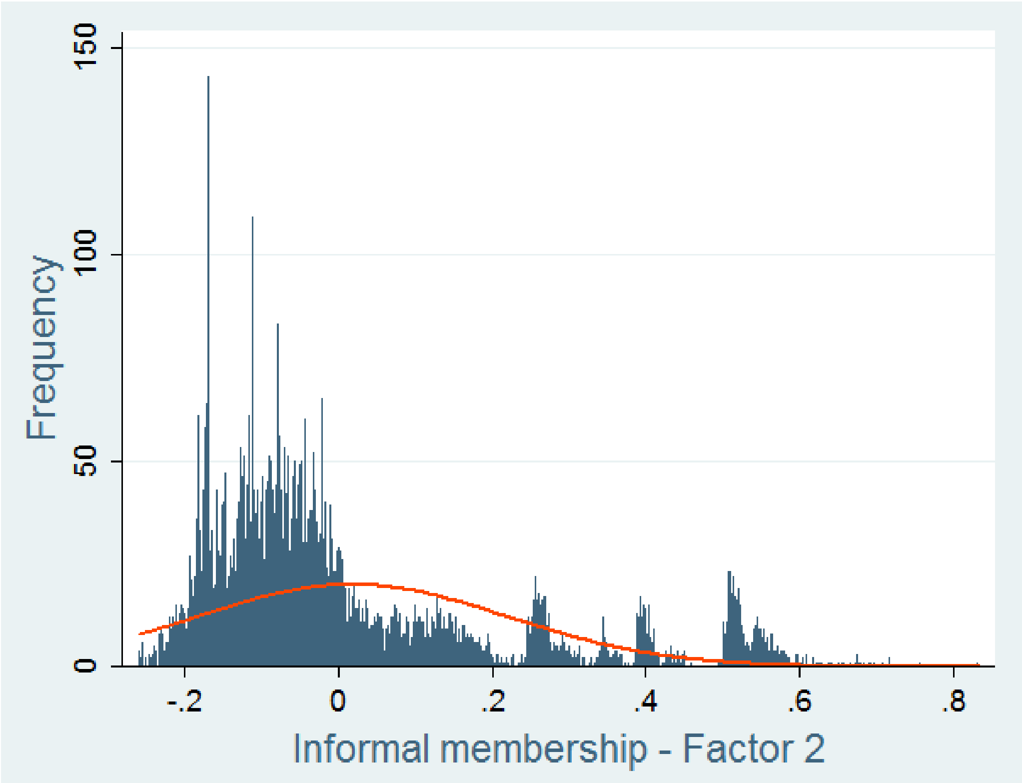


Figure 2.5: Frequencies for Social networks factor (Factor 3) - normal distribution, 2001

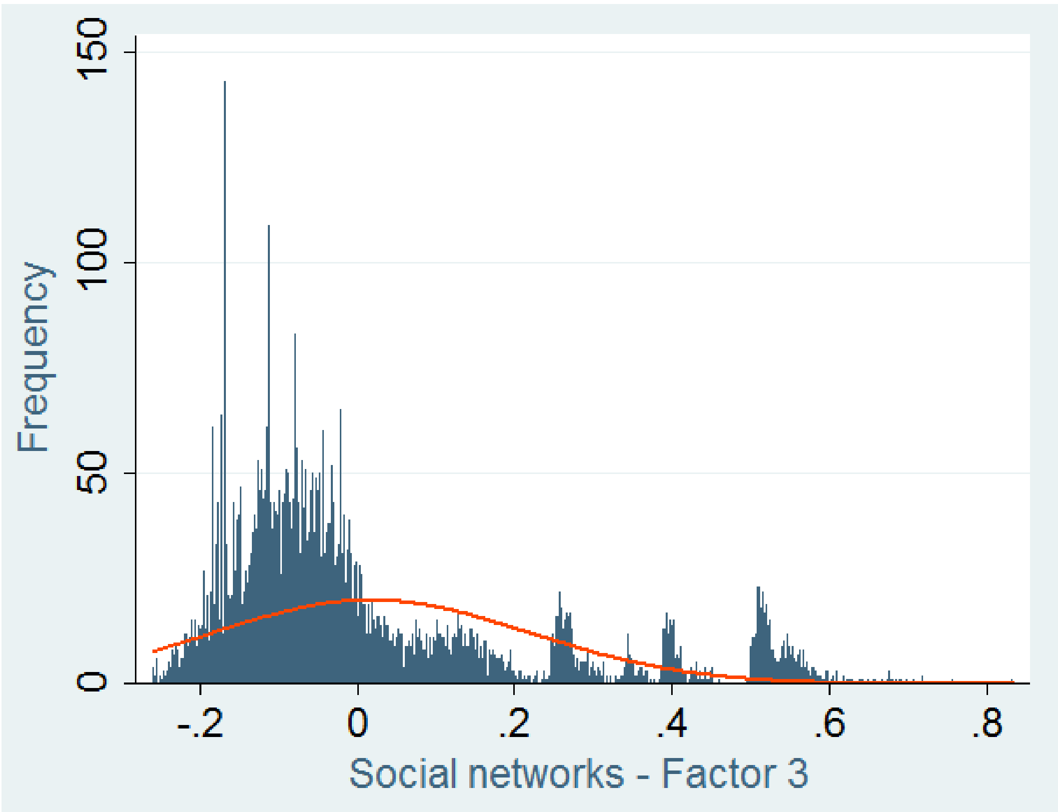
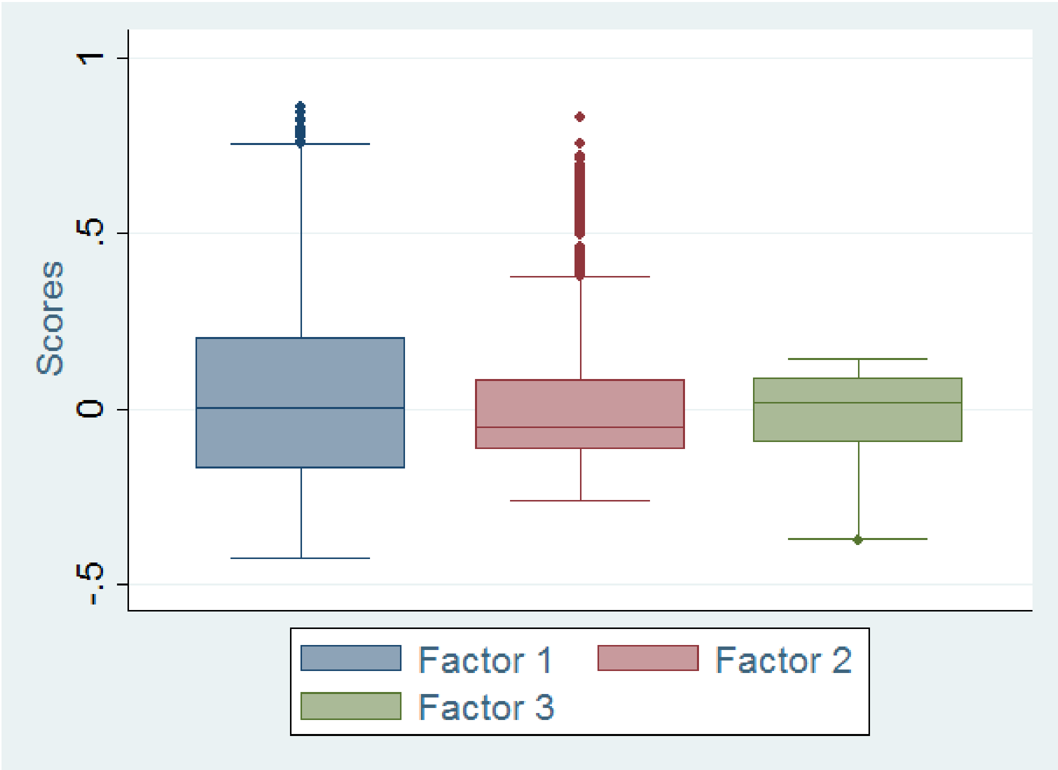


Figure 2.6: Box plots for the Factor 1, Factor 2 and Factor 3 – 2001



Starting from the distribution, we can explore the box plots in order to identify immediately if there is the need to deal with outliers or not. They clearly show that not one of the latent factors can be defined as normally distributed and the box plots also show the proportion of outliers. The box plots show the outliers and the comparison of the distributions for the three factors together. We can observe how the three factors are different mainly in the larger values and Factor 1 seems to be the most skewed between them. But, despite the outliers appearing as several and important, given the data about mean, standard deviations as from Table 2.4 and especially minimum and maximum value, it is possible to keep the outliers thanks also to the tight range of scores.

The second aspect to check, and the most important, can be described by looking at the three graphs of the frequencies (Figure 2.3, 2.4 and 2.5). They show a notable number of observations on negative scores for all three factors considered (especially the third factor of *Social networks*) and, from Table 2.4 about descriptive statistics, we can see that Factor 3 even shows a negative mean. Besides, for all three factors, the mean is close to 0 (or less) with a remarkable standard deviation if related to the scores and the minimum and maximum values. Indeed, we can see how, apart for Factor 1 that has a more spread distribution, Factor 2 and Factor 3 show a high number of frequencies just below their mean.

According to the theory, the factor scores are composite variables that retain information about an individual's placement on the factors. Di Stefano *et al*. (2009) explain clearly how the score of 0 for a factor means that the individual's ratings of the importance for that attribute is close to the average of the sample. Similarly, a negative score places the individual under the average rating for that factor but this does not necessarily mean a negative value. Undoubtedly, the sample average is not necessarily the middle of the scales.

Various authors have proposed several ways to deal with this aspect but in a more exploratory fashion. They identify several methods mainly classified according to two main groups: *non-refined* methods and *refined* methods. The non-refined methods are simple and cumulative procedures to provide information about an individual's placement on the factor distribution. They are easy to compute and interpret but are less exact than the refined methods which are more complex techniques that also provide standardized results. These methods are based on a sum of scores, standardization or cut-off above an arbitrary value. One first drawback is the risk of reduction of the variability in the scores that creates problems for cross-loadings or confer similar weights to scores that, in actuality, have none. Besides, another perhaps more important drawback is the possibility that these methods can create factors correlated between them.

Refined methods, according to De Stefano *et al*. (2009), may be applied when principal component and common factor extraction methods are used with EFA, so they are even more focused on an exploratory approach. After considering these methods, more relate to an exploratory analysis and are used mainly in psychological studies I prefer to opt for retaining the original factors. Apart from the uniform recoding of all the variables carried out, it is also quite clear from the univariate counts that most of my variables show an important percentage of frequencies on null or low items, due to the topic in itself (for example, despite of the sum up of variables, membership variables record many 0 from many individuals). A further rescaling or averaging would have reduced further the variability of my factors score.

Continuing further with the analysis of the results, the literature suggests that for categorical analysis further interpretation of the results has to be done for the standardized results. This is due to the set of the variances of the factors to the standardized value of one. In this way, the interpretation of factor-loading estimates is based on the squared standardized factor loadings (Byrne, 2012). According to Schreiber (2006), interpretation of coefficients and residuals are often underestimated in most parts of studies where only fit indices are reported. The following Tables report the standardized model results with the parameter estimates by factor with P-values confirming strongly their significance (Table 2.5) and the estimates about *R-Square* (Table 2.6).

Table 2.5: Standardized Model Results – 2001

	Estimates	**S.E.**	**Est./S.E.**	**P-Value**
FACTOR 1 by				
MEMBER1	0.334	0.013	25.492	0.000
ACTIVE1	0.388	0.018	22.125	0.000
MEMBER3	0.139	0.017	8.33	0.000
ACTIVE3	0.115	0.018	6.377	0.000
GOVERN1	0.318	0.013	23.687	0.000
GOVERN2	0.288	0.014	20.133	0.000
GOVERN3	0.162	0.016	10.35	0.000
VOTE1	0.686	0.011	60.212	0.000
VOTE2	0.618	0.015	41.445	0.000
VOTE3	0.659	0.011	59.116	0.000
VOTE4	0.767	0.013	61.257	0.000
FACTOR 2 by				
MEMBER2	0.345	0.014	24.132	0.000
MEMBER4	0.972	0.003	279.148	0.000
MEMBER5	0.083	0.016	5.318	0.000
RELIG	0.909	0.005	194.049	0.000
ACTIVE2	0.376	0.014	27.178	0.000
ACTIVE4	0.969	0.004	275.4	0.000
ACTIVE5	0.092	0.015	5.953	0.000
FACTOR 3 by				
NEIGH	0.155	0.015	10.539	0.000
MEET	0.247	0.017	14.914	0.000
CARENR	0.065	0.018	3.639	0.000
HELP1EXT	0.729	0.009	79.627	0.000
HELP2EXT	0.467	0.012	39.031	0.000
HELP3EXT	0.497	0.012	40.072	0.000
HELP1NET	0.913	0.006	144.159	0.000
HELP2NET	0.854	0.008	113.754	0.000
HELP3NET	0.833	0.008	102.463	0.000
HELP4NET	0.802	0.008	96.437	0.000
HELP5NET	0.902	0.006	146.729	0.000
LIKENBRD	0.213	0.02	10.938	0.000
MOVE	0.085	0.014	6.038	0.000
Standardized results for latent variables				
FACTOR 1	with			

FACTOR 2	0.282	0.014	20.706	0.000
FACTOR 3	0.058	0.014	4.068	0.000
FACTOR 2	with			
FACTOR 3	0.111	0.014	7.688	0.000

Table 2.6: R-Square Estimates -2001

Variable	Estimate	S.E.	Est./S.E.	P-Value	Residual variance
MEMBER1	0.112	0.009	12.746	0.000	0.888
MEMBER2	0.119	0.010	12.066	0.000	0.881
MEMBER3	0.019	0.005	4.165	0.000	0.981
MEMBER4	0.945	0.007	139.574	0.000	0.055
MEMBER5	0.007	0.003	2.659	0.008	0.993
ACTIVE1	0.150	0.014	11.062	0.000	0.85
ACTIVE2	0.141	0.010	13.589	0.000	0.859
ACTIVE3	0.013	0.004	3.188	0.001	0.987
ACTIVE4	0.940	0.007	137.7	0.000	0.06
ACTIVE5	0.008	0.003	2.976	0.003	0.992
RELIG	0.827	0.009	97.025	0.000	0.173
GOVERN1	0.101	0.009	11.844	0.000	0.899
GOVERN2	0.083	0.008	10.066	0.000	0.917
GOVERN3	0.026	0.005	5.175	0.000	0.974
VOTE1	0.471	0.016	30.106	0.000	0.529
VOTE2	0.382	0.018	20.723	0.000	0.618
VOTE3	0.435	0.015	29.558	0.000	0.565
VOTE4	0.588	0.019	30.628	0.000	0.412
NEIGH	0.024	0.005	5.269	0.000	0.976
MEET	0.061	0.008	7.457	0.000	0.939
CARENR	0.004	0.002	1.819	0.069	0.996
HELP1EXT	0.532	0.013	39.813	0.000	0.468
HELP2EXT	0.218	0.011	19.516	0.000	0.782
HELP3EXT	0.247	0.012	20.036	0.000	0.753
HELP1NET	0.834	0.012	72.079	0.000	0.166
HELP2NET	0.730	0.013	56.877	0.000	0.27
HELP3NET	0.693	0.014	51.232	0.000	0.307
HELP4NET	0.643	0.013	48.218	0.000	0.357
HELP5NET	0.813	0.011	73.365	0.000	0.187
LIKENBRD	0.045	0.008	5.469	0.000	0.955
MOVE	0.007	0.002	3.019	0.003	0.993

According to Byrne (2012), to interpret the factors and the loadings the *R-square* estimates must be explored. In the case of continuous variables, we should look at the standardized parameters in order to see how much the proportion of variance in the *observed* variables is explained by the underlying factor. For categorical variables, the interpretation is based on the squared standardized factor loadings. All the variables are significant with a P-value of 0.000, obtained immediately from the first attempts at fitting the model. This can confirm that the theoretical hypothesis for the model is correct.

Consequently, for example, *member1* for Factor 1 has a parameter estimate equal to 0.334. Its squared value is equal to the rounded off value of 0.112 in Table 2.6. This means that the 11 percent of the variance in the underlying latent aspect of *member1* can be explained by the construct of Factor 1. At the same time, if we subtract the squared standardized loading (0.112) from 1.00, we obtain a value of 0.888, which is the residual variance for the same variables (last column on the right of Table 2.6).

Starting from Factor 1, Table 2.5 strongly suggests an important proportion of the variances of the variables representing voting such as first group (range from 0.61 to 0.76) and then, quite uniformly, on the membership's variables (all of them about 0.30 on average apart for *member3* and *active3*). Despite this low loading (and a corresponding low R-square, as we can see from Table 2.6) when I tried an alternative model removing them from Factor 1, I had problem of linear dependences between other variables. The significant p-values, both at parameter estimates level and at R-square level, and the strong theoretical justification suggested that they needed to be kept in the factor. In a symmetric way, from Table 2.6, we can see that Factor 1 explains on average almost half of the variance for the underlying variables representing voting behaviour (*vote1n*, *vote2n*, *vote3n* and *vote4n*).

The same argumentation can be retained for Factor 2, about informal membership. *Member5* and *active5* have a low loading estimate (around 0.10) while the other variables show higher and, in some cases, important estimates (about 0.9 for some variables and 0.3 for others). In this case, even more so than the previous one, the variance explained by these variables is really low (but significant) (Table 2.5 and 2.6). But again, also in this case, the removal of these variables showed the rising of linearity among other variables and, consequently, an important decrease in the overall fit indices. Factor 2, finally, fits the variables that appear to show the highest importance in all the models: *member4* (parameter estimate at 0.972 and R-square at 0.945).

Finally, for the third factor on informal networks, *carenr* is the variable that exhibits a low loading and, consequently, variance. Its significance and the fact that, as described previously, it is the only variable for informal caring fitting the model, suggested that I should keep it in the model. Lower loadings and R-square but with significant P-values are reported for the four variables representing neighbourhood relations (*likenbrd*, *move*, *neigh* and *meet*) confirming their better fit in this last factor concerning informal networks.

Finally, Table 2.5 reports also the P-values (all significant at 0.001) for the correlation between the latent factors previously described in paragraph 3.3.1.

In reference to the general model fit (Hu and Bentler, 1995; Di Stefano and Hess, 2005), literature provides different indices that can be divided in two bigger classes: absolute and relative goodness of fit indices.

Absolute fit indices determine how a priori model reproduces the data. Among them, the main used are the Chi-Square Value Test and the Root Mean Square Error of Approximation (RMSEA). The first one indicates the difference between observed and expected covariance matrices. Therefore, P-Value closer to zero indicate a better fit. RMSEA is the index that better deals with sample size issues and it analyses the discrepancy between the hypothesized model and the population covariance matrix. In addition to this case, smaller values indicate a better goodness of fit, with values lower than 0.6 indicating acceptable models.

The second group of indices is the relative fit indices. Among them, the more used are the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) called also Non-Normed Fit Index (NNFI). Generally speaking, the relative fit indices compare the chi-square of the hypothesized model to a baseline model or null model, where usually in the null model all the variables are uncorrelated (with a large chi-square and, therefore, a poor fit).

The CFI analyse the discrepancy between the data and the hypothesized model adjusting also for sample size issues. With a range between zero and one, larger values greater than 0.9 indicate acceptable models. The TLI analyses the discrepancy between the chi-square of the hypothesized model and the null model avoiding negative biases of the Normed Fit Index (NFI).

With regard to 2001 model, the most common indices show goodness of fit, all of them under the suggested cut-off values:

1. **Chi-Square Value Test**: 5307.950 with 397 degrees of freedom and a P-Value equal to 0.000, the value suggested for a good goodness of fit;
2. **RMSEA**: 0.028 with 1.0 probability of RMSEA to be under the critical value of 0.05. Hu and Bentler (1999) suggested a cut off value of <0.06;
3. **CFI**: 0.98, >0.95 (Hu and Bentler 1999);
4. **TLI** or **NNFI**: 0.976, >0.95 (Hu and Bentler 1999).

Schreiber *et al*. (2006) underline that RMSEA, CFI and TLI are the preferable fit indices to consider for one-time analysis.

2.5 Model for 2011: strategy, analysis and results

In this paragraph, I am going to explain the procedure I followed for 2011. Keeping in mind the description done in paragraph 2.2 about the CFA methodology and the UKHLS dataset used in this case, we can consider valid also for this model the variables and measurements of SC dimensions presented in the literature and in accordance with the empirical studies listed in the previous paragraphs. Many of them will be similar to the 2001 case but other different dimensions will be tested. All the variables used, with labels and recoded names are reported in Appendix A (Table A1 and A3). As for 2001, I build my theoretical latent factors based on several variables available from survey and I confirm them with the current CFA, applying the general strategy summarized in Figure 2.1.

Starting from the membership and active membership dimension UKHLS survey present the same variables of BHPS. Therefore, they have been summarized and recoding in the same way as for 2001. We can see from the Table A5 in Appendix A, the univariate proportions and the counts for these categorical variables.

Focusing on the religious aspect, it has been possible recoding and testing more variables than the previous case: *relig*, *att_relig* and *imp_relig*. While the *relig* is the same as for 2001 measuring the belonging to a religion, the other two variables are about:

- *att_relig*: frequency of attendance at religious services (recoded like an ordinal variable depending on the frequency of attendance),
- *imp_relig*: if religion makes a difference to life.

With regard to the second dimension on citizenship and voting behaviours, UKHLS presents a wide range of variables about this topic. After the recoding, I even obtain fourteen variables on *generalized trust* and *institutional trust* (Ackomak and ter Weel 2005, 2009; Dakhli and De Clercq 2004): voting behaviours and expectations, personal skills and level of participation in political life, statements on government actions and principles and a uniform number of sources of information on political news (see Table A3 in the Appendix A).

The discussion about the third dimension is more complex for 2011 case. Given the rich availability of data, I recoded a wide set of variables representing:

- general health condition with variables about illness, sleeping condition, sense of happiness, confidence, levels of satisfaction for different aspects of life,

- providing care to someone else (both in the household and outside);
- the amount of caring (hours per week);
- frequencies of contacts, types of contacts and types of help - given or received - by typologies of family members;
- two variables representing informal networks not related to the family dimension: two variables, *closefr* (recoded like an ordinal variable with respect to the number of close friends listed by the interviewee) and *gout* (frequency of going out with friends)

Some of them are attributable to definitions of health and caring and informal networks, both familial and not, as previously described while the other variables about personal health conditions have been used more in recent years in studies about SC and its effect on individual mental health. Other some variables, according to the studies previously described, can be considered as measurements of kinds of actions not as proper care provision or healthcare, but at least linked to the same amount and type of SC required for an individual to be willing to care for other persons (especially if belonging to the same nuclear family).

Also for the fourth dimension, a richer set of variables was available. In first instance, I recoded and tested twelve variables for several aspects of local mutual trust – derived from a local result of a more generalized and horizontal trust between individuals (Ackomak and ter Weel 2005, 2009; Dakhli and De Clercq 2004).

On the crime topic, differently from BHPS, UKHLS survey contains several variables about feeling unsafe, avoiding actions, being insulted or threatened, being physically attacked in different public places and for different reasons (ethnicity, gender and so on). There are also questions on reasons for not going out and socialising. However, the structure of the survey and the necessary recoding allowed me to test only two variables: *fearcrime* and *safedark*, recoded as ordinal variables representing the level of fear and the feeling of being safe walking alone at night.

Regarding the last aspect of use of social media, it is possible to notice again a different and richer availability of data. This difference in availability can also reflect a more general and theoretical aspect: the development and global diffusion of the use of social media as a recent event. After the recoding, I obtained four variables:

- *socnet_chat*: an ordinal variable representing the number of hours spent chatting per week on social websites;
- *mobile*: having or not a personal mobile;
- *internetfr*: how often internet is used;
- *hrstv*: number of hours in front of tv per day.

Finally, a smaller group of variables representing the time spent at the current address after the last move was introduced and relevant variables created. In addition, in this case I have a wider availability of variables. Apart for the variable about liking or not the neighbours, four dummy variables about the last moving were created for different time lapse: moved in the last three years, between three and five years ago, between five and ten years and more than 10 years ago.

Through these examples, it is quite clear that the amount of time spent in a specific place is an important and fundamental asset to build and maintain SC with people.

2.5.1 Strategy for building up the model

As for 2001 model and according to De Stefano *et al*. (2009), before describing the internal structure of the factors with regard to their loadings, I will describe the several steps done before the fitting of the final model as presented in Figure 2.7.

My first main hypothesis was a model mirroring the four dimensions just descried: one factor about membership (both passive and active) and religious aspects, a second factor about citizenship and voting behaviours, a third factor about health and caring and relations with close friends and a last factor about relations with neighbours, crime, use of social media and mobility.

At the first test, the model did not converge and from MIs was quite clear that the groups of variables representing personal satisfaction and family network and contacts did not result significant in the model.

The richness of information of the UKHLS allows me to consider a model for 2011 with one or more factors corresponding to other dimensions that are deeply related to SC. The addition of these variables and corresponding latent factors prevented a proper fit of the model and, in some case, even its convergence. This can be due to high correlation between the variables used[7] and the small number of observations available even after the recoding and so on. The first big dimension is related to the satisfaction towards several aspects of personal life and temporary personal difficulties also experienced at a physical level (loss of sleep, inability to overcome

[7] This aspect is frequently due to the structure of the original questionnaire where, following an initial filter question, in the case of a positive answer the individual is asked to complete a sub-section of questions on singular aspects, creating in this way variables almost multicollinear.

difficulties, being under stress and so on). We saw previously how, especially – but not exclusively – in psychological studies, individual levels of personal satisfaction with job or relationships seem to be positively related to a higher level of SC. Consequently, all the variables representing temporary difficulties were supposed to be strongly related and loaded on a proper factor. Without focusing further on this interesting topic, as an example Brehm and Rahn (1997) find evidence of a correlation between classical dimensions of SC (like interpersonal trust and civic engagement) and psychological involvement with the community, cognitive skills and general satisfaction (linked also to the economic personal condition).

A second significant dimension that was hypothesized as related to health and caring is composed by all the variables for the kinds of contacts, supports and help provided internally to the family network: to and from the mother, father or children. The integration of the factor representing health and caring or the creation of a proper factor with only these variables has made the model non-converging. This is probably due to variances or residual variances approaching 0.

After the removal of the factors loading on these variables (on personal and psychological dimensions and health and caring the family network), the model immediately fitted with RMSEA Index around 0.080 and CFI/TLI around 0.70. Therefore, the second attempt was done on a model with three factors confirmed respect to the hypothesis and the theory, and where the factor about health and caring and informal networks was removed. In particular, the third factor resulted loading onto composite and different variables. I make the hypothesis, confirmed by the results, that crime variables would load on the same factor of a neighbourhood's relations and networks, according to the literature. So *fearcrime* and *safedark*, recoded as ordinal variables representing the level of fear and the feeling of being safe walking alone at night, perfectly fitted loading on the last third factor of 2011 model.

Other hypotheses for this third factor were done respect to the use of social media. If in 2001 it was enough to question individuals on the use of a mobile phone, for 2011 we have different variables available about the use of social media.
My hypothesis, also according to the literature, is to load these variables on the last factor corresponding to informal networks. Indeed, this hypothesis has been confirmed by the fit of the model. Accordingly, *socnet_chat* (an ordinal variable representing the number of hours spent chatting per week on social websites) and *mobile* result significantly linked to the other variables relating to neighbourhood relations, friends' networks and crime. As we can see from the results set out in the following sections, the sign of the relation is negative: many authors state that even if we can consider social media as an alternative means of interaction, their intensive use subtracts real time spent outside with people[8].

Not confirmed by the analysis was my hypothesis about mobility's variables (recoded according to different lapses of time since the moving up to more than 10 years). They were supposed to load on the same factor of neighbourhood's relations and Bridging

[8] As we can see from the Mplus syntax in the Supplementary Material section, other variables about use of internet and TV but they were removed according to Indices in order to make the model converge.

SC but, as in the previous case of variables for family networks, the high multicollinearity between them or, most probably, again variances or residual variances approaching 0 made the model non-converging.

Focusing on a more general level, I have already anticipated how in this case further specifications for the correlations of residuals between variables and cross-loadings of factors on variables of other variables have been kept only if necessary to have a more generalizable model. As for 2001, correlating the residuals of the variables of membership with their reciprocal active counterpart helped the fitting of the model in a crucial way. I then correlated some variables in the second factor about closeness and support to a party, interest and preparation about politics and trust towards Government variables. The last specification was about the third factor and the variables about fear of crime, going out and use of social media.

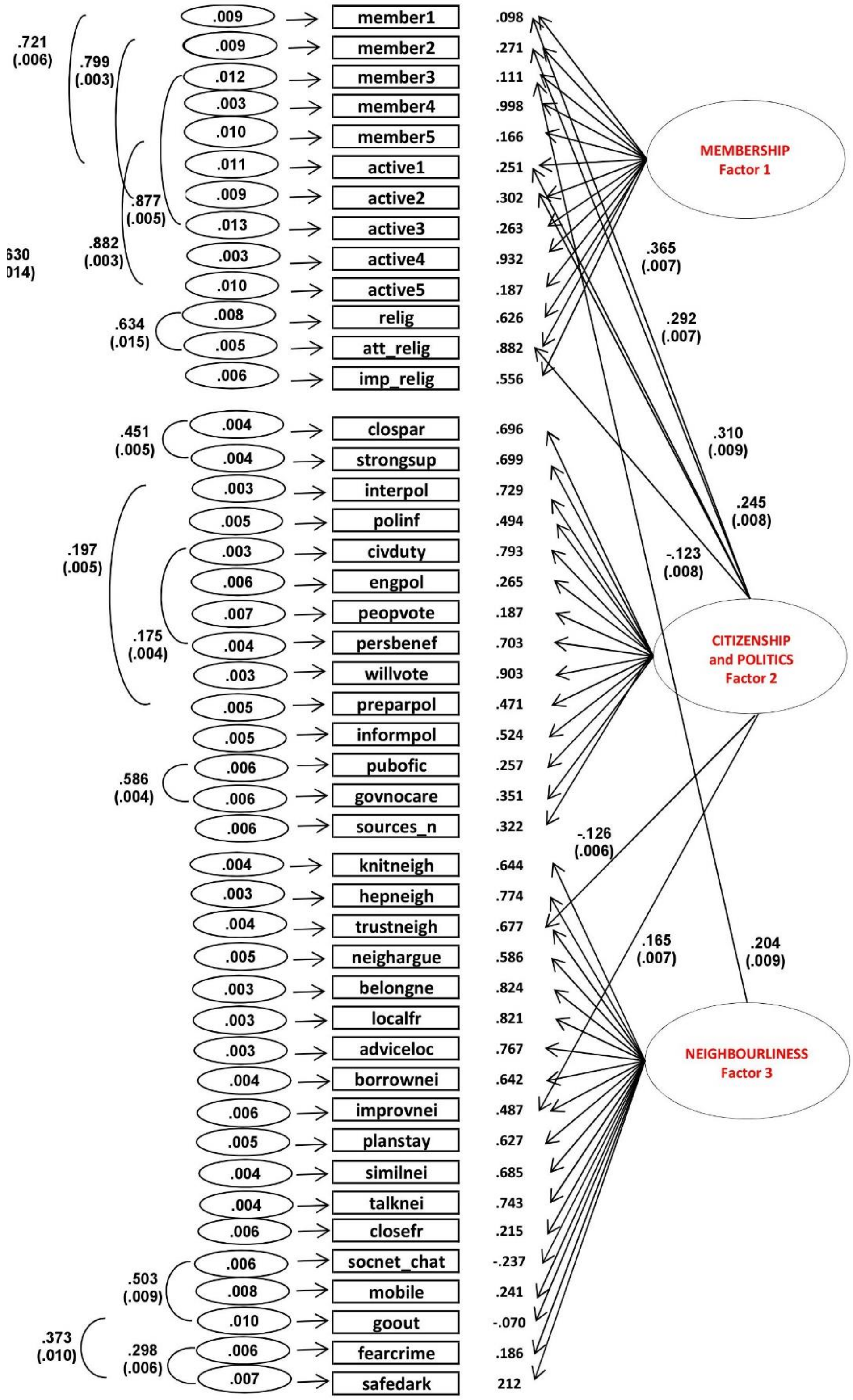
Another interesting point arose from the modification indices: is the main cross-loading of Factor 2 about citizenship, politics and voting on the variables about membership and active membership in Factor 1 that are about these types of organizations: *member* and *active* 1 and 2 (parties, trade unions, professional organizations and so on), confirming a strong correlation. The third factor about neighbourliness shows only a necessary cross-loading on variable about membership to tenants and resident groups from Factor 1.
All these specifications of covariances and cross-loadings have been added to reach highest values of goodness of fit. How Mplus syntax shows in Supplementary Material section and according to MIs, (BY and WITH statements).

For 2011, differently from 2001, defining cross-loadings helped the statistical improvement of the model. From Fig. 2.7 it is possible to see how mainly the second factor has different cross-loadings

Following on all the strategy, the final model proposed for 2011 has three factors named according to the characteristics of the variables on which they load:

- Factor 1: *Membership*;
- Factor 2: *Citizenship and politics*;
- Factor 3: *Neighbourliness.*

Figure 2.7: Final model for 2011



Values shown between variables and paths from Factors represent standardized β coefficients. Other values are error terms. Significance at 0.001

Even if it is not possible to compare in a proper way the two models because of the differences between the two samples (it is not longitudinal study) and in variables used, we can only touch on a comparison. There are differences between 2001 and 2011 for membership, religious and political variables. Membership to formal and informal organizations are separated in 2001 (no matter the active characterization) while in 2011 they underlie the same factor. Political variables in 2001 are based in the same factor of formal organization membership (active or non-active) whilst in 2011 they are better explained by proper a factor apart. Only the religious aspect confirms a similarity between the two years: the variables underlie the same factor of membership[9]. These variables fit better when they are linked with membership, because of their personal and intimate aspect and at the same time the stronger bridging aspect.

Health and caring factor has not been confirmed at all for 2011 apart for one variable for care provided to non-resident individuals (not belonging to the household). Even more, the removal of all these variables allow the full model of three factors to converge for the first time. This is also probably due to the number of observations available and the high number of frequencies on 0 values that create a sizeable problem of linearity between all variables considered.

[9] As will be described later, even if between 2001 and 2011 we can notice this different loading of the latent factors on variables. We will see also that a complex cross-loadings of factors on variables belonging to other factors and correlation between residuals of original variables has been necessary for an important improvement in the goodness of fit of the indices. This fact confirms further this strong link between them.

2.5.2 Results: analysis of three factors characteristics and main indices of goodness of fit

Figure 2.7 shows the complexity of the model: forty-five variances and fourteen residual covariances and seven cross-loadings for a total of sixty-six parameters to estimate. According to Schreiber *et al*. (2006) and Pohlmann (2004) for one sample analysis, the ratio between the number of observations and the number of parameters to be estimated should be higher than ten. The ratio is 753.62, higher than the threshold value indicated. As for the model for 2001, a pre-analysis includes discussion on sample statistics, software and estimator used. Table A5 in Appendix A shows the univariate proportions and counts for the categorical variables. The recoding presents the same uniformity for all the variables like in 2001: increasing levels of SC for each variable are parallel to the increase of the values of the items.

If we look at the following Tables (Table 2.7 and 2.8), we can examine correlation and covariances between factors.

Table 2.7: Correlation matrix for the latent variables – 2011

	Factor 1	Factor 2	Factor 3
Factor 1	1.000		
Factor 2	0.288	1.000	
Factor 3	0.197	0.203	1.000

Table 2.8: Covariance matrix for the latent variables – 2011

	Factor 1	Factor 2	Factor 3
Factor 1	0.010		
Factor 2	0.020	0.484	
Factor 3	0.012	0.091	0.415

As expected, the correlation is positive between all factors. The highest correlation is between Factor 1 and Factor 2 and it can be easily interpreted as the stated relationship, at theoretical level, between attitudes to being a member (and active member) and level of interest in politics and sense of citizenship. Factor 3, loading on all the variables relating to neighbourhood relations, informal networks and fear of crime, is correlated with both the factors with really close intensity (0.197 with Factor 1 and 0.203 with Factor 2). This aspect can be explained by the fact that, as explained, informal networks, more formal relations with neighbours and sense of being safe are reciprocally influenced by personal attitudes and behaviours regarding the wider levels of society (such as associations, organisations and institutions).

The covariances between factors are quite low. This means that Factor 2 and Factor 3 have high variance while Factor 1 seems less dispersed. Standard deviation in Table 2.7 and box plots in Figure 2.11 strongly confirm the lowest dispersion of data for Factor 1. The higher covariance, on the other hand, is between Factor 2 and Factor 3 (0.091) showing again a strong theoretical link.

Table 2.9: Descriptive statistics for Factor1, Factor 2 and Factor 3 – 2011

	Obs.	**Mean**	**Standard Deviation**	**Min.**	**Max.**
Factor 1 *Membership*	49,739	0.0073	0.07	-0.146	0.297
Factor 2 *Citizenship and politics*	49,739	0.0180	0.62	-2.129	2.724
Factor 3 *Neighbourliness*	49,739	-0.0143	0.53	-2.02	1.283

Figure 2.8: Frequencies for Membership factor (Factor 1) - normal distribution, 2011

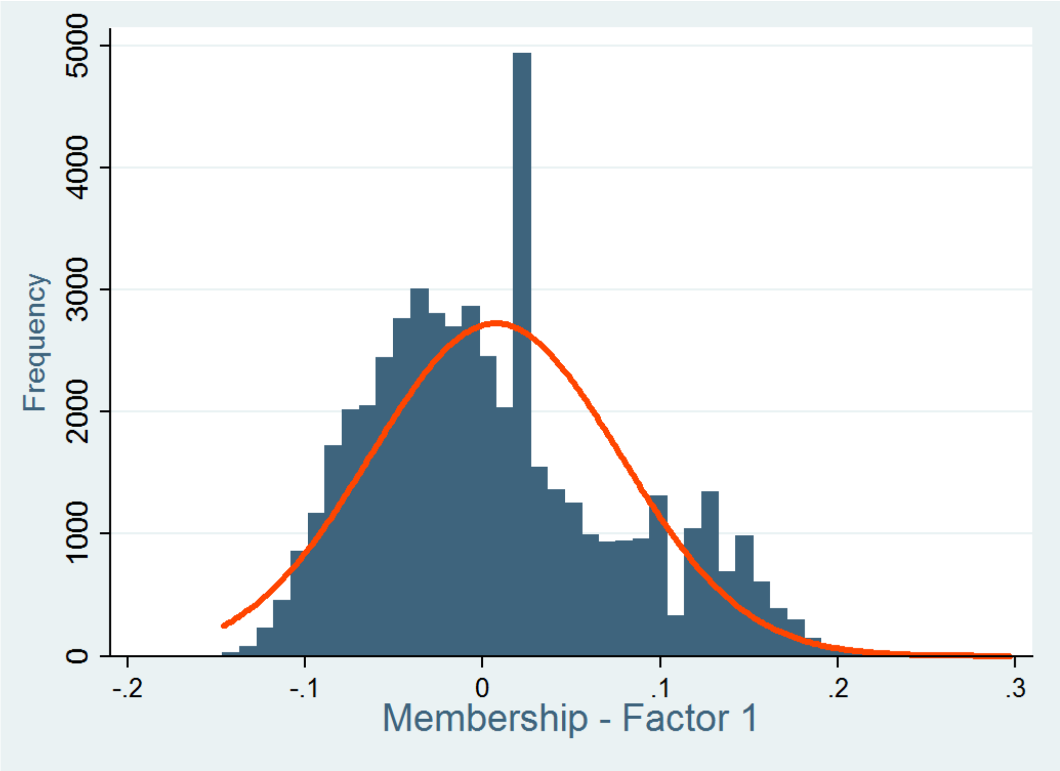


Figure 2.9: Frequencies for Citizenship and politics factor (Factor 2) - normal distribution, 2011

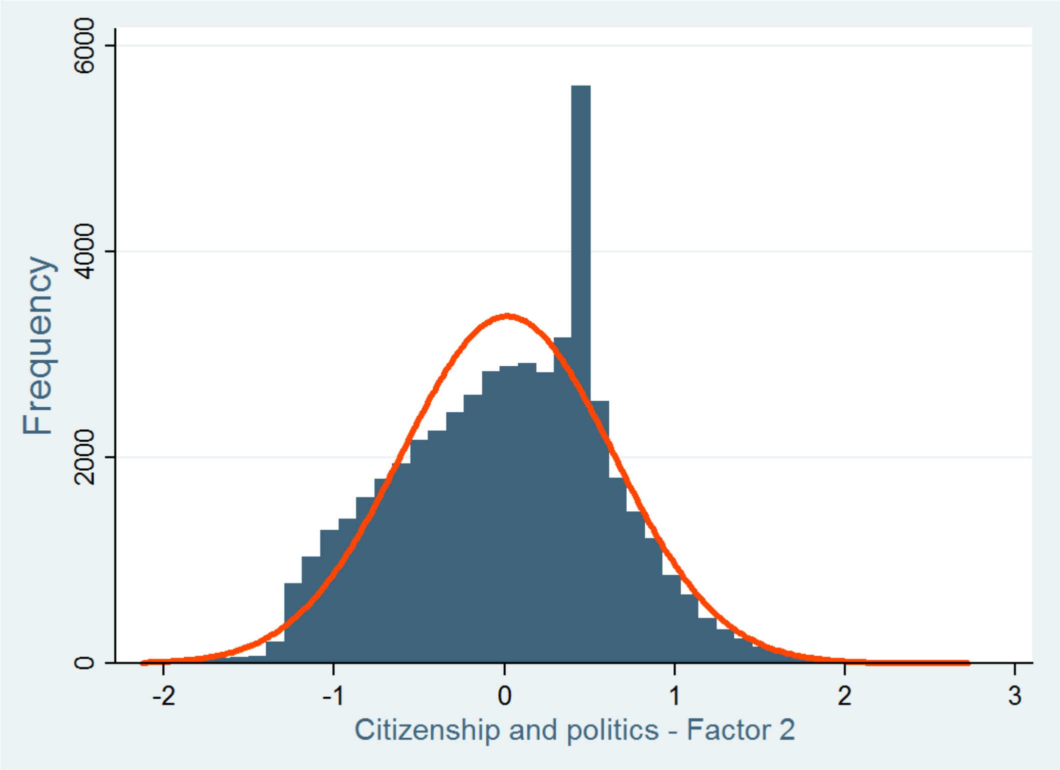


Figure 2.10: Frequencies for Neighbourliness factor (Factor 3) - normal distribution, 2011

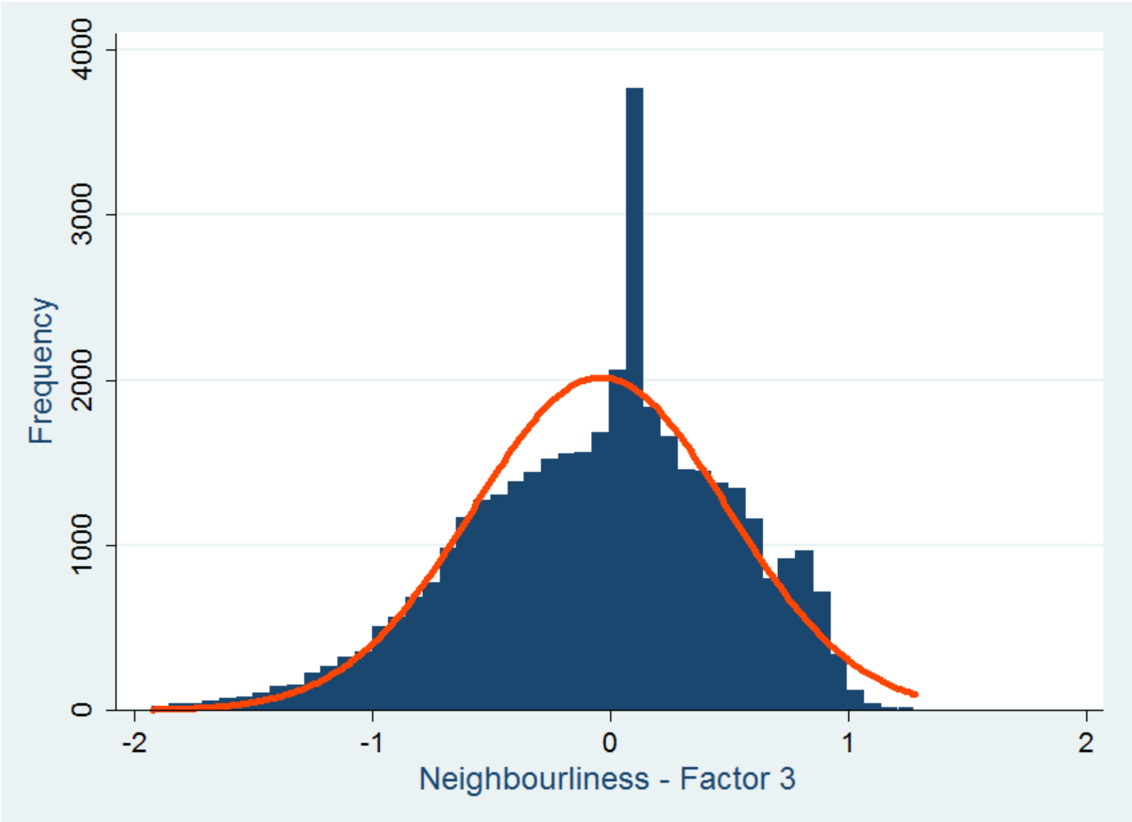
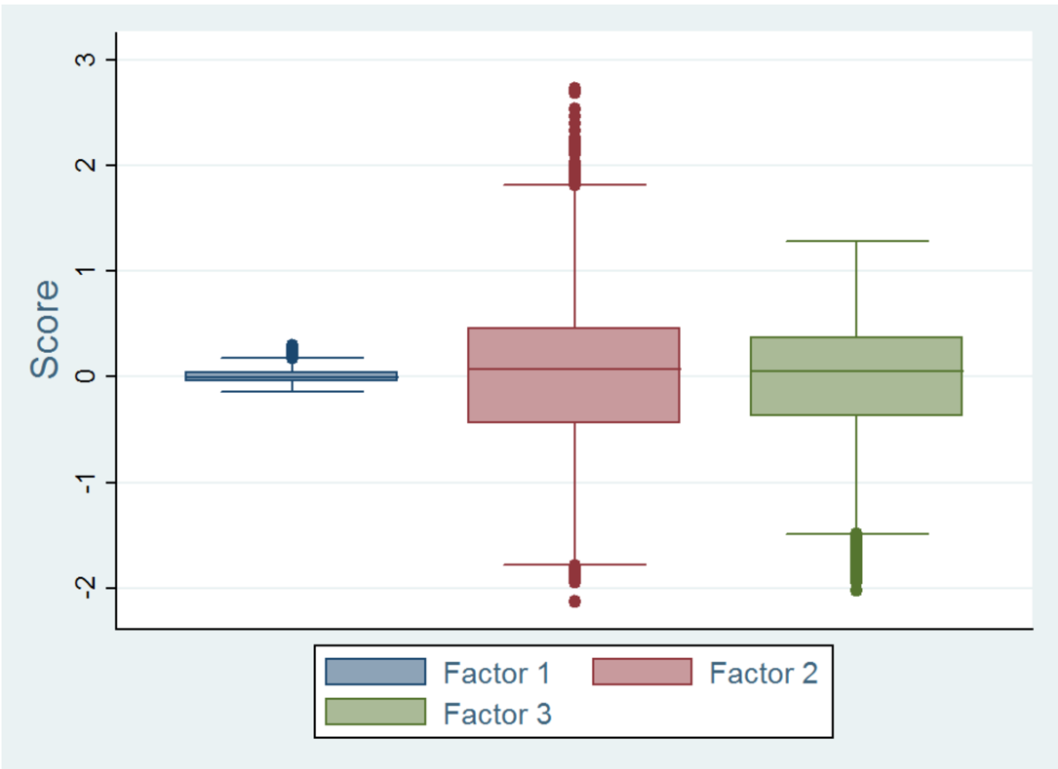


Figure 2.11: Box plots for the Factor 1, Factor 2 and Factor 3 – 2011



Examining carefully both graphs and box plots, we can notice that latent factors in 2011 are close to a normal distribution even though the range of the factors' scores is wider. Even if outliers are present and expected, we can also state that skewness and anomalies in the factors are minor. This can depend on a large sample (in this case the number of observations is more than double) and lower dispersion of data. As previously described, this also means that an individual's rating of importance for that attribute is close to the average of the sample (De Stefano, 2009). As in the previous case, I opted for retaining the original factors.

The structure of the model can be also confirmed looking at the correlation matrix (Table A7 in Appendix A).

Higher correlations are present between variables of membership and their active counterpart in an expected way (*member1* and *active1*, *member3* and *active3*, *member4* and *active4*, *member5* and *active5*). Already explained from a theoretical perspective, in this case the variables about religious aspects also seem to be highly correlated with some of the membership variables with which they load on the same factor. Empirically, this justifies the strong structure of the first factor.

The other large group of variables highly correlated belongs to the second factor about citizenship and politics. *Willvote*, the variable about voting in the next election is naturally highly correlated with variables representing support to party and attitudes to voting (*clospar*, *strongsup*, *interpol,cividuty*). The variables representing supporting a party – *clospa* - and degree of closeness – *strongsup* - are naturally highly correlated (0.937) but the elimination of one of the two creates problems to the fitting of the models.

The last notable correlation is related to the third factor. *Belongne*, the variable representing how much the individual feels they belong to the neighbourhood where he/she lives, is correlated with other several variables representing relations between them: accepting advice from them (*adviceloc*), number of local friends (*localfr*), similar feelings to neighbours (*simileni*) and talking to them (*talknei*) with a coefficient slightly above 0.6.

Focusing on goodness of fit of the model and a more general view of it, we can see how the 2011 model required much less residual covariances and cross-loadings in relation to the previous model. This aspect is extremely important as it implies two main considerations: the model in this way is less current data dependent and the factors explain more variance because it has been not reduced by adding correlations between the residuals (Geiser, 2013). The first consideration is especially important. The model is more generalizable and less dependent on this specific sample and its characteristics. This aspect is fundamental also for the following Multilevel Modelling framework. Deeply linked with this result is the choice to not pursue the perfect fit according to the highest thresholds of the fitting indices, as I am going to describe shortly.

As in the previous case, the interpretation of the factors extracted and their loadings can be achieved on standardized results for categorical variables. All the variables in this case were also significant immediately with a P-value of 0.000, showing a right theoretical model was hypothesized.

Table 2.10: Standardized Model Results – 2011

	Estimate	S.E.	Est./S.E.	P-Value
Factor 1 by				
MEMBER1	0.098	0.009	10.842	0.000
MEMBER2	0.271	0.009	31.131	0.000
MEMBER3	0.111	0.012	9.275	0.000
MEMBER4	0.998	0.003	346.892	0.000
MEMBER5	0.166	0.010	17.047	0.000
ACTIVE1	0.251	0.011	23.313	0.000
ACTIVE2	0.302	0.009	35.036	0.000
ACTIVE3	0.263	0.013	20.787	0.000
ACTIVE4	0.932	0.003	287.864	0.000
ACTIVE5	0.187	0.010	19.246	0.000
RELIG	0.626	0.008	79.295	0.000
ATT_RELIG	0.882	0.005	174.962	0.000
IMP_RELIG	0.556	0.006	89.922	0.000
Factor 2 by				
CLOSPAR	0.696	0.004	177.892	0.000
STRONGSUP	0.699	0.004	187.250	0.000
INTERPOL	0.729	0.003	215.998	0.000
POLINF	0.494	0.005	100.416	0.000
CIVDUTY	0.793	0.003	235.840	0.000
ENGPOL	0.265	0.006	44.080	0.000
PEOPVOTE	0.187	0.007	27.616	0.000
PERSBENEF	0.703	0.004	188.135	0.000
WILLVOTE	0.903	0.003	309.475	0.000
PREPARPOL	0.471	0.005	89.581	0.000
INFORMPOL	0.524	0.005	112.028	0.000
PUBOFIC	0.257	0.006	43.734	0.000
GOVNOCARE	0.351	0.006	62.405	0.000
SOURCES_N	0.322	0.006	56.917	0.000
Factor 3 by				
KNITNEIGH	0.644	0.004	164.080	0.000
HELPNEIGH	0.774	0.003	221.700	0.000
TRUSTNEIGH	0.677	0.004	168.107	0.000
NEIGHARGUE	0.586	0.005	118.308	0.000
BELONGNE	0.824	0.003	293.070	0.000
LOCALFR	0.821	0.003	302.826	0.000
ADVICELOC	0.767	0.003	246.775	0.000
BORROWNEI	0.642	0.004	158.209	0.000

IMPROVNEI	0.487	0.006	86.626	0.000
PLANSTAY	0.627	0.005	134.616	0.000
SIMILNEI	0.685	0.004	174.185	0.000
TALKNEI	0.743	0.004	202.702	0.000
CLOSEFR	0.215	0.006	36.402	0.000
SOCNET_CHA	-0.237	0.006	-39.552	0.000
MOBILE	0.214	0.008	26.565	0.000
GOOUT	-0.070	0.010	-7.104	0.000
FEARCRIME	0.186	0.006	30.343	0.000
SAFEDARK	0.212	0.007	31.399	0.000
Standardized results for latent variables				
Factor 1	with			
Factor 2	0.288	0.007	41.354	0.000
Factor 3	0.197	0.007	28.628	0.000
Factor 2	with			
Factor 3	0.203	0.006	34.793	0.000

Table 2.11: R-Square Estimates -2011

Variable	Estimate	S.E.	Est./S.E.	P-Value	Residual variance
MEMBER1	0.164	0.005	31.038	0.000	0.836
MEMBER2	0.204	0.006	34.015	0.000	0.796
MEMBER3	0.063	0.005	13.496	0.000	0.937
MEMBER4	0.995	0.006	173.446	0.000	0.005
MEMBER5	0.028	0.003	8.523	0.000	0.972
ACTIVE1	0.204	0.008	26.869	0.000	0.796
ACTIVE2	0.193	0.006	31.992	0.000	0.807
ACTIVE3	0.069	0.007	10.394	0.000	0.931
ACTIVE4	0.869	0.006	143.932	0.000	0.131
ACTIVE5	0.035	0.004	9.623	0.000	0.965
RELIG	0.392	0.010	39.647	0.000	0.608
ATT_RELI	0.731	0.008	97.060	0.000	0.269
IMP_RELI	0.309	0.007	44.961	0.000	0.691
CLOSPAR	0.484	0.005	88.946	0.000	0.516
STRONGSU	0.488	0.005	93.625	0.000	0.512
INTERPOL	0.532	0.005	107.999	0.000	0.468
POLINF	0.244	0.005	50.208	0.000	0.756
CIVDUTY	0.629	0.005	117.920	0.000	0.371
ENGPOL	0.070	0.003	22.040	0.000	0.930
PEOPVOTE	0.035	0.003	13.808	0.000	0.965
PERSBENE	0.494	0.005	94.068	0.000	0.506
WILLVOTE	0.815	0.005	154.738	0.000	0.185
PREPARPO	0.221	0.005	44.790	0.000	0.779
INFORMPO	0.275	0.005	56.014	0.000	0.725

PUBOFIC	0.066	0.003	21.867	0.000	0.934
GOVNOCAR	0.123	0.004	31.203	0.000	0.877
SOURCES_	0.104	0.004	28.459	0.000	0.896
KNITNEIG	0.415	0.005	82.040	0.000	0.585
HELPNEIG	0.599	0.005	110.850	0.000	0.401
TRUSTNEI	0.510	0.005	97.475	0.000	0.490
NEIGHARG	0.344	0.006	59.154	0.000	0.656
BELONGNE	0.679	0.005	146.535	0.000	0.321
LOCALFR	0.674	0.004	151.413	0.000	0.326
ADVICELO	0.589	0.005	123.387	0.000	0.411
BORROWNE	0.412	0.005	79.105	0.000	0.588
IMPROVNE	0.297	0.006	50.916	0.000	0.703
PLANSTAY	0.393	0.006	67.308	0.000	0.607
SIMILNEI	0.469	0.005	87.093	0.000	0.531
TALKNEI	0.552	0.005	101.351	0.000	0.448
CLOSEFR	0.046	0.003	18.201	0.000	0.954
SOCNET_C	0.056	0.003	19.776	0.000	0.944
MOBILE	0.046	0.003	13.282	0.000	0.954
GOOUT	0.005	0.001	3.552	0.000	0.995
FEARCRIM	0.035	0.002	15.171	0.000	0.965
SAFEDARK	0.045	0.003	15.700	0.000	0.955

As we can see from Table 2.10 and 2.11, Factor 1, the factor representing membership, active membership and religious beliefs, loads strongly on variable *member4* (0.998) and its active counterpart *active4* (0.932). These variables address religious membership and together with the other variables for religious beliefs that also have high loadings (*relig*, *att relig*, *imp_relig*) may show that this dimension is very important in relation to the other memberships. This result is similar to 2001 where also, in that case, *member4* and *active4* showed a high proportion of variance perfectly explained by their underlying factor (Factor 2). Confirmation arises also from Table 2.11 where 99.5 percent for *member4*, 86.9 percent for *active4*, 73.1 percent for *att_relig* of variance is explained by Factor 1.

Factor 2 mainly explains the variance of variables representing a party's proximity (around 0.70 for parameter estimates of *clospar* and *strongsup*) and some of the variables for voting (0.79 for *civduty*, 0.70 for *persbenef*, 0.90 for *willvote* as estimates of parameter and 62.9%, 49.4% and 81.5% for the squared proportion of variance). After these high values, we can observe still quite uniform and normal loadings that are never less than 0.2 for the other variables (with only one exception for *peopvote*).

Factor 3 on *Neighbourliness* shows how the first big group of variables representing relations within the neighbourhood are well explained by the latent factor. All the parameter estimates are quite high (from 0.60 upward with only one exception of 0.48) and from Table 2.11 at least 30 percent of their variance is explained on average by the construct. The second group of variables is about the fear of crime, use of media and relations with friends. All of them show an average loading of the estimates around 0.2. We can notice the negative relations between the variable for frequency of use of social media and the variable representing going out with friends. One theoretical justification for this last negative loading can be related to the possibility that the informal networks of friends does not coincide with the network of neighbours. Then the time spent with one network is inversely related to the time spent with the others. Some evidence confirming this arises from qualitative studies (Southerton, 2003). Another possible reason is that active people spend less time on social networks: they do it because they also have more information on what to do. Then, they usually spend time actively than being at home and spending time on social networks. This can be attributed to the overlap between SC and cultural capital, as described by Bourdieu (1986). Li (2015) finds this inverse correlation between active individuals and time spent in front of the television.

The main cross-loadings are for Factor 2 on variables loading on the other two factors. When I attempted to correlate the latent Factor 2 with Factor 1 and Factor 3 instead of making explicit so many cross-loadings, the results did not improve as much as when I correlated the residuals of the variables. Therefore, correlating the latent factors has not resulted as a necessary element in the fitting of this model and cross-loadings are all significant with a P-Value of 0.000 (as we can see from the diagram of the model in Figure 2.7). The lack of setting covariances between factors is also confirmed by the low covariances (Table 2.6). As already described, factors for 2011 seem to have a higher internal variance as well as a higher independence from the variation of the other factors. Finally, all these correlations and cross-loadings are also justified both at the empirical level (all of them are significant at 99 percent) and theoretical (same dimension and definition of SC) but are definitively less than the previous case, especially with regard also to the higher number of original variables involved in the model.

In reference to the general model fit (Hu and Bentler, 1995 and Di Stefano and Hess, 2005) and according to Schreiber *et al*. (2006) and like already described RMSEA, CFI and TLI are the preferable fit indices to consider for one-time analysis, the most common indices show the following values:

1. **Chi-Square Value**: 83918.581 with 921 degrees of freedom and a P-Value equal to 0.000;

2. **RMSEA**: 0.043 with 1.0 probability of RMSEA to be under the critical value of 0.05. Hu and Bentler (1999) suggested a cutoff value of <0.06;
3. **CFI**: 0.946, slightly close to the suggested cut off value of 0.95 (Hu and Bentler 1999);
4. **TLI** or **NNFI**: 0.941, slightly close to the suggested cut off value of 0.95 (Hu and Bentler 1999).

2.6 Conclusions

The models for SC factors proposed in this work confirm the idea of the multidimensionality of this concept. All the dimensions, both classical and more recent, have found strong confirmation: membership and active membership, voting behaviour, citizenship, party's closeness, informal networks, neighbourliness, use of social media and crime control.

Following Li (2015b) and Li *et al*. (2005) and the findings on the parallel operation of trust and membership (parallel but not causal), we can see how the models tested in this work produce an important conclusion: their fit was not possible if one variable or one factor would have been removed or changed. This means that SC has been confirmed as a complex and multi-construct capital. Indeed, as suggested at the conference on the state of SC in Britain (2015), we can identify different indicators interacting: informal support is necessary for health and happiness with small-scale economic benefits, whereas voluntary organizations may have a civic value but at the same time widening the social network. It is more effective to refer to *types* or *classes* of SC according to the degree to which the clear majority of individuals have 'mix and match' profiles of SC rather than 'high or low' levels.

However, each of the factors identified also work on their own as proper further variables, showing that each dimension has its own reliability and consistency on which it is possible to carry out analysis, confirming for example Swales (2015) that analysing participation in communities' associations find that it follows different and autonomous trend respect to the membership to other types of organizations. Therefore, we can see how membership and neighbourliness may rely on complementary but different patterns. In relation to previous studies, this work shows higher levels of richness and complexity both at theoretical and at empirical levels. All the dimensions analysed and the factors hypothesized have been derived from a careful literature review and have been strongly confirmed as significant both for 2001 and for 2011. The number of variables involved cannot be compared with previous studies such as the use of BHPS and UKHLS for this kind of analysis and purposes.

According to the 2001 and 2011 models, the classical dimensions of membership identified by the fathers of the concept (Bourdieu, Putnam and Coleman) and the other sub-classifications – active and passive membership and formal and informal organizations (Olson's organisation types) – are still actual and appear to still mirror attitudes and behaviour of people even over a time lapse of ten years. In a similar way, the dimension regards the party's closeness and citizenship remains constant between the two years considered. Despite the different loading of the variables with regard to the structure of the factors, all these kinds of variables result as strongly significant.

Keeping in mind clearly that it is not possible to compare statistically the two models because of the differences in sample, data and variables, we can just highlight some differences between the two models. The most important one is demonstrated in the third factor for both years. While variables representing caring and informal networks load significantly on a proper factor in 2001, the same does not occur for 2011 where instead variables about neighbourliness load on to a proper, strong factor. In the same factor, finally, we have seen how variables representing more recent topics are significantly explained in their variance component by this factor. Here we are referring to crime control and use of social media, new important aspects for 2011.

Apart from similarities and differences in relation to the variables, the structure of the latent factors also deserves further thought. The 2001 model has factors that are more dispersed yet slightly more correlated to each other, with higher covariance. The 2011 factors' covariances are low but they present high internal variance and a distribution approaching a more normal one. We can conclude then that, apart for a wider sample (that could explain this difference), while they have been created to capture a common pattern of variance they are also reliable on their own. These conclusions confirm the idea that the choice of using a variances analysis involving the identification of different factors relying on different dimensions of SC seems better than using other approaches involving the choice of variables or the building of a unique index of SC. Theoretically and empirically, SC is confirmed once more to be deeply linked with many aspects of the individual's life and their networks.

Appendix A

Table A1: Membership and active membership variables – the first recoding process to sum up original variables and obtain main variables for CFA – year 2001 and 2011

Final variable	Survey variable	Label
member1	Orgma	Member of political party
	Orgmb	Member of trade union
	Orgmo	Member of professional organisation
	Orgmp	Member of pensioners' organisation
member2	Orgmc	Member of environmental group
	Orgmd	Member of parents' association
	Orgmg	Member of voluntary service group
	Orgmh	Member of other community group
	Orgmk	Member of women institute
	Orgml	Member of women group
	Orgmm	Member of another organisation
member3	Orgme	Member of tenants or residents group
	Orgmi	Member of social group
member4	Orgmf	Member of religious group
member5	Orgmj	Member of sports club
	Orgmq	Member of Scout/Guides organisation
active1	Orgaa	Member of political party
	Orgab	Member of trade union
	Orgao	Member of professional organisation
	Orgap	Member of pensioners' organisation
active2	Orgac	Member of environmental group
	Orgad	Member of parents' association
	Orgag	Member of voluntary service group
	Orgah	Member of other community group
	Orgak	Member of women institute
	Orgal	Member of women group
	Orgam	Member of another organisation
active3	Orgae	Member of tenants or residents group
	Orgai	Member of social group
active4	Orgaf	Member of religious group
active5	Orgaj	Member of sports club
	Orgmq	Member of Scout/Guides organisation

Table A2: Other variables used – year 2001

Variable	Label
Relig	Attendance at religious services
govern1	Government reflects will of the people
govern2	Ordinary people can't influence gov't
govern3	Government ought to impose earnings ceiling
vote1	Supports a particular political party
vote2	Strength of support for stated party
vote3	Level of interest in politics
vote4	Voted in June 2001 general election
neigh	Frequency of talking to neighbours
meet	Frequency of meeting people
carenr	Provides care for non-resident person
help1ext	Someone outside HH can help if depressed
help2ext	Someone outside HH can help find job
help3ext	Someone outside HH can borrow money from
help1net	Is there someone who will listen
help2net	Is there someone to help in a crisis
help3net	Is there someone you can relax with
help4net	Anyone who really appreciates you
help5net	Anyone you can count on to offer comfort
likenbrd	Likes present neighbourhood
move	Prefers to move to a new house

Table A3: Other variables used – year 2011

Variable	Label
relig	Belong to a religion
att_relig	Attendance at religious services ordinal
imp_relig	Religion makes difference to life ordinal
clospar	Closer to a party if not a supporter
strongsup	Strongness of support/being closer
interpol	Interested in politics
polinf	Your vote will make a difference
civduty	Voting is a duty for citizen
engpol	Too much effort being involved in politics/public affairs
peopvote	People around me generally vote in elections
persbenef	Feel personal satisfaction when vote
willvote	Will vote before the next interview
preparpol	Feel qualified to participate in politics
informpol	More informed on politics than other people
pubofic	Public officials don't care about what people like me think
govnocare	People like me cannot say anything to the government
sources_n	Number of sources of news
knitneigh	This is a close-knit neighbourhood
helpneigh	People willing to help neighbours
trustneigh	People in this neighbourhood can be trusted
neighargue	People in this neighbourhood don't get along
belongne	Feel like belonging to neighbourhood
localfr	Local friends mean a lot
adviceloc	Can ask advice to neighbours
borrownei	Can borrow things from neighbours
improvnei	Feel like belong to neighbourhood
planstay	Plan to stay per year
similnei	Feel similar to neighbours
talknei	Stop and talk with neighbours
closefr	How many close friends_ordinal
socnet_chat	Number of hours spent chatting between Monday to Friday on social networks
mobile	Have a personal mobile
goout	Go out with friends when feel like it
fearcrime	Level of worry of a crime_ordinal
safedark	Feel safe walking in the dark

Table A4: Univariate proportions and counts for observed variables – year 2001

Variable	Category	Proportion	Counts	Variable	Category	Proportion	Counts
MEMBER1	1	0.762	12,088	GOVERN2	1	0.781	12,378
	2	0.202	3,195		2	0.219	3,478
	3	0.034	541	GOVERN3	1	0.830	13,155
	4	0.002	31		2	0.170	2,701
	5	0.000	1	VOTE1	1	0.610	9,677
MEMBER2	1	0.839	13,306		2	0.390	6,179
	2	0.132	2,090	VOTE2	1	0.749	11,870
	3	0.024	377		2	0.251	3,986
	4	0.004	70	VOTE3	1	0.600	9,510
	5	0.001	13		2	0.400	6,346
MEMBER3	1	0.878	13,917	VOTE4	1	0.322	5,105
	2	0.117	1861		2	0.678	10,751
	3	0.005	78	NEIGH	1	0.225	3,562
MEMBER4	1	0.872	13,833		2	0.775	12,294
	2	0.128	2,023	MEET	1	0.119	1,894
MEMBER5	1	0.818	12,971		2	0.881	13,962
	2	0.179	2,832	CARENR	1	0.884	14,021
	3	0.003	53		2	0.116	1,835
ACTIVE1	1	0.918	14,556	HELP1EXT	1	0.187	2,960
	2	0.075	1,186		2	0.813	12,896
	3	0.007	111	HELP2EXT	1	0.408	6,473
	4	0.000	2		2	0.592	9,383
	5	0.000	1	HELP3EXT	1	0.283	4,494
ACTIVE2	1	0.837	13,264		2	0.717	11,362
	2	0.140	2,213	HELP1NET	1	0.089	1,406
	3	0.020	325		2	0.911	14,450
	4	0.003	47	HELP2NET	1	0.100	1,589
	5	0.000	7		2	0.900	14,267
ACTIVE3	1	0.899	14,262	HELP3NET	1	0.095	1,511
	2	0.098	1,557		2	0.905	14,345
	3	0.002	37	HELP4NET	1	0.120	1,895
ACTIVE4	1	0.875	13,869		2	0.880	13,961
	2	0.125	1,987	HELP5NET	1	0.102	1,616
ACTIVE5	1	0.804	12,753		2	0.898	14,240
	2	0.191	3,027	LIKENBRD	1	0.073	1,165
	3	0.005	76		2	0.927	14,691
RELIG	1	0.767	12,160	MOVE	1	0.316	5,006
	2	0.233	3,696		2	0.684	10,850
GOVERN1	1	0.712	11,285				
	2	0.288	4,571				

Table A5: Univariate proportions and counts for observed variables – year 2011

Variable with category	Proportion	Counts	Variable with category	Proportion	Counts
MEMBER1			PREPARPO		
Category 1	0.783	35,325	Category 1	0.477	21,626
Category 2	0.179	8,076	Category 2	0.257	11,659
Category 3	0.035	1,599	Category 3	0.266	12,036
Category 4	0.003	129	INFORMPO		
Category 5	0.000	5	Category 1	0.479	21,741
MEMBER2			Category 2	0.319	14,470
Category 1	0.796	35,934	Category 3	0.202	9,174
Category 2	0.159	7,161	PUBOFIC		
Category 3	0.035	1,594	Category 1	0.498	22,517
Category 4	0.008	362	Category 2	0.292	13,193
Category 5	0.002	73	Category 3	0.209	9,460
Category 6	0.000	8	GOVNOCAR		
Category 7	0.000	2	Category 1	0.512	23,189
MEMBER3			Category 2	0.238	10,762
Category 1	0.910	41,057	Category 3	0.250	11,306
Category 2	0.087	3,942	SOURCES_		
Category 3	0.003	135	Category 1	0.010	467
MEMBER4			Category 2	0.568	25,923
Category 1	0.876	39,527	Category 3	0.331	15,097
Category 2	0.124	5,607	Category 4	0.091	4,174
MEMBER5			KNITNEIG		
Category 1	0.823	37,142	Category 1	0.214	9,760
Category 2	0.174	7,858	Category 2	0.279	12,752
Category 3	0.003	134	Category 3	0.508	23,199
ACTIVE1			HELPNEIG		
Category 1	0.903	40,761	Category 1	0.095	4,360
Category 2	0.085	3,841	Category 2	0.191	8,728
Category 3	0.011	475	Category 3	0.714	32,605
Category 4	0.001	47	TRUSTNEI		
Category 5	0.000	3	Category 1	0.094	4,260
ACTIVE2			Category 2	0.254	11,540
Category 1	0.795	35,866	Category 3	0.653	29,674
Category 2	0.160	7,233	NEIGHARG		
Category 3	0.035	1,566	Category 1	0.081	3,692
Category 4	0.008	361	Category 2	0.200	9,128
Category 5	0.002	82	Category 3	0.719	32,774
Category 6	0.000	18	BELONGNE		
Category 7	0.000	1	Category 1	0.077	3,117
ACTIVE3			Category 2	0.283	11,488

Category 1	0.928	41,889	Category 3	0.641	26,059
Category 2	0.070	3,156	LOCALFR		
Category 3	0.002	82	Category 1	0.105	4,281
ACTIVE4			Category 2	0.330	13,430
Category 1	0.876	39,536	Category 3	0.564	22,946
Category 2	0.124	5,591	ADVICELO		
ACTIVE5			Category 1	0.243	9,880
Category 1	0.826	37,275	Category 2	0.246	9,992
Category 2	0.170	7,653	Category 3	0.511	20,788
Category 3	0.004	199	BORROWNE		
RELIG			Category 1	0.357	14,517
Category 1	0.431	14,037	Category 2	0.230	9,359
Category 2	0.569	18,527	Category 3	0.413	16,787
ATT_RELI			IMPROVNE		
Category 1	0.765	24,343	Category 1	0.062	2,514
Category 2	0.077	2,447	Category 2	0.214	8,681
Category 3	0.158	5,037	Category 3	0.725	29,460
IMP_RELI			PLANSTAY		
Category 1	0.292	5,599	Category 1	0.135	5,487
Category 2	0.349	6,704	Category 2	0.187	7,598
Category 3	0.359	6,884	Category 3	0.678	27,551
CLOSPAR			SIMILNEI		
Category 1	0.450	22,345	Category 1	0.136	5,541
Category 2	0.181	8,993	Category 2	0.249	10,124
Category 3	0.370	18,371	Category 3	0.615	24,979
STRONGSU			TALKNEI		
Category 1	0.500	22,949	Category 1	0.133	5,413
Category 2	0.286	13,147	Category 2	0.194	7,904
Category 3	0.213	9,798	Category 3	0.672	27,343
INTERPOL			CLOSEFR		
Category 1	0.293	13,407	Category 1	0.036	1,645
Category 2	0.614	28,029	Category 2	0.645	29,283
Category 3	0.093	4,250	Category 3	0.252	11,428
POLINF			Category 4	0.056	2,562
Category 1	0.009	340	Category 5	0.006	278
Category 2	0.585	22,688	Category 6	0.003	147
Category 3	0.253	9,808	Category 7	0.001	42
Category 4	0.154	5,968	SOCNET_C		
CIVDUTY			Category 1	0.620	28,430
Category 1	0.021	837	Category 2	0.243	11,151
Category 2	0.195	7,790	Category 3	0.128	5,853
Category 3	0.163	6,494	Category 4	0.010	443
Category 4	0.621	24,744	MOBILE		
ENGPOL			Category 1	0.136	6,254
Category 1	0.360	14,257	Category 2	0.864	39,612

Category 2	0.365	14,462	GOOUT		
Category 3	0.275	10,889	Category 1	0.079	3,951
PEOPVOTE			Category 2	0.921	45,773
Category 1	0.106	3,799	FEARCRIM		
Category 2	0.336	12,030	Category 1	0.000	6
Category 3	0.558	20,015	Category 2	0.060	2,754
PERSBENE			Category 3	0.347	15,918
Category 1	0.030	1,185	Category 4	0.592	27,141
Category 2	0.247	9,805	SAFEDARK		
Category 3	0.259	10,270	Category 1	0.145	6,673
Category 4	0.464	18,372	Category 2	0.125	5,740
WILLVOTE			Category 3	0.729	33,452
Category 1	0.202	8,031			
Category 2	0.141	5,599			
Category 3	0.658	26,174			

Table A6: Correlation matrix of variables for 2001

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. MEMBER1														
2. MEMBER2	0.267													
3. MEMBER3	0.168	0.141												
4. MEMBER4	0.215	0.374	-0.015											
5. MEMBER5	0.246	0.091	0.125	0.079										
6. ACTIVE1	0.762	0.248	0.144	0.251	0.196									
7. ACTIVE2	0.186	0.867	0.071	0.355	0.062	0.262								
8. ACTIVE3	0.088	0.049	0.858	-0.042	0.089	0.165	0.075							
9. ACTIVE4	0.152	0.335	-0.039	0.938	0.045	0.292	0.384	-0.025						
10. ACTIVE5	0.170	0.063	0.063	0.048	0.896	0.211	0.113	0.123	0.077					
11. RELIG	0.094	0.239	-0.118	0.890	0.001	0.200	0.278	-0.095	0.891	0.006				
12. GOVERN1	0.104	0.029	0.005	0.072	0.072	0.111	0.018	-0.005	0.059	0.069	0.099			
13. GOVERN2	0.160	0.119	-0.014	0.143	0.125	0.130	0.110	-0.047	0.117	0.120	0.089	0.466		
14. GOVERN3	-0.051	-0.054	0.037	-0.014	0.004	0.012	-0.058	0.036	-0.007	0.008	0.036	0.491	0.233	
15. VOTE1	0.164	0.101	0.112	0.143	-0.016	0.159	0.076	0.111	0.103	-0.020	0.072	0.258	0.158	0.200
16. VOTE2	0.128	0.103	0.067	0.125	-0.034	0.192	0.075	0.106	0.090	-0.033	0.108	0.262	0.163	0.221
17. VOTE3	0.277	0.250	0.110	0.151	0.090	0.280	0.181	0.084	0.131	0.083	0.081	0.211	0.239	0.090
18. VOTE4	0.262	0.273	0.150	0.289	0.034	0.261	0.252	0.104	0.278	0.001	0.260	0.149	0.129	0.043
19. NEIGH	-0.018	0.084	0.182	0.077	-0.039	0.028	0.105	0.173	0.115	-0.028	0.065	-0.006	-0.036	0.021
20. MEET	-0.082	-0.001	0.069	0.021	0.104	-0.003	0.031	0.092	0.043	0.124	0.044	0.017	-0.012	-0.009
21. CARENR	0.086	0.176	0.123	0.145	0.002	0.085	0.185	0.139	0.132	0.022	0.113	-0.061	0.013	-0.045
22. HELP1EXT	0.046	0.064	0.010	0.085	0.102	0.008	0.065	0.020	0.078	0.113	0.015	0.022	0.074	-0.038
23. HELP2EXT	0.060	0.010	-0.017	-0.032	0.169	0.025	0.025	0.017	-0.023	0.173	-0.038	0.044	0.086	-0.001
24. HELP3EXT	0.060	-0.027	0.015	-0.048	0.126	0.001	-0.007	0.030	-0.037	0.133	-0.073	0.017	0.050	0.000
25. HELP1NET	0.045	0.081	0.016	0.128	0.100	0.012	0.078	0.011	0.118	0.084	0.066	0.057	0.055	0.028
26. HELP2NET	0.062	0.091	0.038	0.120	0.149	0.028	0.092	0.023	0.118	0.152	0.055	0.102	0.088	0.030

27. HELP3NET	0.005	0.019	0.019	0.065	0.119	0.013	0.052	0.044	0.057	0.117	0.036	0.048	0.038	-0.020
28. HELP4NET	0.062	0.092	-0.015	0.131	0.102	0.026	0.085	-0.015	0.123	0.086	0.098	0.077	0.095	0.046
29. HELP5NET	0.005	0.033	0.016	0.123	0.049	-0.011	0.049	0.043	0.107	0.049	0.107	0.035	0.060	0.035
30. LIKENBRD	0.116	0.148	0.087	0.170	0.125	0.112	0.116	0.018	0.162	0.107	0.162	0.130	0.117	0.088
31. MOVE	0.023	0.090	0.047	0.169	0.003	0.064	0.096	0.010	0.165	-0.013	0.198	0.074	0.037	0.073

	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
16. VOTE2	0.784															
17. VOTE3	0.485	0.603														
18. VOTE4	0.5600	0.515	0.457													
19. NEIGH	0.099	0.074	0.021	0.157												
20. MEET	-0.029	-0.021	-0.075	-0.084	0.183											
21. CARENR	0.070	0.008	0.030	0.136	0.090	0.161										
22. HELP1EXT	-0.002	-0.023	-0.007	-0.014	0.112	0.223	0.053									
23. HELP2EXT	-0.066	-0.050	0.024	-0.107	0.018	0.188	-0.025	0.541								
24. HELP3EXT	-0.040	-0.069	-0.012	-0.055	0.049	0.195	0.035	0.598	0.556							
25. HELP1NET	0.073	0.032	0.039	0.052	0.098	0.185	0.016	0.730	0.453	0.513						
26. HELP2NET	0.030	-0.012	0.035	0.070	0.121	0.201	0.015	0.639	0.469	0.610	0.795					
27. HELP3NET	0.015	-0.011	-0.015	-0.005	0.123	0.165	0.025	0.584	0.373	0.405	0.740	0.687				
28. HELP4NET	0.052	0.052	0.085	0.092	0.093	0.107	0.058	0.507	0.373	0.384	0.666	0.647	0.726			
29. HELP5NET	0.076	0.013	0.009	0.050	0.146	0.189	0.034	0.630	0.387	0.451	0.797	0.759	0.765	0.775		
30. LIKENBRD	0.088	0.040	0.089	0.185	0.142	0.046	0.002	0.139	0.072	0.081	0.162	0.196	0.151	0.177	0.166	
31. MOVE	0.095	0.072	0.017	0.174	0.145	0.007	-0.001	0.029	-0.042	-0.042	0.108	0.086	0.097	0.083	0.093	0.83

Table A7: Correlation matrix of variables for 2011

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. MEMBER1																		
2. MEMBER2	0.307																	
3.MEMBER3	0.223	0.192																
4. MEMBER4	0.222	0.369	0.064															
5.MEMBER5	0.219	0.127	0.170	0.069														
6. ACTIVE1	0.761	0.333	0.195	0.284	0.180													
7. ACTIVE2	0.230	0.838	0.132	0.327	0.096	0.381												
8. ACTIVE3	0.140	0.153	0.859	0.057	0.146	0.270	0.231											
9. ACTIVE4	0.170	0.316	0.040	0.937	0.043	0.325	0.390	0.103										
10. ACTIVE5	0.168	0.104	0.117	0.057	0.885	0.236	0.183	0.199	0.118									
11. RELIG	0.004	0.101	0.019	0.699	-0.053	0.096	0.102	0.014	0.610	-0.058								
12. ATT_RELI	0.054	0.182	-0.066	0.839	-0.048	0.139	0.168	-0.039	0.779	-0.035	0.786							
13. IMP_RELI	0.000	0.070	-0.086	0.546	-0.125	0.073	0.059	-0.067	0.479	-0.125	0.433	0.627						
14. CLOSPAR	0.219	0.196	0.143	0.182	0.079	0.220	0.157	0.125	0.161	0.067	0.158	0.087	0.055					
15. STRONGSUP	0.225	0.190	0.145	0.177	0.078	0.226	0.154	0.129	0.157	0.070	0.142	0.073	0.048	0.937				
16. INTERPOL	0.322	0.275	0.129	0.187	0.155	0.302	0.236	0.130	0.176	0.141	0.074	0.066	0.036	0.559	0.591			
17. POLINF	0.087	0.068	0.044	0.102	0.009	0.093	0.072	0.055	0.102	0.021	0.144	0.126	0.054	0.341	0.342	0.294		
18. CIVDUTY	0.279	0.270	0.171	0.264	0.071	0.263	0.242	0.138	0.246	0.064	0.218	0.188	0.074	0.546	0.540	0.511	0.465	
19. ENGPOL	0.135	0.134	0.028	0.061	0.074	0.151	0.136	0.041	0.063	0.080	-0.042	0.024	-0.002	0.119	0.141	0.269	0.101	0.089
20. PEOPVOTE	-0.033	0.019	0.030	0.072	0.032	-0.005	0.029	0.029	0.083	0.033	0.108	0.102	0.034	0.103	0.093	0.043	0.118	0.159
21. PERSBENEF	0.200	0.181	0.126	0.203	0.013	0.194	0.158	0.109	0.194	0.020	0.205	0.165	0.068	0.528	0.524	0.449	0.451	0.733
22. WILLVOTE	0.285	0.299	0.153	0.268	0.123	0.259	0.259	0.120	0.251	0.115	0.223	0.196	0.080	0.661	0.642	0.605	0.455	0.752
23. PREPARPOL	0.294	0.221	0.072	0.118	0.130	0.264	0.179	0.063	0.112	0.126	0.004	0.055	0.033	0.321	0.341	0.540	0.186	0.325
24. INFORMPOL	0.254	0.193	0.088	0.126	0.130	0.237	0.151	0.090	0.106	0.108	0.041	0.078	0.068	0.365	0.381	0.571	0.184	0.352
25. PUBOFIC	0.121	0.120	-0.006	0.113	0.125	0.124	0.137	0.004	0.102	0.130	0.013	0.069	0.003	0.115	0.115	0.162	0.164	0.149
26. GOVNOCARE	0.178	0.160	0.004	0.146	0.121	0.166	0.171	0.012	0.139	0.133	0.026	0.087	0.025	0.161	0.162	0.235	0.213	0.236
27. SOURCES_N	0.287	0.215	0.092	0.092	0.202	0.225	0.222	0.111	0.097	0.220	-0.061	-0.036	-0.068	0.144	0.162	0.324	0.067	0.179
28. KNITNEIGH	-0.032	0.041	0.068	0.064	-0.006	0.003	0.058	0.078	0.065	0.010	0.113	0.105	0.045	0.037	0.038	-0.025	0.069	0.069
29. HELPNEIGH	0.096	0.140	0.099	0.123	0.094	0.092	0.144	0.086	0.118	0.098	0.075	0.080	-0.010	0.109	0.106	0.089	0.090	0.178

30. TRUSTNEIGH	0.122	0.174	0.115	0.145	0.113	0.109	0.167	0.083	0.140	0.105	0.115	0.076	-0.029	0.179	0.165	0.151	0.106	0.251
31. NEIGHARGUE	0.168	0.179	0.104	0.114	0.148	0.122	0.172	0.088	0.106	0.145	0.016	-0.015	-0.074	0.119	0.117	0.149	0.060	0.201
32. BELONGNE	0.051	0.147	0.165	0.173	0.055	0.087	0.150	0.163	0.143	0.058	0.180	0.147	0.058	0.121	0.106	0.045	0.120	0.187
33. LOCALFR	0.021	0.159	0.176	0.166	0.014	0.066	0.169	0.178	0.158	0.026	0.183	0.147	0.062	0.100	0.091	0.042	0.112	0.184
34. ADVICELOC	0.007	0.126	0.153	0.135	0.006	0.046	0.135	0.149	0.129	0.015	0.132	0.117	0.047	0.062	0.056	-0.007	0.078	0.134
35. BORROWNEI	0.076	0.127	0.099	0.043	0.102	0.083	0.143	0.108	0.047	0.108	0.005	0.031	-0.012	0.044	0.047	0.052	0.050	0.089
36. IMPROVNEI	0.151	0.224	0.144	0.142	0.116	0.169	0.234	0.170	0.158	0.117	0.052	0.091	0.033	0.121	0.122	0.174	0.113	0.219
37. PLANSTAY	0.067	0.131	0.192	0.115	-0.007	0.070	0.128	0.163	0.104	-0.017	0.115	0.055	0.015	0.119	0.101	0.030	0.093	0.195
38. SIMILNEI	0.030	0.053	0.167	0.073	0.004	0.034	0.066	0.147	0.058	-0.002	0.144	0.063	-0.015	0.093	0.077	-0.012	0.098	0.178
39. TALKNEI	0.056	0.160	0.221	0.124	0.003	0.090	0.180	0.216	0.123	0.015	0.113	0.060	0.021	0.084	0.078	0.037	0.092	0.172
40. CLOSEFR	0.131	0.167	0.135	0.130	0.161	0.140	0.175	0.141	0.125	0.148	0.016	0.041	-0.027	0.068	0.068	0.107	0.014	0.088
41. SOCNET_CHAT	-0.041	-0.057	-0.147	-0.136	0.109	-0.051	-0.034	-0.089	-0.111	0.116	-0.185	-0.109	-0.100	-0.174	-0.150	-0.095	-0.099	-0.273
42. MOBILE	0.142	0.163	0.179	0.080	0.247	0.141	0.131	0.195	0.085	0.235	-0.023	0.024	-0.057	0.026	0.036	0.073	0.050	0.063
43. GOOUT	0.197	0.133	0.005	-0.038	0.271	0.095	0.132	0.009	-0.018	0.266	-0.189	-0.087	-0.104	-0.125	-0.084	0.073	-0.048	-0.075
44. FEARCRIME	-0.048	-0.010	-0.014	0.001	0.006	-0.037	-0.015	0.002	-0.023	-0.002	-0.047	-0.018	-0.021	-0.013	-0.015	-0.048	-0.022	-0.022
45. SAFEDARK	0.191	0.090	0.053	-0.036	0.262	0.103	0.077	0.068	-0.026	0.258	-0.163	-0.054	-0.102	0.026	0.038	0.131	0.005	0.016

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
20. PEOPVOTE	-0.025																	
21. PERSBENEF	0.114	0.106																
22. WILLVOTE	0.156	0.207	0.649															
23. PREPARPOL	0.298	-0.010	0.300	0.363														
24. INFORMPOL	0.203	-0.030	0.321	0.380	0.688													
25. PUBOFIC	0.242	0.057	0.166	0.190	0.172	0.107												

26. GOVNOCARE	0.277	0.031	0.235	0.270	0.255	0.173	0.676											
27. SOURCES_N	0.201	-0.012	0.140	0.227	0.264	0.232	0.153	0.203										
28. KNITNEIGH	-0.016	0.227	0.078	0.045	0.003	-0.004	0.047	0.024	-0.029									
29. HELPNEIGH	0.007	0.244	0.147	0.184	0.071	0.055	0.116	0.103	0.043	0.645								
30. TRUSTNEIGH	-0.002	0.292	0.199	0.255	0.119	0.104	0.119	0.114	0.025	0.504	0.671							
31. NEIGHARGUE	0.072	0.186	0.139	0.224	0.080	0.048	0.131	0.135	0.124	0.381	0.569	0.590						
32. BELONGNE	-0.032	0.243	0.168	0.186	0.020	0.023	0.077	0.069	-0.017	0.554	0.564	0.534	0.422					
33. LOCALFR	-0.028	0.203	0.164	0.155	-0.002	0.007	0.029	0.016	-0.016	0.511	0.510	0.447	0.330	0.719				
34. ADVICELOC	-0.032	0.190	0.126	0.113	-0.025	-0.020	0.030	0.030	-0.011	0.466	0.505	0.418	0.297	0.609	0.724			
35. BORROWNEI	0.024	0.121	0.075	0.086	0.066	0.059	0.070	0.080	0.068	0.387	0.491	0.355	0.281	0.469	0.544	0.627		
36. IMPROVNEI	0.094	0.117	0.176	0.225	0.140	0.114	0.104	0.125	0.140	0.254	0.369	0.329	0.288	0.400	0.422	0.394	0.442	
37. PLANSTAY	-0.088	0.158	0.162	0.155	-0.011	0.010	-0.013	-0.021	-0.066	0.322	0.361	0.398	0.285	0.572	0.489	0.425	0.322	0.348
38. SIMILNEI	-0.097	0.208	0.166	0.135	-0.042	-0.029	0.003	-0.006	-0.055	0.377	0.423	0.457	0.337	0.613	0.553	0.502	0.372	0.380
39. TALKNEI	-0.035	0.175	0.152	0.138	-0.009	0.016	-0.001	-0.002	0.008	0.478	0.506	0.415	0.336	0.614	0.638	0.596	0.539	0.438
40. CLOSEFR	0.076	0.055	0.076	0.110	0.079	0.070	0.082	0.090	0.161	0.098	0.146	0.134	0.157	0.136	0.157	0.155	0.142	0.129
41. SOCNET_CHAT	0.131	-0.055	-0.234	-0.186	0.019	-0.057	0.071	0.058	0.146	-0.073	-0.113	-0.206	-0.089	-0.192	-0.177	-0.151	-0.034	-0.049
42. MOBILE	0.070	0.075	0.037	0.097	0.081	0.049	0.091	0.114	0.183	0.061	0.143	0.121	0.167	0.146	0.151	0.117	0.143	0.184
43. GOOUT	0.129	-0.057	-0.109	-0.043	0.130	0.061	0.099	0.120	0.310	-0.123	-0.032	-0.093	0.069	-0.138	-0.138	-0.124	0.070	0.133
44. FEARCRIME	-0.001	0.078	-0.024	-0.010	-0.006	-0.009	0.066	0.038	-0.069	0.159	0.186	0.265	0.168	0.182	0.109	0.102	0.065	0.025
45. SAFEDARK	0.108	0.057	0.003	0.069	0.182	0.143	0.146	0.164	0.186	0.116	0.191	0.215	0.213	0.159	0.067	0.064	0.191	0.201

	37	38	39	40	41	42	43	44
38. SIMILNEI	0.611							
39. TALKNEI	0.476	0.544						
40. CLOSEFR	0.061	0.088	0.120					
41. SOCNET_CHAT	-0.301	-0.245	-0.209	0.065				
42. MOBILE	0.028	0.113	0.143	0.269	0.145			
43. GOOUT	-0.203	-0.138	-0.064	0.154	0.504	0.360		
44. FEARCRIME	0.143	0.137	0.103	0.015	-0.034	0.068	-0.074	
45. SAFEDARK	0.070	0.083	0.100	0.109	0.138	0.298	0.349	0.326

Supplementary Material

Missing data analysis

As anticipated at paragraph 2.2.2, in this section I am going to check the Missing At Random (MAR) assumption for 2001 sample, given the difference in the observations between the full sample and the sample of the estimated model with listwise deletion (18,867 versus 15,586). The first step was to run Independent Sample T-Tests to compare means of several variables between the two groups: full sample and sample with missing data from the estimates of the models. The hypothesis is that there is no difference between the means, meaning that missing data are random and they do not bias results.

I used variables about sex, age and educational attainment from the original dataset[10]. I choose them because they do not have missing data at all and they can give information about the full sample. The hypothesis of no differences between the means of the groups has been rejected.

I then run logistic regressions to investigate it further. Dependent variable is a coded 1 for observations with at least a missing value and 0 otherwise. It then identifies the two samples.

Results are reported in the following table:

[10] Age is a continuous variable; Sex is a dummy variable with 1 for Female, Education level is a categorical variable with 6 values: No qualifications, other qualifications, GSCE, A-level, Degree, Other higher degree.

Table SM.1.1: Logistic regression results

	Results
Constant	-2.881*** (0.119)
Age	0.020*** (0.001)
Sex	0.237*** (0.048)
Educational level	-0.190*** (0.019)

Without focusing on the sign of the relation, we can see that all the variable are significant, leading to the rejection of MAR assumption (indeed, they should have resulted not significant).

To check if differences between these means are significantly greater than zero I then finally check descriptive statistics of the two samples respect to the variables used in the logistic regression and also respect to the original variables used for the Factor Analysis. As we can see from the following table, the differences are small. It seems big differences between the two samples are not relevant, therefore this adjust for MAR assumption (or MCR).

Table SM.1.2: Descriptive statistics

	Full sample			Sample without missing		
Variable	**Observations**	**Mean**	**Std. dev.**	**Observations**	**Mean**	**Std. dev.**
age	18867	45.348	18.575	15860	44.417	17.835
sex	18867	1.542	0.498	15860	1.540	0.498
educational level	18867	2.584	1.273	15860	2.634	1.253
member1	17908	0.264	0.519	15860	0.276	0.530
member2	17908	0.195	0.491	15860	0.196	0.491
member3	17908	0.123	0.343	15860	0.127	0.348
member4	17908	0.130	0.336	15860	0.128	0.334
member5	17908	0.177	0.390	15860	0.185	0.397
active1	18056	0.088	0.309	15860	0.089	0.311
active2	18056	0.189	0.466	15860	0.191	0.468
active3	18056	0.099	0.306	15860	0.103	0.311
active4	18056	0.126	0.331	15860	0.125	0.331

active5	18056	0.189	0.403	15860	0.200	0.412
relig	18867	0.230	0.421	15860	0.233	0.423
govern1	18063	0.282	0.450	15860	0.288	0.453
govern2	18062	0.214	0.410	15860	0.219	0.414
govern3	18061	0.169	0.374	15860	0.170	0.376
vote1	18061	0.387	0.487	15860	0.390	0.488
vote2	18061	0.250	0.433	15860	0.251	0.434
vote3	18061	0.389	0.488	15860	0.400	0.490
vote4	18061	0.671	0.470	15860	0.678	0.467
neigh	18061	0.774	0.418	15860	0.775	0.417
meet	18059	0.880	0.325	15860	0.881	0.324
carenr	18059	0.113	0.317	15860	0.116	0.320
help1ext	18056	0.788	0.408	15860	0.813	0.390
help2ext	18051	0.561	0.496	15860	0.592	0.492
help3ext	18003	0.687	0.464	15860	0.717	0.451
help1net	17988	0.887	0.317	15860	0.911	0.284
help2net	17481	0.893	0.310	15860	0.900	0.300
help3net	17477	0.899	0.302	15860	0.905	0.294
help4net	17373	0.875	0.331	15860	0.880	0.324
help5net	17316	0.894	0.307	15860	0.898	0.303
likenbrd	17191	0.921	0.269	15860	0.927	0.261
lkmoven	17188	0.678	0.467	15860	0.684	0.465

Mplus syntax for Model 2001

```
INPUT INSTRUCTIONS
TITLE: CFA with Categorical Outcome Variables;
listwise=on;
variable: names are member1 member2
member3 member4 member5 active1
active2 active3 active4 active5 relig
govern1 govern2 govern3 vote1 vote2
vote3 vote4 neigh meet carenr help1ext
help2ext help3ext help1net help2net help3net
help4net help5net likenbrd move;
categorical are member1 member2
member3 member4 member5 active1
active2 active3 active4 active5 relig
govern1 govern2 govern3 vote1 vote2
vote3 vote4 neigh meet carenr help1ext
help2ext help3ext help1net help2net help3net
help4net help5net likenbrd move;
missing are all (-9999);
analysis: type is general;
iterations=1000;
```

```
convergence=0.00005;
difftest;
model: F1 by member1 active1 member3 active3
govern1 govern2 govern3
vote1 vote2 vote3 vote4;
F2 by member2 member4 member5
relig active2 active4 active5;
F3 by neigh meet carenr
help1ext help2ext help3ext
help1net help2net help3net help4net
help5net likenbrd move;
member1 with active1;
member2 with active2;
member5 with active5;
member1 with active5;
member1 with member2;
member1 with member5;
member2 with member5;
member1 with active2;
active1 with active2;
active1 with member2;
member3 with active3;
active1 with active3;
member1 with active3;
member1 with member3;
member2 with member3;
govern1 with govern2;
govern1 with govern3;
govern3 with govern2;
vote1 with vote2;
vote2 with vote3;
vote3 with member2;
vote4 with member2;
vote4 with active2;
help2ext with help3ext;
help2ext with member5;
help2ext with active5;
help2net with help3ext;
help1net with help2net;
help1net with help3net;
likenbrd with move;
move with member2;
move with active2;
move with relig;
move with vote4;
F1 with F2;
F1 with F3;
output: sampstat modindices stdyx standardized;
tech4;
```

Mplus syntax for Model 2011

```
INPUT INSTRUCTIONS
Title: CFA with Categorical Outcome Variables;
listwise=on;
Variable: names are member1 member2 member3 member4
member5 active1 active2 active3 active4 active5 relig
att_relig imp_relig clospar strongsup interpol polinf
civduty engpol peopvote persbenef willvote preparpol
informpol pubofic govnocare sources_n knitneigh
helpneigh trustneigh neighargue belongne localfr
adviceloc borrownei improvnei planstay similnei
talknei closefr socnet_chat mobile gout fearcrime
safedark lkhere moved moved_yr moved_3yr
moved_3_5yr moved_5_10yr moved_10yr jobsat satisheal
satisinc satisleis satislif happyrel likenei
internetfr hrstv health illness carehome careout
careout_n careout_par careout_net carehrs caresome
concenslee useful decis stress diffic enjoy problem
unhappy confid worthless happy healthsum moth_see
moth_con moth_cont fath_see fath _con fath_cont
chil_see chil_con child_cont help_par help_child
help_fr child visit_fam;
Missing are all (-9999);
categorical are
member1 member2 member3 member4
member5 active1 active2 active3 active4 active5
relig att_relig imp_relig clospar strongsup interpol
polinf civduty engpol peopvote persbenef willvote
preparpol informpol pubofic govnocare sources_n
knitneigh helpneigh trustneigh neighargue belongne
localfr adviceloc borrownei improvnei planstay similnei
talknei closefr socnet_chat mobile goout fearcrime
safedark;
usevariables are member1 member2 member3 member4
member5 active1 active2 active3 active4 active5
relig att_relig imp_relig clospar strongsup interpol
polinf civduty engpol peopvote persbenef willvote
preparpol informpol pubofic govnocare sources_n
knitneigh helpneigh trustneigh neighargue belongne
localfr adviceloc borrownei improvnei planstay similnei
talknei closefr socnet_chat mobile goout fearcrime
safedark;
Analysis:
Type = general missing;
coverage=.06;
iterations=10000;
convergence=0.00005;
difftest;
estimator=WLSMV;
model:F1 by member1 member2 member3 member4
```

```
member5 active1 active2 active3 active4 active5
relig att_relig imp_relig;
F2 by clospar strongsup interpol polinf
civduty engpol peopvote persbenef willvote
preparpol informpol pubofic govnocare sources_n;
F3 by knitneigh helpneigh trustneigh
neighargue belongne localfr adviceloc borrownei
improvnei planstay similnei talknei
closefr socnet_chat mobile goout fearcrime
safedark;
member1 with active1;
member2 with active2;
member3 with active3;
member5 with active5;
strongsup with clospar;
F2 by member1;
F2 by member2;
F2 by active1;
F2 by active2;
F2 by att_relig;
F2 by trustneigh;
F2 by improvnei;
F3 by member3;
att_relig with relig;
persbenef with civduty;
informpol with preparpol;
preparpol with interpol;
informpol with preparpol;
govnocare with pubofic;
safedark with fearcrime;
safedark with goout;
goout with socnet_chat;
output: sampstat modindices stdyx standardized;
tech4;
tech10;
```

CHAPTER 3 – SOCIAL CAPITAL AND LOCAL AREA EFFECTS: A MULTILEVEL MODEL ANALYSIS

3. 1 Introduction

The previous chapters have highlighted widely how SC is a complex and multidimensional concept. However, especially, how it is deeply related to several personal and individual life dimensions: from membership to crime, from neighbourliness to voting behaviours, from use of social media to trust towards the Institution, particularly Government. Based on these evidences, a natural question arises: how much do other personal characteristics influence it? What might be the other aspects affecting its development?

In order to address these questions, I turn my view to a particular modelling approach that has increasingly been used over the last few years: Multilevel Models. These kinds of models have been developed to take into account the hierarchical and clustered nature of particular kinds of data. The cluster dimension can be referred to different levels: human (individuals, households), organisations (firms, schools, hospitals, etc.) or geographical (local area, regions, country).

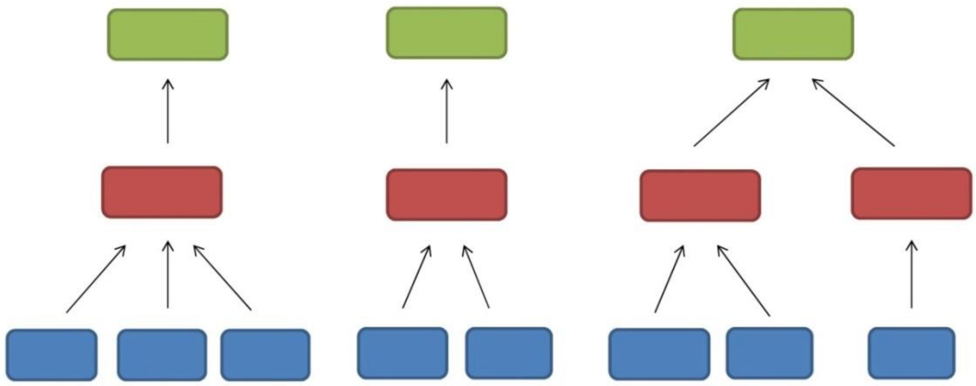
As will be described later in the chapter, it is possible to identify an increasing interest on this last level. Indeed, the geographical level, especially at local and restricted areas, can catch all the aspects related to living in communities that naturally affects individual's networks, acquaintances and possibilities to modify them.

Having made clear that SC is an asset dependent both on individual aspects and local dimensions, I will try to identify, using a Multilevel Model (MM hereafter), which personal characteristics are correlated with SC (in a positive and negative way) and at which geographical level we can start explaining its differentials. The three factors of SC developed previously with Confirmatory Factor Analysis will be our dependent variables. The MM will be developed from a Null Model, in order to see the level of variance that can be explained from differences at local area without any covariates involved until the more complex Random Slope Models, where SC varies between all the individuals and are contemporary to all of the areas involved in the analysis.

3.2 Multilevel Models: The theoretical assumptions and empirical implications

Multilevel Models - also called Random Effect Models, Hierarchical Linear Models or Mixed Models just to cite some titles – are models that, as Goldstein (2010) widely described, catch the clustered and hierarchical structure of observational data collected in biological and human sciences. From experimental collections of data to the national survey data, many data show a hierarchy consisting of units grouped at different levels. This particular structure can be represented graphically like a diagram, as in Figure 3.1, where each colour represents a different level (in this case, up to 3 levels).

Figure 3.1: Diagram representation of clustered data



One basic example occurs when same individuals and units are measured on more than one occasion. These examples are frequent in clinical trials and laboratory experiments or weather collection data. In this case, the lowest level 1 are the measurements occasions nested at level 2 represented by individuals.

Longitudinal studies can be also applied to analyse, in this case, changes over the time of an individual and differences between individuals (growth studies). With observational data, the possible hierarchies may depend on the topic. In a study about school performances, the lowest level can be represented by students or pupils and their test scores, level 2 can be the classes and the last level can be the schools.

In organizational and management studies, level 1 can be the employers nested in Departments, and at the highest third level, in Companies. At a more geographical level, as frequently occurs within survey data, the level 1 of the observations can be individuals (or households) nested in Areas further clustered in Regions. Whichever the topics and the levels identified, MM have been increasingly used, due to the fact that they take into account this aspect, making it the central idea of the model. The statistical implications of these procedures are evident. While the standard statistical analysis is based on the assumptions that observations are independent between themselves, in MM the clustering process of observations in larger analytic units results in:

- a higher correlation between observations in the same units than the average. It is also called the *within effect* and it is analysed in the *fixed part* of the model at level 1;
- the lower similarities between observations from different clusters than the average. It is called *between effect* and it is analysed in the *random part* of the model, corresponding to all the other levels hypothesized (level 2 and 3 and so on).

These factors can be caused from selective factors involved in the grouping of individuals or joint patterns of variances with respect to similar influences or mutual interactions (Kish, 1967:163).

The underestimation of this aspect can lead to underestimated standard errors and missed identification of the influences of higher levels. Moreover, MM helps to take also into account another important interdependence between observations: the interdependence originating from the survey design. Important and large surveys, such as the U*nderstanding Society* survey that I intend to use for this purpose, use methods that on the one hand try to produce correct and representative estimates and on the other hand decrease the impressive costs. One of the methods for the latter aspects is the choice, during the sampling procedure, of observations reducing the distance that the interviewer should cover. Therefore, frequently we can have individuals and/or households quite close each other.

It is quite evident in this case how some kind of territorial dependency and similarity may be present between observations. All the surveys present some statistical adjustment and weights to account for non-independence of observations from the same cluster. Instead of simple post-hoc corrections, MM allows us to consider it and even insert it in the analysis as a valid aspect in itself, deserving the status of a further source of variation (Brunton-Smith and Sturgis, 2011 and Goldstein 2010).

Further, it is the statistical structure of a MM in itself that helps to take in account all these aspects. In this study, a two-level model grouping individual within local areas will include residuals both at first and at second level. Therefore, the residual variance will be partitioned into a *between-areas (*level 2) component (variance of the area-level residuals) and a *within-areas* (level 1) component (variance of the individual-level residuals). The area-level residuals of level 2 – also called 'area effect' – will represent all the unobserved area characteristics that may affect the individual outcomes identifying correlations between outcomes for individual of the same area (elaboration on University of Bristol, Centre for Multilevel Modelling online material)[11].

[11] http://www.bristol.ac.uk/cmm/learning/multilevel-models/what-why.html. Accessed on October 2015.

3.2.1 Theoretical assumptions and types of Multilevel Models

Before showing how MM are built according to an even higher level of complexity, it is important to indicate the main assumptions of these kind of models, this will be important also for some goodness of fit check.

As summarized by Dedrick *et al*. (2009) and Snijders T. A. B. and Berkhof J. (2007), MM holds the same assumptions of linear regression but they are modified to fit the hierarchical nature of data design. Furthermore, some of these assumptions that I am going to list, are also important for the post estimation process. The main assumptions are:

- *linearity*: function forms are linear at each level;
- *normality of level 1 residuals*. They are assumed to be independently and normally distributed with covariance Σ. Lack of normality can bias the standard errors at all the levels, leading to a question about the validity of the statistical tests and the accuracy of the confidence intervals identified. The check of the normality of distribution of the residuals at this level is considered a good check of the goodness of fit. Where this assumption does not hold because of the type of the outcome variable (binary, ordinary or multinomial), other models are appropriate, like the Hierarchical Generalized Linear Model (Dedrick *et al*., 2009).

  Therefore, level 2 random effects have a multivariate normal distribution;
- *homoscedasticity*: level 1 residual variance is constant;
- *independence*. Level 1 residuals and level 2 residuals are uncorrelated. Besides, independence is also considered between observations: at a higher

level, they are independent of each other (Rabe-Hesketh S., Skrondal A., 2012).

After pointing out the main theoretical assumptions behind MM, as briefly suggested in the introduction, it is possible to see how they can be built going from a simple model up to more complex ones (Snijders and Bosker, 1999; Rabe-Hesketh S., Skrondal A., 2012). An important precondition is that, as suggested previously, MM are composed of a *fixed part* and a *random part*. Intuitively, the fixed part corresponds to the level 1, where observations are analysed in respect to a variable or more (covariates) referred to the observations in themselves. The clustered and hierarchical nature of the data and its effects on level 1 observations is mainly investigated in the random part of the model, corresponding to the level 2 of the model (and further higher levels where considered). At this level, therefore, different hypotheses can be made in respect of the intercept component, as well as the slope component of the model (Hayes, 2006).

Therefore, the unexplained residual variance is partitioned into two parts: higher-level variance between higher level entities and lower-level variance within these entities and between observations. Each level has a residual term and it is the residual at a higher level which is the so-called *random effect* (Bell and Jones, 2015).

Starting from the simplest case, the first model is called Null Model (or Baseline Model, Intercept-only Model or Unconditional Model) and is used to calculate the Intra Class Correlation Index (ICC hereafter) that indicates the degree of similarities between observations within the same cluster. Called also Variance Partition Coefficient (VPC hereafter), statistically it indicates the proportion of variance in our outcome variable that is accounted for by the clustering. It gives us mainly the percentage of the total residual variance after controlling for all the covariates. It is also useful to understand if data are really clustered or not, therefore indicating if MM are required with respect to a classical OLS regression. If it is close to 0, it means that there is no variation and that the clusters' means will not differ from the mean of normal regression. If it approaches to 1, it means that there is no variance to explain at the individual level, meaning that all the individuals are the same.

Formally, the equation of a Null Model is:

$$y_{ij} = \beta_0 + u_{0j} + e_{ij} \qquad [3.1]$$

With

$e_{ij} \sim N(0, \sigma_e^2)$ and [3.1 b]

$u_{0j} \sim N(0, \sigma_u^2)$

where i is for the observations, in this case individuals, and j is for the cluster, in this case geographical areas. β_0 is the overall mean across areas – the intercept constant across areas, u_{0j} is the area effect of area j on the outcome y and it identifies the between-area variance (level 2) in the random part of the model and e_{ij} is the within-area between-individual (level 1) variance residual that identify the within-area variance.

The Null Model, by definition, does not include a fixed part with explanatory variables and this explains the subscript 0 for the individual observations in the term u_{0j}. For the following formulation of the ICC, we can call it σ_u^2 . Contemporarily, e_{ij} is the within estimated variances of the residuals at area-level in the random part of the model and it varies between areas[12]. We can call it σ_e^2.

The formula of the ICC allows us to understand better the calculus of the variance explained only at cluster level:

$$ICC = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \qquad [3.2]$$

To test if there is a significant area effect, the Null Model can be compared with a Single-level Model (carrying out a likelihood ratio test). The Single-level Model is a model without the random area effect:

$$y_{ij} = \beta_0 + e_{ij} \qquad [3.3]$$

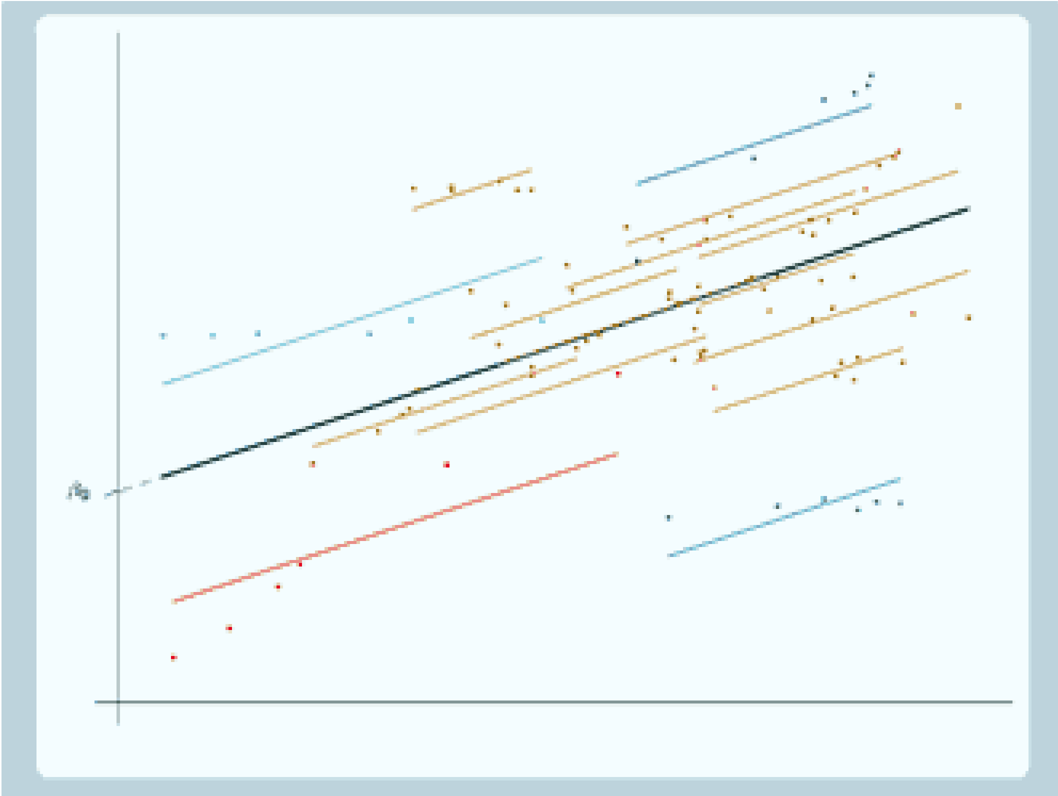
After the illustration of these basic MM used to identify the necessity or not of a clustered model, the further step in building a MM is the Random Intercept Model (RIM hereafter). Assuming that most of the phenomena under investigation in the social sciences depend on different variables at the same time, the RIM includes different predictors at level 1 that are assumed to vary and affect the outcome.

[12] If we would like to calculate the magnitude of the variation among clusters in their mean individuals' levels, we could calculate the plausible values range for these means based on the between variance we obtain from the model and this calculus would include also the intercept from the fixed part of the model.

The model is so composed by two components: an observational level and a cluster level with residuals terms for each level. Respectively, as previously described, they identify a *between effect* of variation at observation level (level 1) and a *within effect* of variation at cluster level (level 2). In my specific case, I identify an effect of portioning variance between-area effect and a within-area effect with individuals at level 1 and geographical area at level 2 (this is the reason why RIM is also called the Variance Component Model). In this first case, intercepts are allowed to change according to the hypothesis that there is random variation between individuals in different areas but that the effects of explanatory variables are assumed to be the same for each area. A plot of the predicted area regression lines would show a series of parallel lines next to each other, one for each area[13]. A simple example of univariate Random Intercept Model is showed below:

[13] http://www.bristol.ac.uk/cmm/learning/multilevel-models/what-why.html. Accessed on October 2015.

Figure 3.2 : Example of a plot of regression lines of a general univariate Random Intercept Model



Source: Centre for Multilevel Modelling, University of Bristol, online training material

Formally the RIM presents this notation:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij} \qquad [3.4]$$

where β_0 is the intercept, $\beta_1 x_{ij}$[14] is the explanatory variable of the model. The fitted line for the explanatory variable at level 1 identified will differ from its average line in its intercept by an amount of the level 2 residuals $\hat{u}._{0j}$for area j. Finally, e_{ij} is the within-area residual that identify the within-area variance (at level 1). Here, we can

[14] In a Multivariate Random Intercept Model, we can just add other explanatory variables in the formula ($\beta_2 x_{ij}$ + …), more generally identified with $\beta_n x_{ij}$ according to the number of explanatory variables of the model.

still see how the slope of the area lines is assumed to be fixed at $\beta_1$for the respecting variable x_1.

Looking deeper at the image above, we can interpret it relating to the formula just explained. The overall average line β_0 corresponds to the single level regression model intercept and it is considered the fixed part of the model. The difference with a classical single level regression model is the random part of the model, which in the figure corresponds to the parallel and multi coloured lines. In this random part, the overall intercept line is still β_0 but each area line has an intercept of $\beta_o + u_j$ that identifies each line parallel to the overall average line. Including the parameter u_j from the random part of the model, we are allowing it to vary determining the 'random intercepts'. Interpreting the parameters, in the random part we finally estimate σ_u^2 and σ_e^2 that, according to [3.1 b] are the variance of area effect u_j term (between-area variance at level 2) and the variance residual at area-level term e_j(within-area between-individual variance at level 1).

RIM can be built in a more and more complex way. The first option consists of adding more explanatory variables, as for a single level regression model. We can also add ***interactions*** between them. This affects variation at level 2 (may increase, decrease or stay the same) but level 1 variation and total residual variation will either decrease or stay at the same.

A second option is adding explanatory variables defined only at level 2. They are actually called ***contextual effects***. These variables frequently can be or aggregated at a higher level with respect to the level of the explanatory variables used in the model or different variables at a different level of aggregation not yet included. They are also called *ecological variables* when they are inserted in the model to see which 'external' variables regarding the nature of the cluster can 'indirectly' affect the outcome.

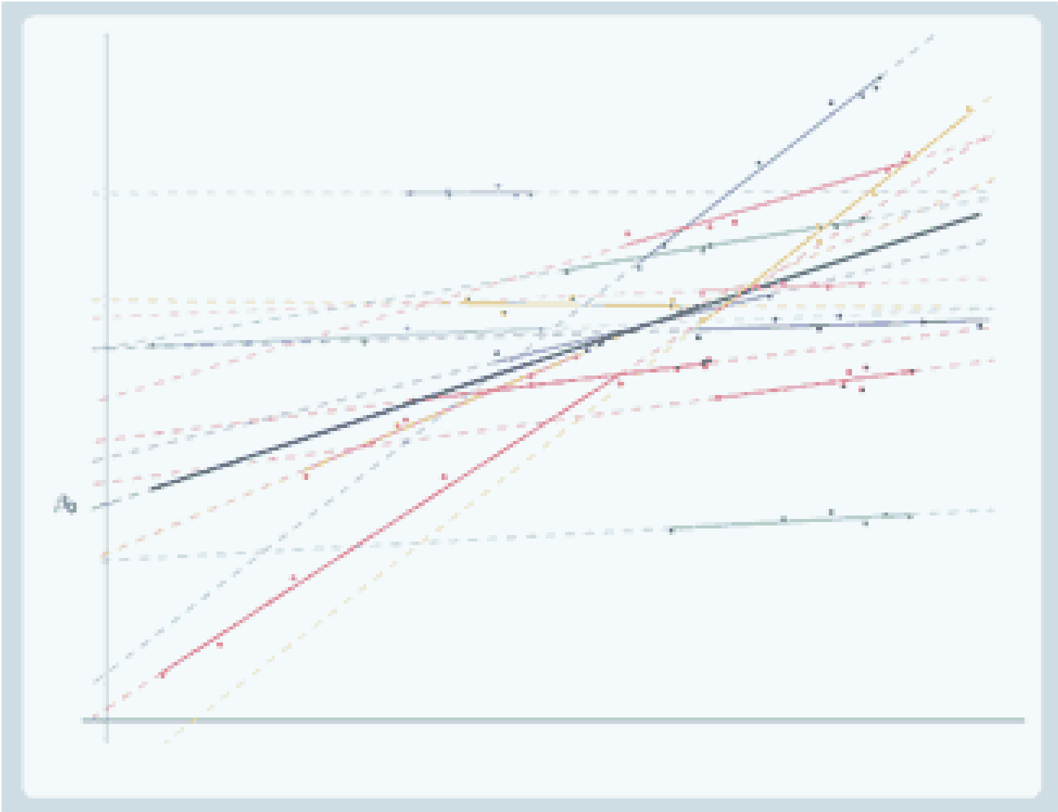
These group-level variables are used to see if the dependent variable may be influenced by other unmeasured group factors. The difference with the between effect measured lies not only in the possibility that different data can be used but mainly because they do not vary from observations to observations within the level 2 group.

Assuming that changes in the outcome are the same for all the areas can reduce one of the principal advantages implicit in MM. Because of this consideration, frequently, especially in the social sciences, RIMs are increasingly made more complex, allowing different slopes at clusters' levels. We obtain the so-called Random Slope Model or Random Coefficient Model (RSM and RCM hereafter), where changes in the outcomes are different not only within the areas but also between them. They indeed allow all individual level coefficients to vary across areas, not just the intercept term (Goldstein, 2010).

Graphically, the regression lines of a RSM are[15]:

[15] http://www.bristol.ac.uk/cmm/learning/multilevel-models/what-why.html. Accessed on October 2015.

Figure 3.3: Example of a plot of regression lines of a univariate Random Slope Model



Source: Centre for Multilevel Modelling, University of Bristol, online training material

The formal notation is the following:

$$y_{ij} = \beta_0 + \beta_{1j}x_{1ij} + u_{0j} + u_{1j}x_{1ij} + e_{0ij} \qquad [3.5]$$

where the difference with the previous model is the added term $u_{1j}x_{1ij}$. Indeed, the different slopes for all the areas are given exactly by these slope residuals (the subscript of *u* indeed changes). In this way, the coefficient for the explanatory variable x_1, for example, will be $\beta_{1j} = \beta_1 + u_{1j}$, with β_{1j} varying between all the areas j. This coefficient is composed of the average effect across all areas, β_1, and the residual difference from the average effect in each area u_{1j}.

The residual term has a variance $\sigma_{u1}^2$with the usual assumptions and it is the term representing the unexplained variation. So, at the area-level, the model fits the average intercept and the difference from the average intercept, but also the average coefficient and the residual difference from the average coefficient in each area.

Besides, because now there are two error terms at the area-level, there is an additional covariance term between the unexplained intercept and the coefficient variance. We can represent the complex area-level variation with the following variance matrix Ω_u:

$$\begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} \sim N(0, \Omega_u), \Omega_u = \begin{bmatrix} \sigma_{u0}^2 & . \\ \sigma_{u01} & \sigma_{u1}^2 \end{bmatrix} \quad [3.6]$$

The difference between figure 3.2 and 3.3 is clear: we have different intercepts and different slopes. Explanatory variable has a different effect for each area, meaning that each area line has a different slope. While β_0 can be interpreted in the same way of RIM, the difference is aboutβ_1, the slope of the average line: its average increases (or decreases) across all the areas in y for a one unit of change in x_1 differently. Consequently, we have three variances parameters: σ_{u0}^2 is the variance in intercepts between areas, σ_{u1}^2 is the variance in slopes between areas while σ_{u01} is the covariance between intercepts and slopes.

As for the previous case of RIM, also for RSM we can add more explanatory variables, interactions and corresponding random slopes. We can add random slopes on one predictor, several or all of them. Number of units at level 2 in the datasets may affect the decision on which explanatory variables it is possible to test the random slope. A final option applicable to the MM is the test of ***cross-level interactions***. In this case, RSM allow us to see if there is an effect on the outcome variable produced by the interaction of a level 1 variable and a level 2 variable.

Generally, we can then easily state that MM can be applied to complex phenomena and sometimes it can be difficult to select the variables of interests and even the levels of study. By the way, some general guidelines can help in preventing mistakes that can make the model biased or wrong.

The first suggestion is about considering or not random coefficients for variables that do not vary at a lower level. If a covariate is not significant at level 1, is it advisable to add it like a random slope in the model? Unfortunately, there is no a general agreement about this aspect in the literature. Theoretical reasons and empirical evidence can only drive and inspect this aspect, considering other problems like over identification (Douglas, 2004; Rabe-Hesketh and Skrondal, 2012).

Besides, it may be tempting to allow different covariates to have random slope. However, the number of the parameters increases rapidly with the number of random slopes. Indeed, for each random effect (intercept or slope), we should consider that we have a variance parameter and a covariance parameter for each pair of random effects. This procedure can bias the identification of the model (frequently with over specification) and lead to difficult and long computational timing. It is recommended to be more flexible in the fixed part of the model than in the random part. (Rabe-Hesketh, Skrondal, 2012).

3.2.2 Other theoretical issues

After describing the theoretical models and the types of MM and before beginning the empirical part, it is important to underline an aspect that is becoming central in the debate about MM. Several authors are focusing on solving this issue that can also have an impact at the empirical level. One of the first steps in building MM is concerned with checking the real necessity of using them. Indeed, even we can assume a hierarchical and clustered nature of the data, in reality, for several reasons, this cannot be always valid. We can summarize this problem as the choice between the Random Effects (RE) model and the Fixed Effects (FE) model, where FE can be identified by using a normal OLS regression.

Following a primary paper by Fielding (2004) where doubts were raised about the Hausman Test as a crucial test to determine if FE or RE should be used, Bell and Jones (2015) have recently published an important paper where they show several important proofs and considerations concerning this problem.

They begin by identifying why RE are not widely used yet: the exogeneity assumptions of RE, that is, that the residuals are independent of the covariates - $Cov\left(x_{ij}, u_j\right) = 0$ and $Cov\left(x_{ij}, e_{ij}\right) = 0$ - often do not hold in RE. Therefore, they investigate why, frequently, RE has been abandoned in favour of FE estimation. Alternatively, this endogeneity is composed of the two different 'between' and 'within' effects. But FE reduces it, using a procedure that, according to the two authors, causes an omitted variable bias: all higher-level variance (and corresponding 'between' effect) is controlled out using higher-level entities included in the model as dummies. They end up estimating only the 'within' effect.

In order to test if this type of form of bias exists in the RE, the Hausman Test is frequently carried out. However, according the author, it is problematic when it is used to address if FE or RE have to be used:

> …it is simply a diagnostic of one particular assumption behind the estimation procedure usually associated with the random effects model…it does not address the decision framework for a wider class of problem…' (Fielding in Bell and Jones, 2015). '...to reiterate: the Hausman test is not a test of FE versus RE; it is a test of the similarity of within and between effects. A RE model that properly specifies the within and between effects will provide identical results to FE… (Bell and Jones, 2015).

Without focusing excessively on this aspect, after addressing all the principal objections along with simulations, their main points which are important for this work are:

- that understanding the difference between the within-effect and between-effect is crucial when choosing modelling strategies;
- that the drawback of RE – correlated lower-level covariates and higher-level residuals – is omitted-variable and can be solved with a Mundlak's formulation (1978)[16] or following simplification (not yet widely used, by the way). In this way, FE and RE are identical because the only difference disappears: the incorrect specification which ignores the correlation between the effects and the explanatory variables;
- once this problem is solved, RE is much better than FE because they are also readily extendable with random coefficients, cross-level interactions and

[16] Mundlak's proposes a partition of the effect of lower-level covariates into two parts and he simply adds one additional term in the model for each varying covariates that accounting for the between effect: that is, the higher-level mean and it is treated in the same way as any higher-level variable. The Raudenbush method in Bell and Jones (2015) rearranges Mundlak's formulation in a simpler way, using group-mean centering.

complex variance functions and they are flexible and generalizable. Disregarding these aspects would lead to a poor model and misleading results;

- RE is explicit in relation to context and heterogeneity, rather than controlling out (as explicit models do) and this aspect is more interesting and real for researchers and policy makers;
- endogeneity is treated as a substantive phenomenon and not ignored;
- once the above aspects have been proved, FE remains only a special and restrict case of RE that results therefore more appropriate;
- in cross-sectional cases such as this study, heterogeneity at the individual level can also be explicitly modelled by including additional random coefficients and cross-level interactions, using RCM.

Finally, the authors underline how this problem is as much philosophical as it is statistical. RE are not perfect, especially when higher-level units are few. But the choice depends on the intelligence of the researchers and on their research questions. Where context and ecological outcomes are as important as the individual characteristics in shaping the phenomena in the research questions, then RE are definitively more appropriate.

3.3 Social Capital Factors: Multilevel Models and first empirical issues

3.3.1 Introduction and empirical use of MM in Social Capital literature

In this paragraph, after the theoretical introduction to MM and before presenting the models I intend to examine, I will briefly outline the empirical literature concerning the application of MM on SC topics. As we have seen from the previous chapter, the literature relating to singular aspects and dimensions of SC is impressive and wide ranging. It seems that the increasing popularity of MM goes simultaneously with its applications on this topic and all the dimensions previously described and tested: from membership to trust, from voting attitudes to caring, from use of media to fear of crime.

Not only that, in this introduction I am going to list some of the studies where MM has also been applied to those individual characteristics that, according to my hypothesis, affect levels of SC. Such dimensions, that this work will use MM as explanatory variables at the individual level, are mainly age, sex, ethnicity, marital status, educational attainment, socio-economic classification, employment status, religion and so on. Further references will also be present in the final sections with the interpretation of the results.

We can then see how MM has been applied to study the role of ethno-cultural diversity and SC in fostering national cohesion across European Societies (Reesken and Wright, 2013) finding that nationalism is related to lower levels of SC. MM has been also used in a study on self-rated health, welfare regimes and trust in Europe (Rostila, 2007). The author finds how contexts with low trust are detrimental for the health of distrustful individuals. Subramanian *et al*. (2006) confirms the same results for an American case whereas Carpiano (2007) on the same topic finds more ambiguous results, using data from Los Angeles survey on Family and Census data. In addition, neighbourhood mortality rates seem to be affected from SC through health (Lochner *et al*., 2003).

The studies linking SC and health with applications of MM are numerous. This is also due to the fact that Health Studies have originally been one of the first topics where MM have been applied and widely used. In addition, it is mainly in this field of studies that MM are applied to research questions about SC, health (and different issues: obesity, smoking, drinking, diseases and so on) and geographical dimension, leading to further estimations at the small area level. SC, its sociodemographic characteristics, health and small area analysis for England has been carried out beginning with the use of MM by Mohan *et al*. (2005). Important previous references for this study are Wing *et al*. (1992) which develops an analysis on American State Economic Areas and Morris *et al*. (1996), which applies the analysis at the Local Education Authority level in England. Many other studies can be reported since even as recently as the 90s, when MM began to be used more widely.

In relation to neighbourliness, MM has been applied to understand differentials of SC. Subramanian *et al*. (2003) attempted to discover if individual, socioeconomic and demographic attributes may eventually affect SC levels. The results suggest that even controlling for many of these variables (age, sex, race, marital status, income, education) significant neighbourhood differences remain linked to personal trust. This confirms the idea that SC is a true contextual construct.

Snelgrove *et al.* (2009), using BHPS in a longitudinal study and applying classical concepts of SC (generalized trust, civic participation and mutual reciprocity), attempt to examine the association between area SC indicators and individual poor self-rated health. They find positive association between higher self-rated health in areas where social trust is higher, after controlling for individual characteristics. After adjustment, they do not find evidence for an association between area civic participation and self-rated health.

Another study sees the application of MM to the topic of SC and demographic characteristics. Perna and Titus (2005) study the relationship between ethnicity and parental involvement as SC in the college enrolment in the USA, finding a lower share of actual enrolment for African Americans and Hispanic Americans than White. Many further studies can be listed here, showing that this topic is highly interesting and important, although there is not the within the confines of this study to discuss this further.

3.3.2 Empirical specifications

After showing the applications of MM in SC studies, I am going to present the model concerning the three factors of SC for 2011 that I have previously created with CFA (see Chapter 2):

- Factor 1: Membership
- Factor 2: Citizenships and Politics
- Factor 3: Neighbourliness

The choice to focus only on 2011 Factors has been mainly for two reasons:

- the entire PhD project referred to the ONS Project ‘Beyond 2011’, aiming to find an alternative way of estimation to the use of Census covariates. Indeed, the next chapter will see the estimation of Small Area Estimates of SC levels for England and Wales based on the MM presented in this chapter;

- as described widely in the previous chapter, the CFA for 2011 presents a higher level of generalizability. Therefore, I consider it to be more appropriate to begin with such a model in order to allow for the further estimations at the small area-level.

I will use the Understanding Society survey[17] in a cross-sectional study. The year upon which I focus is 2011 (the last Census year). I consider all the respondents interviewed in 2011, wave 3, because questions about SC on which Factors of SC were created are presented only within this wave. This is despite the fact that their observations interviewed in 2011 are also in the year 2, wave 2 (following the new scheme of timing of data collection). The MM proposed in this chapter will use variables about individual characteristics from all the waves of UKHLS and from the previous BHPS and its 18 waves. The final sample, useful for this analysis, contains 37,932 observations.

The geographical level I choose as a cluster unit for level 2 is the Middle Super Output Area (MSOA). This choice is justified by several reasons:

- lower levels, like Lower Super Output Area (LSOA) or the smallest Output Area (OA) would have been an interesting level of analysis but the number of units and the number of observations per unit available in the survey is too low in relation to the total number of areas. Besides, OAs are not available in the Secure Data Access for this survey; a special licence is required in order to access them;

[17] For further description of the surveys please see paragraphs 2.1 and 2.3.2, Chapter 2.

- MSOA seems to be the first geographical level with a high enough number of units available in the survey for a proper MM and for the following step of estimation at small area level;
- other classifications are available. While the current one is a Census classification, administrative Geography, Electoral Geography, Health Geography and Postcodes are available. Between the several Census classifications (Built-Up area/sub-division, Enumeration District, Output Area, Statistical Wards/CAS Ward and ST Ward, Super Output Area and Workplace Zone), the choice of the Super Output Area is also constrained by its more easily corresponding availability in the UKHLS. The geographical merge required between Census and survey for MM and the further analysis at the small area can then be carried out;
- theoretically, beyond the statistical reasons, MSOA seems to be the right level of analysis for a phenomenon such as SC, one that is strong locally. Higher levels of analysis, like Local Authority Districts or Government Offices Regions (GOR) are probably much too wide and large for such a type of capital analysis.

The analysis is carried out only for England and Wales due to the difficult comparison of data and geographies with Scotland and Northern Ireland. The number of MSOA for England in 2011 are 6,791 and for Wales are 410. Other principal information about MSOA 2011 are:

Table 3.1: Information about MSOA 2011 for England and Wales

	Minimum Population	Maximum Population	Minimum number of Households	Maximum number of Households
MSOA	5,000	15,000	2,000	6,000

The sample of MSOA for this study consists in 5255 different areas with 426 MSOA that have only one observation in the survey. All the residual areas have therefore at least two observations. In the following table I show the frequencies of the areas:

Table 3.2: Frequencies table for MSOA in the survey

Number of repeated observations for MSOA	Frequencies	Percentage	Cumulative percentage
0	426	1.12	1.12
1	4,829	12.73	13.85
2	4,829	12.73	26.58
3	4,141	10.92	37.50
4	3,654	9.63	47.13
5	3,173	8.36	55.50
6	2,763	7.28	62.78
7	2,383	6.28	69.07
8	2,009	5.30	74.36
9	1,703	4.49	78.85
10	1,435	3.78	82.63
11	1,186	3.13	85.76
12	995	2.62	88.38
13	821	2.16	90.55
14	665	1.75	92.30
15	559	1.47	93.78
16	460	1.21	94.99
17	372	0.98	95.97
18	302	0.80	96.77
19	234	0.62	97.38
20	181	0.48	97.86
21	143	0.38	98.24
22	113	0.30	98.53
23-62	556	1.53	100

One last specification regards the choice about the cluster units to be analysed. SC is an individual resource but strictly linked and dependent on the individual’s networks. The primary and natural network for a person is the family; therefore, my first idea was to structure the MM with 3 levels: individuals nested in households (HH) nested in MSOAs.

The number of households in this sample is 20,971 and, according to the number of members, we have the following frequencies:

Table 3.3: Households by number of members

Numbers of members in the HH	Frequencies	Percentage	Cumulative percentage
1	8,170	21.54	21.54
2	19,834	52.29	73.83
3	5,835	15.38	89.21
4	2,816	7.42	96.63
5	840	2.21	98.85
6	252	0.66	99.51
7	105	0.28	99.79
8	40	0.11	99.89
9	9	0.02	99.92

Slightly more than the half of the sample is composed by single individuals or by HH with only two members, so I considered that, even if being in a couple can have a strong effect on the level of SC, I may not be able to identify a 'familial' effect according to a wider definition. Besides, there are some notable statistical reasons supporting the choice to do not include the HH level.

As eloquently described by Twigg *et al*. (2000), they are the following: the proportion of one person HHs in the sample, the violation of the normality assumptions and the difficulty in separating out the between-HH and within-HH effect, because of the confounding across levels.

With regard to this particularly study, the choice of a two-level model rather than three-level model has been leaded mainly by the predictive goal of this research.

The following use of models for Small Area Estimates mainly suggests keeping the models as simpler as possible. The main risk may be the attribution of HH level variation to MSOA level, as the following table show:

Table 3.4: ICC for Null Models by type of MM for each factor

ICC for type of MM	Factor 1	Factor 2	Factor 3
MM – 2 levels			
MSOA	0.130	0.120	0.130
MM – 3 levels			
MSOA	0.070	0.064	0.085
HH | MSOA	0.500	0.370	0.470

From the table above where first tests on the Baseline Models hypothesizing 3 levels are reported, we can see how important the HH effect is, and also how much it reduces the variation explained at MSOA level in the Null Models. I estimate that HH and MSOA random effects compose approximately the 50 percent, 37 percent and 47 percent of the total residual variance for Factor 1, Factor 2 and Factor 3. However, these 3 levels MM reduce considerably the variance explained at MSOA level. While it is about 6-8 percent for the 3 levels MM, it increases by more or less 5 percent in 2 levels MM.

Therefore, this preliminary analysis opens definitively other possibilities of investigations that, for the reason explained, cannot be conducted herein[18].

Other empirical specifications are valid for all of the three factors: all individual covariates have been selected from the survey based on existing research on the individual correlates of SC. Besides, possible interactions have been tested with the

18 Other practical aspects can be considered to support this choice: availability of Census covariates at HH level and their cross-interactions for the Small Area Estimates, geographical details' availability for these covariates and lastly, especially for RCM, the intensive computational needs of a three-level model with contextual effects and cross-level interactions.

usual analysis, especially in the case of unexpected insignificance variables (according to hypothesis and literature). One last important specification concerns the case of Random Coefficient Models. For completeness of presentation, I described them in the previous paragraph about the theoretical and empirical applications of all the types of MM. RCM have been tested using all the individual predictors for all the three factors but they did not show any particular or interesting result. I chose therefore to do not present them. Further considerations on this aspect will be outlined in the Conclusions.

Following on empirical specification, as for Chapter 2, I check also if the change in the number of observations in all the models is due to at least Missing At Random data (MAR) (see par. 2.2.2, Ch. 2 for further explanations). The software I use in this study, Stata, and the command I decided to use for these models – xtmixed – the listwise deletion is the default option for handling missing data: observations with at least one missing data on the variables of interest are dropped out automatically. As we saw in the previous chapter (par. 2.2.2, Ch. 2), it reduces the availability of data but it seems a better option for studies where full information about variance of the phenomena is needed for less biased results. I also thought that in any case the number of observations retained in all the models would have been enough for defining a good sample size. Therefore, at the end of the Appendix of this chapter, I present a Supplementary Material section, where results and relative interpretations are shown in order to see if the main, strict assumption of listwise deletion, the Missing At Random Data or Missing Completely At Random, hold in my models.

Finally, for all the RIMs and CEM the main reference is the formula 3.4 while formula 3.2 is the reference for ICC.

3.3.3 Creation of the contextual variables: a Principal Component Analysis

One further empirical aspect is the creation of the contextual variables. In the description of the types of MM, I briefly outline that one specification of RIM can present *contextual effects* that are mainly area-level covariates. As Brunton-Smith and Sturgis (2011) and Brunton-Smith (2011), to identify such ecological variables I use a Factor Analysis. Therefore, I will retain the richness and completeness of the Census' information with a smaller number of variables (Components of Factors) that are linear combinations of the originals. I choose as method the Principal Component Analysis to create Components (PCA) with Census variables at MSOA level according to the same definition of the variables in the survey. This method, different from the CFA, does not assume any a priori or theoretical assumptions about the factor and it is a more exploratory analysis. Oblimin rotation (with Kaiser normalization when rotating factors – for an easier interpretation) is assumed. I chose this oblique type of rotation with respect to the more famous Varimax because it assumes correlated factors, an assumption that I consider more realistic.

Furthermore, beyond the Census variables, I include variables from the Neighbourhood Statistics about: benefits and tax credits, rates of crimes, density per area and distance travelled to work. These variables have been standardized in order to allow the comparison of effect size between variables of different scales. Definitions are available in Appendix B. Generally, all the variables chosen from the

Census can be referred to characteristics that, according to my hypothesis, identify more heterogeneous areas in term of compositions.

Finally, before showing the results, I indicate the formula I use to build an Ethnicity Index. I assume, indeed, that there is an important ethnicity effect that would not result in significance in the Factor analysis because of the low counts of some ethnicities in the areas. So instead of keeping the original variables divided by all the ethnicities used in the Census (White, Black, Asian, Mixed and Other ethnicities (Arab and other ethnic groups), I build a kind of index measuring the degree of ethnical homogeneity of an area with a range from 0 to 1 where 1 is an area with the highest level of ethnic diversity, following the more famous Herfindhal Index structure. According to Brunton-Smith (2011) this formula gives us the probability of two randomly selected individuals from the same locality being of different ethnic origin. The ethnic fractionalisation index has been regularly used to identify differences of ethnic heterogeneity.

The formula is:

$$1 - \sum_{i=1}^{n} s_i^2 \qquad [3.7]$$

Specifically, in this case:

$$1 - \left[\left(\frac{White}{Tot\ Resident}\right)^2 + \left(\frac{Mixed}{Tot\ Resident}\right)^2 + \left(\frac{Black}{Tot\ Resident}\right)^2 + \left(\frac{Asian}{Tot\ Resident}\right)^2 + \left(\frac{Other}{Tot\ Resident}\right)^2 \right] \qquad [3.8]$$

The steps to choose the variables are:

- reference to the Communalities results for the choice about which variables to retain[19]. Communality is the total amount of variance that an original variable shares with all the other variables included in the analysis. A threshold of 0.5-0.6 is recommended. Some of the variables show a low communality but, according to the literature, it is also possible to retain variables that are important theoretically. Sex, crime rates and commuting variables are important to SC theory. Besides, despite the low communality, the final model maintains really good level of goodness of fit;

Table 3.5: Communalities of PCA

Variable	Extraction
Male	.006
Crime_propert	.147
WTA_less10km	.176
Crime_violen	.195
Muslim	.458
Emplo	.501
Age30_44	.567
Single	.606
Tax_credit	.612
Christ	.628
Qual_lev2	.629
Born_EU	.632
Density	.635
No_care	.653
House_terrac_flat	.710
Resid_5more	.741
House_owned	.759
Sick_benef	.763
Rooms_2	.773
Health_good	.777
Qual_lev3	.807
Born_UK	.828
Indu_terz	.846
Nssec_prof	.891
Income_benef	.892

[19] Instead of reporting the huge correlation matrix between all the variables, Communalities results are another kind of indicator of variance and covariance. This indeed indicates the amount of the variance shared between the variables and it gives easier readable indications. Indirectly, Communalities suggest the correlation between variables.

- reference to the main goodness of fit of model's indicators:
    - Kaiser-Meyer-Olkin's measure of sampling-adequacy. It varies between 0 and 1. A value of 0 indicates that the sum of partial correlations is large relative to the sum of correlations, indicating diffusion in the pattern of correlations. In this case, factor analysis is inappropriate. 1, on the other side, a value close to 1 indicates that patterns of correlations are compact, so factor analysis is appropriate. Values greater than 0.5 are appropriate but ranges that are more specific have been indicated: 0.5-0.7 is mediocre, 0.7-0.8 is good, 0.8-0.9 is great and higher than 0.9 is superb. The PCA analysis presents a KMO of **0.850**.
    - Bartlett's Test of Sphericity: this test has a null hypothesis stating that the original correlation matrix is an identity matrix (meaning that all the correlations between variables would be zero). In this case, a factor analysis would be useless, because we need variables that are correlated between them, showing a kind of relationship. The P-Value for this analysis is 0.000, so it is highly significant: we reject the null hypothesis. PCA is appropriate.

Before showing the Components, in the following tables we can see the eigenvalues associated with each factor representing the variance explained by that particular linear component:

Table 3.6: Eigenvalues and variance explained

	Initial eigenvalues		
Component	**Total**	**% of Variance**	**Cumulative %**
1	8.583	34.333	34.333
2	6.649	26.597	60.930

3	1.884	7.535	68.465
4	1.717	6.869	75.334
5	1.293	5.173	80.507
6	1.079	4.317	84.824
7	.643	2.573	87.396
8	.590	2.359	89.756
9	.545	2.181	91.937
10	.374	1.498	93.435
11	.315	1.259	94..694
12	.218	.873	95.567
13	.183	.732	96.299
14	.164	.657	96.955
15	.149	.595	97.550
16	.117	.467	98.017
17	.096	.384	98.401
18	.077	.309	98.710
19	.075	.302	99.012
20	.067	.268	99.280
21	.052	.209	99.489
22	.050	.200	99.689
23	.034	.135	99.824
24	.028	.110	99.934
25	.017	.066	100.00

As we can see, the first two factors already explain about the 60 percent of the variance. And, indeed, the analysis presents two main factors. According to the twenty-five variables used and the results that I am going to interpret, I name the two factors:

- Factor 1: ***Heterogeneity***. It reflects the level of heterogeneity of the areas with respect to structural variables of population (personal characteristics), crime rates of the areas and housing profile.
- Factor 2: ***Economic profile***. It reflects more closely the aspects of the areas relating to Employment status and related characteristics and economic conditions, strictly dependent on economic status too.

These two factors are composed in this way:

Table 3.7: Component Matrix

Covariates measures	Heterogeneity	Economic profile
1 or 2 rooms house	**.877**	-.061
Born in UK	**-.866**	-.277
House owned	**-.843**	.219
Terraced or flat house	**.842**	.009
Resident since 5 year or more	**.819**	.265
Christian	**-.789**	-.070
Population density	**.784**	.142
Single	**.755**	-.189
Born in Europe	**.746**	.274
Do not provide care	**.723**	.360
Muslim	**.668**	-.109
Age 30-44	**.625**	.420
Violence against the person	**.431**	-.099
Criminal damage	**.356**	-.145
Male	**.078**	.008
High and Low Managerial, Administrative and Professional and Intermediate, Small Employers, Own Account (NS-SEC classif.)	-.277	**.902**
Tertiary sector worker	.194	**.899**
Good health	.025	**.881**
Degree and higher qualifications	.217	**.872**
Benefits for disability	.321	**-.813**
A levels	.119	**.784**
Benefits for economic disadvantages	.598	**-.731**
Employed	-.194	**.680**
Tax credit	.453	**-.638**
Distance travelled to work: less than 10 km (WTA classif.)	.220	**.357**

The final results show some interesting considerations and one main limitation.

The first, big factor explaining the 35 percent of the variance is composed by all the individual characteristics that we consider, for the most part, natural: place of birth, sex and age. Other characteristics are still personal but, in some way, referring to long lasting dimensions of an individuals' life that can change slightly easier than the previous listed dimensions: religion, caring and marital status. Finally, the other variables sharing most of the variance with these variables are those ones regarding the housing aspect. Examining more closely the composition of the factor, we can see that areas that have higher proportions of people between thirty and forty-four, single, Muslim, resident for a relevant time, Europeans and not involved in the important activities of caring and (not strongly important with a really low score loading) males with an average size of house, terraced house or flat, in populated and dense areas, will have higher degree of *heterogeneity*. This aspect can also justify the higher counts of crime for these areas (keeping in mind that these variables are at Local Authority District level, a higher level than MSOA, leading maybe to a loss of details and disaggregation). Areas with a lower degree of *heterogeneity* will have lower scores than English born proportions, house owners and Christians[20].

The second factor explains around about 25 percent of the total variance. It is composed of characteristics that are acquired at some point in the life, that can be also changeable (in a less disruptive way) and that can be improved during the lifetime: qualification levels, employment status, socio-economic classification and industrial characterization of the job. These characteristics are obviously linked to the other variables about economic status: the claim of benefits and credits.

According to the structure of this second factor, areas that will have a high score on *Economic profile* have also high proportions of employed people, in the professional socio-economic class, with good health and quite educated (high school and higher levels – graduate and postgraduate) and not commuting in an important way. On the other side, these areas will have minor proportions of people claiming different kinds of benefits and credits.

[20] Given the low counts of the other religions at small area level, I use only Muslim as a main 'other religion' than Christian.

The only main limitation is about some minimal cross-loadings between the two factors. The overall goodness of the model validates to maintain the final structure presented.

Finally, the final factors have been tested in the MM after being standardized. Standardization facilitates the interpretation of the estimated coefficients when the variables have different units of measurement, exactly as in this case.

3.3.4 Main hypothesis

Before starting to present the results for the three dimensions of SC, I would need to specify some theoretical hypothesis. As anticipated in the Introduction to this work, I am carrying out a work that I would define as 'exploratory and descriptive'. The aim here is to identify which aspects influence SC in order to be able to estimate it. There is a main, strong, theoretical assumption that holds behind all of the empirical hypotheses: SC like a 'multidimensional' concept', will probably bring us different results for the three factors (CIS 15, 2015). It will show higher levels by some measures and low on others.

Specifying empirical assumptions and hypothesis, for the purpose of this study, I will then build up the MM testing individual covariates and area-level covariates. All the individual covariates are personal characteristics that, according to the literature, have been found to be significantly linked to SC. I will also indicate my main expectations about the sign and the strength of the relations, trying to report references. These references will help to try to understand the results or, at least, to give an idea about the possible reasons behind it, keeping the question and the debate open.

In the building process, several variables have been tested with reference to the literature[21], and some of them have not resulted significant at all: ownership of the house, type of house (flat, terraced, etc.), number of cars, heating, years of residence, size of the household, type of employment, number of hours worked per week, commuting and industrial sector of the employment. But some of these variables resulted significant in the components at Census area-level, as we have just seen. Interesting observations can be completed but, for obvious reasons, this is not the appropriate context and further research can be considered in the future.

The use of area-level covariates is a further step done to fully understand how much the context affects this important capital. Also in this case, the choice of the variables used in the PCA has been done according to the classical hypothesis and studies. Regarding the factors of the main hypothesis, the three factors are positively correlated with an increasing trend of the age and with the male gender and the main ethnicity being British. I also expect positive correlation with a good state of health, a status of employed in professional and high qualified occupations and with higher educational attainment and with housing conditions where houses have more than 3 rooms (signalling a certain level of good economic wellbeing).

[21] The choice of the variables and their recoding has been done also according the availability in the Census for the SAE final work.

With regard to other characteristics, I expect different correlation related to the three factors: the number of hours spent caring may have a negative correlation with the factor about membership and citizenship and positive with neighbourliness, whereas providing care should present a positive sign with all of them, suggesting a natural propensity to relate to other people. Having dependent children can show a different sign respect to the three factors. I instead expect negative sign respect to the variable about moving.

Finally, ethnicity and religious aspects can show different sign according to the factors. The literature, indeed, reports conflicting results about how much diversity and religions hamper or trigger SC. Overall, as I will specify later on, I expect slightly different results for the third factor about neighbourliness with respect to the first two factors about membership and citizenship.

I will now present the building up of the MM and their results by factors. Descriptive statistics are available in Appendix B with the definitions of the survey variables used. I also report descriptive statistics of variables with respect to the geographical level, to see how much they vary within and between MSOAs (Table B2).

3.4 Factor 1 – Membership: Results

According to Table 3.4, the Null Model shows that 13 percent of the variance for this factor can be explained by differences between MSOA. In Social Sciences, there is no a general agreement about a threshold but we can consider this variability quite good given the type of phenomena and the very restricted area-level analysed.

With regard to the initial issue, if a Variance Component Model has to be used instead of a classical OLS, I can show graphical proofs supporting the use of MM. Indeed, as previously described, the Hausman test can be frequently misleading, taking the choice of a Fixed Effects Model. I report in Appendix B two graphs that are frequently used to show the real need of MM. Figure B1 shows the variation of Factor 1 between MSOA with the corresponding confidence interval whereas Figure B2 shows the MSOA effects in rank, together with their 95 percent confidence intervals. Both graphs support the idea that there is an actual geographical effect to take into consideration to explain the variation of levels of membership between individuals within different areas. Indeed, the variation between areas is quite high therefore, a Variance Component Model is feasible and probably better for a rich and more complete analysis.

3.4.1 The Random Intercept Model

Looking at the following table and according to equation 3.4, we can see the results for the RIM. As explained previously, the addition of individual covariates is included in the model to account for individual level variations in SC factors and also to control for selection bias. To obtain proper estimates of the impact of the area where individuals live, it is therefore necessary to control for potential differences in the individual composition of each area (Hox, 2010). Starting from the RIM, regarding Table 3.4 from this chapter, we can see that the ICC decreases from the 13 percent of the Null Model to the 8.1 percent: adding individual level covariates naturally reduces the unexplained variance because an important part of it is explained by individual covariates. The area contribution appears small when compared to the contribution of differences between individuals within MSOA. However, further unexplained differences can be referred to the area-level, still accounting for the 8 percent of the total residual variance.

If we look at the variance explained by the MSOA level, we can see that it is really low (0.03 percent) whereas is much higher the variance at individual level (0.4 percent).

Table 3.8: Random Intercept Model results – Factor 1

		Model 1: Random Intercept Model (S.E.)
FIXED EFFECTS		
Constant		-0.056*** (0.001)
Gender (Ref: Female)	Male	-0.013*** (0.001)
Age (Ref: Age 16-24)	Age 25-34	0.005** (0.002)
	Age 35-45	0.018*** (0.002)
	Age 46-55	0.026*** (0.002)
	Age 56-65	0.040*** (0.002)
	Age 66 and more	0.065*** (0.002)
Health (Ref: Poor health)	Good health	0.015*** (0.002)
	Fair health	0.006*** (0.002)
Marital Status (Ref: In a couple)	Single	-0.004*** (0.001)
Religion (Ref: Christians)	Muslim	0.027*** (0.002)
	Other religion	0.032*** (0.002)
Caring (Ref: Care someone)	Do not provide care	-0.010*** (0.002)
Caring 2 (Ref: Less than 20 hrs per week)	Care 20 hrs or more per week	-0.006*** (0.002)
Ethnicity (Ref: White)	Not White	0.038*** (0.001)
NS-SEC of Occupations (Ref: Lower Supervisory, Technical, Semi Routine and Routine, Never Worked and Long-Term Unemployed)	High and Low Managerial, Administrative and Professional	0.009*** (0.001)

	Intermediate, Small Employers, Own Account	0.001 (0.001)
Educational Level (Ref: No qualifications)	GSCE level	0.019*** (0.001)
	A level	0.026*** (0.002)
	Degree and higher	0.042*** (0.001)
	Other qualifications	0.016*** (0.002)
Employment Status (Ref: Unemployed)	Employed	-0.007*** (0.001)
Interactions	Employed * Male	0.006*** (0.002)
RANDOM EFFECTS		
MSOA level		0.0003*** (0.000)
Individual level		0.004*** (0.000)
Number of cases		27,744
ICC		0.081

*** P-value <(0.001) **P-value <(0.005) *P-value <(0.010)

Focusing on the covariates, all significant for the category[22], we can see that Factor 1 increases, as I expect, with the increase of the age, good health, higher educational attainment (as confirmed by CIS 15, 2015) and high socio-economic class of occupation, especially for non-White individuals.

It can seem that coefficients are small even if significant. But it is important to keep in mind that, how showed in the previous chapter (see Table 2.7 and Figure 2.7, Ch. 2) the range of this factor is really small: the minimum is -0.146 while the maximum value it takes is 0.297.

With regard to the age variable, older classes of ages show higher and higher coefficients with considerable changes of effects respect to their relative previous

[22] Only one dummy variable about the NS-SEC classification of the occupation is not significant.

cohort: being aged between 56 and 65 increases the average score of Factor1 by 0.04 point, showing the cumulative and long-term perspective dimension of SC.

I can say that it mirrors partially the existing literature. Classical studies say that political involvement increases with age (Milbrath and Goel 1977; van Deth, 1990) while other studies show that SC usually rises and falls with age and only rises with occupations with greater returns to SC and who invest in human capital too (Glaeser *et al*., 2002) (van Deth, 2000) (CIS 8, 2015).

About the education level, having a degree or higher attainments show the highest coefficient of the model: it increases the average level of Factor 1 by 0.04 point, almost half point in the score. All the other lower levels are significant too, showing decreasing but important coefficients too.

Being religious, on average, has a positive and increasing effect on the level of Factor 1 of SC. In this case, it is mandatory notice that this positive correlation is coherent with the hypothesis but that also may depend on the fact that the Factor created with the CFA includes few variables about the general attitudes and beliefs towards religion. These results confirm Putnam's intuition on religious dimension and the findings of van Oorschot et al. (2006) about the strong correlation of SC accumulation and religious beliefs and attitudes.

Referring to the ethnicity variables, generally the literature finds that membership is higher between non-White, showing in any case differences between the groups (CIS 15, 2015). CIS 8 (2015) finds that active membership of White people is average with respect to other ethnicities (higher than some groups and lower than others). Unfortunately, given the data, it was impossible to further disaggregate the analysis by groups (which seems to matter, given the different results for groups found in the cited study) but the overall effect for this factor is that general membership is higher than average in areas where diversity is higher too. In a following report (CIS 15, 2015) there are evidences about an increasing gap between the participation in voluntary organisations by ethnic background. Being a non-White British has a positive effect: increase the average score of Factor 1 by 0.038 point, one of the highest coefficients.

Unexpectedly, the being male decreases the mean score of Factor 1 by 0.013 point.
While I hypothesize that membership and especially active membership would have been higher between males (for several reasons: status, less time constraints than women, propensity to go out etc.), results are different. This can be justified by the types of organizations included. This hypothesis has been suggested by some of the current literature. There is controversial evidence about the relation between gender and SC, particularly about the membership dimension and related topics. In an interesting work about gender and social capital, several authors explain that since ever the lens of interpretation of SC phenomena, we suffer of a certain kind of 'gender inequalities blindness' that has two important consequences: 1) definitions of SC are more male oriented and 2) studies and evidence, consequently, do not report enough information on this relation. In this specific case, for example, membership has been frequently associated with the traditional political membership, typically more related to the male dimension. Instead, other organizations, here considered, show higher female membership: environmental, voluntary, caring, schooling, local,

neighbourliness and so on. Not only that, they also show that frequently women reach a position of leadership in these organizations, showing indeed a real, active membership (that is, accounted in the factor as a proper dimension, together with the number of associations to which the individuals take part) (O'Neill and Gidengil, 2013). In another qualitative and comparative study about membership in groups for natural resource management, the authors find that overall collaboration, solidarity and conflict resolution increase within groups where women are present and, much more, these mixed groups foster norms of reciprocity (Westermann *et al*., 2005).

A more recent study finds a negative, longitudinal trend between being male and SC accumulation (Helliwell, 2006). Stressing the different type of membership between men and women, Li (2015a) shows that women and older (middle-aged) are more involved in voluntary membership and less in 'commercial activities'. The author refers to a famous study of Warde *et al*. in the 1990s that shows that men prefer 'associational activities' like going to pubs, cinemas, night clubs, generally defined as 'going out'. In another study, he confirms that women are more willing to volunteer than men (Li *et al*, 2015).

If these studies may help to explain this identified inverse relation, there is no chance to give a definitive answer. Other studies show indeed that '…women participate less in associational life because they can't, because they won't, or because nobody asked them…' (Norris and Inglehart, 2003, p. 1), that political involvement shows higher levels for men than for women (Milbrath and Goel, 1977; van Deth, 1990) or that men have more power in the workplace (so contacts with men are more productive of good jobs) while women know more about health (so contacts with women have a greater positive effect on health) (Erickson, 2004), confirming again the gap between male positions and female ones. Van Oorschot *et al.* (2006) confirm that SC is highly gendered and controversially correlated with gender dimensions: European women tend to be more involved in family networking and more trust-worthy but they are less engaged in political activities and voluntary organisations than men are. They also trust less people and they bond more than men but bridge do less than they do.

In the same way, also being singles on average decreases the average level of SC for Factor 1. It can be fundamental, in this case, considering how it does affect the real personal network. According to the main hypothesis, I expect such a result because being single naturally decreases the size and the number of nodes of the network. Other studies, indeed, find a positive correlation between being married or living as married and negative correlation with the other forms of being alone (single, divorced, separated, widowed) (Helliwell, 2006). However, there are also studies, like Claibourn and Martin (2000) that find an inverse relation: being single may push individuals to look for higher involvement in associations and organizations or like Costa and Kahn (2001) that find that SC trends decline only among married women and not among single women. Therefore, as well as other variables, we can consider it an open question.

The other negative correlation is between the employment status and being a member (active or not). In an interesting paper, Newton (2001) tries to see how social trust affects membership and citizenship. He states that:

> …social trust does not correlate widely or strongly with the usual set of social, economic, and political variables (income, education, class, gender, age, race, left-right politics, employment status, membership of voluntary

> organizations), but there is a slight tendency for it to be found in some social types. It is more frequently expressed by the 'winners' in society, rather than the 'losers' — that is, it correlates positively, if weakly and patchily, with high income, high education, and high social status, and is more likely to be found in men and the middle-aged, and in those who say they are happy, satisfied with their jobs, and proud of their nation. It is not surprising that those who are doing well in society are more likely to express social trust than those doing less well (Newton, 2001, p. 204).

Therefore, a positive correlation between the employment status and membership, as well as citizenship and the so called 'political capital' (Newton, 2001) has been demonstrated. Glaeser *et al*. (2002) confirm that SC is positively linked with not only occupations but also higher level types of works. We can then state with a certain amount of confidence that there is a positive relation between being employed and SC, especially with regard to its membership dimension. CIS 8 (2015), shows that SC not only facilitates the access to (better) jobs but also that active membership is higher between employed people than unemployed or other inactive.

I then decide to investigate furtherly this aspect and I find a significant and positive correlation for the interaction between gender and employment: with regard to the average population, employed males have higher average score of membership and active membership (by 0.006 point). This result in a better fit than the hypothesis reported above about definitions of membership and gender inequalities. The status composed jointly by male gender and employed make the correlation positive and significant.

The last unexpected result is about the caring variables. If caring too many hours per week (the variable accounts for more than 20 hours per week) can be inversely related to being a member and/or active member because of the amount of remaining free time, it seems that this hypothesis is not valid for the specular aspect: not providing care at all. Many studies, indeed, show that informal caring – and health conditions' improvements - (typical of families, small communities and neighbours) depends positively on the level of SC of the reference group (Lomas, 1998; Hendryx *et al*., 2002; Kawachi *et al*., 1997; Perry *et al*., 2008). Probably, considerations about caring activities cannot be only reduced to the amount of time but also to a certain predisposition to spend time with other people. It may be the case that the types of organizations and associations considered load on different stock of SC (see par. 2.3.1).

To control the goodness of the model, graphical checks have been frequently recommended by the literature, particularly the plot of the level 1 residuals. Testing their Normality assumption is, indeed, one of the main steps. Therefore, graphs and plots of residuals of level are reported in Appendix B.

From Figure B3 A & B, we can see that residuals' distribution approximate to a normal distribution, confirming the goodness of the RIM. A further, less used graphical check is the plot of the residuals against the predicted values. According to Rabe-Hesketh and Skrondal (2012) and Dedrick at al. (2009), if it assumes a cloudy shape around the fitted line, then the model is well fitted. As we can see from Figure B4, it does have this shape.

3.4.2 The Contextual Effects Model

The further step is the extension of the RIM to assess the validity of contextual explanations that focus on the impact of the areas in which people live. This model, as anticipated, is the further step is the Contextual Effects Model (CEM hereafter), where the two Components created with the PCA are tested. Results are the following:

Table 3.9: Contextual Effects Model results – Factor 1

		Contextual Effects Model (S.E.)
FIXED EFFECTS		
Constant		-0.056*** (0.001)
Gender (Ref: Female)	Male	-0.012*** (0.001)
Age (Ref: Age 16-24)	Age 25-34	0.006** (0.002)
	Age 35-45	0.018*** (0.002)
	Age 46-55	0.026*** (0.002)
	Age 56-65	0.040*** (0.002)
	Age 66 and more	0.065*** (0.002)
Health (Ref: Poor health)	Good health	0.014*** (0.002)
	Fair health	0.006*** (0.002)
Marital Status (Ref: In a couple)	Single	-0.005*** (0.001)
Religion (Ref: Christians)	Muslim	0.026*** (0.002)
	Other religion	0.032*** (0.002)
Caring (Ref: Care someone)	Do not provide care	-0.010*** (0.001)
Caring 2 (Ref: Less than 20 hrs per week)	Care 20 hrs or more per week	-0.006** (0.002)
Ethnicity (Ref: White)	Not White	0.037*** (0.002)
NS-SEC of Occupations (Ref: Lower Supervisory, Technical, Semi Routine and Routine, Never Worked and Long-Term Unemployed)	High and Low Managerial, Administrative and Professional	0.008*** (0.001)
	Intermediate, Small Employers, Own Account	0.001 (0.001)

Educational Level (Ref: No qualifications)	GSCE level	0.018*** (0.001)
	A level	0.025*** (0.002)
	Degree and higher	0.040*** (0.001)
	Other qualifications	0.015*** (0.002)
Employment Status (Ref: Unemployed)	Employed	-0.007*** (0.001)
Interactions	Employed * Male	0.006*** (0.002)
CONTEXTUAL EFFECTS (Level 2)		
	Heterogeneity	-0.003*** (0.001)
	Economic Profile	0.005*** (0.001)
	Ethnicity Index	0.003*** (0.005)
	Heterogeneity* Economic Profile	-0.004*** (0.000)
	Heterogeneity* Ethnicity Index	0.056*** (0.000)
RANDOM EFFECTS		
MSOA level		0.003*** (0.000)
Individual level		0.040*** (0.000)
Number of cases		27,744
ICC		0.077

*** P-value <(0.001) **P-value <(0.005) *P-value <(0.010)

Looking at the Table 3.9 of results, we can do some considerations. While I do not register any important change with regard to the level 1 individual covariates (coefficients are more or less similar for all predictors and corresponding signs are confirmed), I have an ICC of 7.7 percent. Adding contextual effects helps explain the variance by less than 1 percent (0.4 percent). The fact that the variance explained by MSOA level and individual level are as the same as for the RIM confirms this hypothesis (0.03 percent and 0.4 percent).

Despite this low improvement of the model in explaining the variance at area-level, all the three variables and almost all their interactions are strongly significant but with really small coefficients. The highest one is the economic profile coefficient, 0.005 whereas the other two coefficients are: 0.003 for the ethnicity index and even a negative coefficient for the degree of heterogeneity of the areas, -0.003.

They suggest that people living in areas where professional and higher administrative occupations are more common as well as higher percentage of healthy and educated people and with a higher ethnic diversity, have a higher than average level of membership, net of all other covariates. But the negative sign for the degree of heterogeneity suggests also that a certain degree of homogeneity about characteristics like gender composition, marital status, religious belonging, and cohort of ages of inhabitants, is important too in order to have an higher than average level of membership SC.

Interactions explored to further understand the effect of these contextual components show that heterogeneous profile of areas show stronger effects than their economic profile, keeping the sigh of the relations negative. However, together with a higher degree of ethnic diversity, it turns to have a positive effect on the increase of Factor 1 by 0.056 point respect to the average level (shortly less than the negative relation does).

Interactions between the contextual effects and most significant variables have been tested but they were not found significant or even if significant, they showed really small coefficients.

Despite the estimated contextual model does not seem to add particular extra value to the investigation of the phenomena, the model seems to have a good fit. As for the RIM, graphs about the distribution of the residuals at level 1 are presented in the Appendix (Figure B5 A & B) together with the plot of the residuals versus the predicted values. Also in this case, the distribution is really close to a normal one, respecting the assumption for this kind of model and the plot shows a 'cloudy' shape, as from hypothesis.

3.5 Factor 2 – Citizenship and Politics: Results

Starting the analysis for this factor from the Null Model, the portion of variance for this factor that can be explained by differences between MSOAs is about 12 percent (Table 3.4). With the same procedure used for the previous factor, I investigate the effect of adding individual covariates then contextual effects and finally testing random coefficients to see how levels of Citizenship and Politics varies at MSOA levels. Figures B7 and B8 (in the Appendix B) once again, as for the first factor, confirm that Variance Component Models are more than suitable for this kind of analysis given the variance between MSOA of levels of citizenship and politics.

3.5.1 The Random Intercept Model

According to the following table of results, we can see that most of the results are quite similar to the Membership's results and, therefore, to the literature: sense of citizenships, trust toward the Institutions, voting behaviours are positively and significant correlated with the increase of the age, a good state of health, being educated and a good socio-economic position.

Table 3.10: Random Intercept Model results – Factor 2

		Model 1: Random Intercept Model (S.E.)
FIXED EFFECTS		
Constant		- 0.603***
		(0.025)
Gender	Male	0.088***
(Ref: Female)		(0.008)
Age	Age 25-34	0.037
(Ref: Age 16-24)		(0.016)
	Age 35-45	0.132***
		(0.016)
	Age 46-55	0.221***
		(0.016)
	Age 56-65	0.391***
		(0.016)
	Age 66 and more	0.543***
		(0.016)
Health	Good health	0.127***
(Ref: Poor health)		(0.015)
	Fair health	0.037
		(0.016)
Marital Status	Single	-0.038**
(Ref: In a couple)		(0.008)
Religion	Muslim	0.015
(Ref: Christians)		(0.018)
	Other religion	-0.070***
		(0.016)
Caring	Do not provide care	-0.056***
(Ref: Care someone)		(0.010)
Caring 2	Care 20 hrs or more per week	-0.075***
(Ref: Less than 20 hrs per week)		(0.017)
Ethnicity	Not White	0.002
(Ref: White)		(0.015)
NS-SEC of Occupations	High and Low Managerial, Administrative and Professional	0.205***
(Ref: Lower Supervisory, Technical, Semi Routine and Routine, Never Worked and Long-Term Unemployed)		(0.010)
	Intermediate, Small Employers, Own Account	0.063***
		(0.011)
Educational Level	GSCE level	0.167***
(Ref: No qualifications)		(0.012)
	A level	0.262***
		(0.013)
	Degree and higher	0.434***
		(0.012)
	Other qualifications	0.110***
		(0.014)
Employment Status	Employed	-0.033***
(Ref: Unemployed)		(0.010)
Number of Rooms in the House	3 or more rooms in the house	0.089***
(Ref: 1 or 2 Rooms)		(0.008)
Interactions	Single*Not White	0.067***

	(0.018)
RANDOM EFFECTS	
MSOA level	0.025*** (0.001)
Individual level	0.288*** (0.002)
Number of cases	27,700
ICC	0.080

*** P-value <(0.001) **P-value <(0.005) *P-value <(0.010)

We can also notice that, generally speaking, the coefficients are higher than the ones for the first factor, identifying then stronger average effects of the same variables for factor 2. This can be also explained by the wider range of score for this factor: minimum value of -2.129 and maximum value of 2.724 (see Table 2.7 and Figure 2.11, Ch. 2).

Age cohorts, as for Factor 1, show an increasing trend with really high coefficients for the oldest ones: being aged between 56 and 55 increases the Factor 2 average score of almost 0.4 point and being oldest (66 and more) increase the same factor's average score of even 0.54 point, the highest coefficient of the model. The only difference is that the first cohort, age 25 -34, is not significant in this case. This fact may be linked to the cumulative and long-term process of this type of SC.

In the same way, having a degree or higher educational attainments increases the mean score by 0.43 point and the trend increases as well starting from the lowest levels to the highest attainments.

The state of health seems having even an amplified effect respect to the case of Factor 1: being in a good state of health increase the average level of citizenship of an individual by 0.12 point.

Also in this case, the correlation between the levels of this particular dimension of SC and caring is negative, confirming the previous hypothesis about the different kind of stock of SC on which the two aspects can differently rely on.

Slightly different is the correlation about the religious aspect. If in the previous case, both being Muslim and belonging to religions other than Christianity and Islam were positively correlated with the average level of being a member or active member[23], in this case belonging or not to the more common religions seems to matter. Being Muslim, indeed, results not significant.

An unexpected result concerns ethnicity. While I show that the White cohort has low levels of membership, I supposed that for this factor the correlation would have been positive: political activities, institutional trust and voting attitudes and behaviours should depend more on being (White) British. This may be related to natural rights related to citizenship but also to a more formal type of involvement and capital – the political capital (Newton, 2001). We may assume that trust towards Institutions and the other variables related to it start to be higher between the other ethnicities. Immigrants from other countries or citizens not White (2nd Generation) can have higher trust because of different reasons related to this type of trust: social mobility,

[23] It is fundamental here to remember that Factor 1 also includes variables about religious attitudes and beliefs and that this aspect could have been the reason for the positive correlation.

employment opportunities, easier acquisition of citizenship and related rights (Bauböck, 2005). Instead, the variable does not result significant at all.
The ethnicity's significance changes with the interaction with the marital status. As well as for Factor 1, being single has a negative correlation on the average level of citizenship and politics, probably due to the more limited networks, but being single and not White shows a significant and positive correlation with the level of this type of SC. It suggests that people living in areas where there are more single and not White people have a higher than average level of citizenship and politics, net of all other covariates, by 0.067 point.

Following on in the analysis, as well as the first factor, the marital status shows a negative correlation but with a small coefficient: 0.038. As anticipated, it shows a positive correlation with the ethnicity. I also tested if interactions with the socio-economic positions would have resulted significant but they did not.

Concerning the employment status, as well for Factor 1, it shows a negative correlation with a small coefficient: 0.03, confirming the previous of idea that it may affect the real availability of time. I also tried to test this hypothesis focusing on the interaction between the employed status and the socio-economic classification of occupation, frequently depending on the educational level, but it does not result significant.

Definitively stronger than for Factor 1 is the effect of the socio-economic conditions: being a manager or high-classified professional increases the mean score of the Factor 2 of 0.2 point.

Finally, the two main differences with regard to the membership factor are due to gender and the number of rooms of the house. As expected, given the strong political trait of this factor, on average, being male increases the average level of citizenship and politics by almost 0.9 point. As well as gender, the new significant variable about number of rooms of the house higher than 3 – that can be interpreted like a kind of indicator of economic wellbeing – shows that people living in areas where the economic wellbeing is higher than the average have a higher than average level of citizenship and politics by 0.089.

Regarding the overall model, checking the ICC index, the variance explained decreases by 4 percent with respect to the Null Model, suggesting how important the individual covariates are in explaining differences of levels of citizenship and politics within areas. The variance explained at MSOA level is about 2.5 percent, whereas what is much higher is the variance explained at the individual level: 28 percent. With respect to Factor 1, we can see that variance explained by the two levels is really higher than in the previous case. This may confirm that this type of model is even more appropriate in explaining differences in citizenship and politics SC.

Besides, as shown in Figure B9 A & B and B10, residuals of level 1 approximate to a normal distribution and the plot of predicted values against residuals show the desirable cloudy shape for a good fitting of the model.

3.5.2 The Contextual Effects Model

As previously discussed, the further step in the analysis sees the test of contextual effects with area-level covariates. In this case, I use the components created with the PCA.

Table 3.11: Contextual Effects Model results – Factor 2

		Contextual Effects Model (S.E.)
FIXED EFFECTS		
Constant		0.574*** (0.025)
Gender (Ref: Female)	Male	0.090*** (0.008)
Age (Ref: Age 16-24)	Age 25-34	0.038 (0.015)
	Age 35-45	0.131*** (0.016)
	Age 46-55	0.219*** (0.015)
	Age 56-65	0.387*** (0.016)
	Age 66 and more	0.536*** (0.016)
Health (Ref: Poor health)	Good health	0.120*** (0.015)
	Fair health	0.033 (0.016)
Marital Status (Ref: In a couple)	Single	-0.039** (0.008)
Religion (Ref: Christians)	Muslim	0.021 (0.018)
	Other religion	-0.073*** (0.017)
Caring (Ref: Care someone)	Do not provide care	-0.058*** (0.010)
Caring 2 (Ref: Less than 20 hrs per week)	Care 20 hrs or more per week	-0.071*** (0.017)
Ethnicity (Ref: White)	Not White	-0.025 (0.016)
NS-SEC of Occupations (Ref: Lower Supervisory, Technical, Semi Routine and Routine, Never Worked and Long-Term Unemployed)	High and Low Managerial, Administrative and Professional	0.193*** (0.010)
	Intermediate, Small Employers, Own Account	0.057*** (0.011)
Educational Level (Ref: No qualifications)	GSCE level	0.162*** (0.012)
	A level	0.253***

		(0.013)
	Degree and higher	0.415***
		(0.012)
	Other qualifications	0.105***
		(0.014)
Employment Status	Employed	-0.034***
(Ref: Unemployed)		(0.010)
Number of Rooms in the House	3 or more rooms in the house	0.087***
(Ref: 1 or 2 Rooms)		(0.009)
Interactions	Single*Not White	0.062***
		(0.018)
CONTEXTUAL EFFECTS (Level 2)		
	Heterogeneity	0.018
		(0.009)
	Economic Profile	0.058***
		(0.004)
	Ethnicity Index	0.011
		(0.009)
RANDOM EFFECTS		
MSOA level		0.022***
		(0.001)
Individual level		0.288***
		(0.002)
Number of cases		27,699
ICC		0.071

*** P-value < (0.001) **P-value <(0.005) *P-value <(0.010)

The level 1 individual covariates confirm the same correlations with slightly different coefficients: the strongest effects on average level of Factor 2 come from age, education, economic condition (house with 3 rooms or more can be considered like a proxy variable), high socio-economic position and gender while religion and ethnicity are not significant or significant but with small coefficient. Health and caring have less important impact as well as the marital status.

The contextual variables are not particularly significant. Only the variables about the economic profile are positively significant. This means that an individual living in areas with those economic conditions (tertiary services, highly educated, etc.) would be expected to have a Citizenship and Politics SC score augmented of 0.058 point. The fact that the coefficient is 14 times its standard error confirms its strong significance.

The ethnicity index does not result significant, as well as the individual covariate about being not White. And also, the component about the heterogeneity profile of the areas is not significant. This aspect can lead to the hypothesis that this type of SC is slightly higher in homogeneous areas by ethnicity and other social conditions.

The inclusion of contextual effects reduces the ICC from 8 percent to 7.4 percent, similar to the first factor about membership where the decrease of the ICC was only of 0.4 percent between the RIM and the CEM. The decreasing variance explained at MSOA level (-0.6 percent) shows in any case that the addition of contextual effects is worthy for this factor (while variance explained at individual level is, as expected, still constant). The reduction in the area effect is around 2.2 percent.

Finally, also in this case, figures B11 A & B and B12 in Appendix B confirms the good fit of the model.

Enlarging our analysis, respect to Factor 1, we can note that for citizenship and politics SC the only contextual variable significant is the economic profile of the areas, whereas all the three variables were significant for Factor 1 but with a higher coefficient than the previous case: 0.003 versus the current 0.058.

3.6 Factor 3 – Neighbourliness: Results

The last dimension under analysis is the neighbourliness one, explained by the third factor created with the CFA. This factor loads on variables about the relations with the neighbours defined like interactions, trust, help, sense of belonging, number of close friends, use of social media, going out and feeling of safeness and intentions to move. I already hypothesized that results for this factor might have been different. We are focusing, indeed, on a dimension more informal, personal and intimate. Besides, this factor tries to catch the variance about more 'local' and less structured aspects than the previous two factors. The hypothesis has been confirmed. Figure B13 and B14 shows that also for this factor the use of a Variance Component Model may be appropriate given the difference between MSOAs. Therefore, also the ICC for the Null Model is around 13 percent as Table 3.4 reports.

3.6.1 The Random Intercept Model

Beginning to add the individual covariates as previously done for the other two factors, we can see from the following table the results of the RIM. As expected, it is quite clear that results are different with respect to the other two factors. Besides, given the particular structure of this factor, we can see that several 'new' variables have been resulted significant for this factor, whereas they were not considered for the other two factors because not significant at all, according to the hypothesis made at the beginning of the chapter (see par. 3.3.4).

Table 3.12: Random Intercept Model results – Factor 3

		Model 1: Random Intercept Model (S.E.)
FIXED EFFECTS		
Constant		-0.368***
		(0.004)
Gender	Male	-0.024
(Ref: Female)		(0.011)
Age	Age 25-34	0.119***
(Ref: Age 16-24)		(0.016)
	Age 35-45	0.231***
		(0.015)
	Age 46-55	0.270***
		(0.015)
	Age 56-65	0.379***
		(0.015)
	Age 66 and more	0.472***
		(0.016)
Health	Good health	0.183***
(Ref: Poor health)		(0.015)
	Fair health	0.079***
		(0.016)
Marital Status	Single	-0.061***
(Ref: In a couple)		(0.013)
Religion	Muslim	0.146***
(Ref: Christians)		(0.017)
	Other religion	0.025
		(0.016)
Caring	Do not provide care	-0.039***
(Ref: Care someone)		(0.009)
Caring 2	Care 20 hrs or more per week	-0.081***
(Ref: Less than 20 hrs per week)		(0.017)
Ethnicity	Not White	-0.073***
(Ref: White)		(0.014)
Educational Level	GSCE level	0.001
(Ref: No qualifications)		(0.012)
	A level	-0.012
		(0.013)
	Degree and higher	0.036***
		(0.009)
	Other qualifications	-0.006
		(0.014)
Employment Status	Employed	0.004
(Ref: Unemployed)		(0.011)
Local School Service	Intermediate	0.167***
(Ref: Poor)		(0.011)
	Good	0.277***
		(0.010)
Other local Services	Intermediate	0.079***
(Ref: Poor)		(0.011)
	Good	0.125***
		(0.011)
Moving	Moved	-0.053***

(Ref: Never moved in life/change address)		(0.012)
Number of Rooms in the House (Ref: 1 or 2 Rooms)	3 or more rooms in the house	0.061*** (0.008)
Children (Ref: Has not 1 or more dependent children under age 18)	Dependent children	0.049*** (0.010)
Interactions	Male * Single	0.057*** (0.015)
	Male * GSCE level	Not significant
	Male * A level	Not significant
	Male * Degree and higher	-0.045*** (0.014)
	Male * Other qualifications	Not significant
	Male * Not White	0.044*** (0.017)
	Employed * Single	-0.053*** (0.015)
RANDOM EFFECTS		
MSOA level		0.028*** (0.001)
Individual level		0.228*** (0.002)
Number of cases		23,228
ICC		0.110

*** P-value <(0.001) **P-value <(0.005) *P-value <(0.010)

Expected positive and significant correlations have been resulted for age – according to an incremental way (confirmed also by Li *et al*, 2015 that find that older people give more), health conditions and belonging to one religion, economic wellbeing (detected by the number of rooms of the house). Like in the previous cases, the oldest cohorts of age show important coefficients, particularly the 56-65 class of age: it increases the average level of Factor 3 by 0.58 point.

Remembering that Factor 3 range's score goes from -2.02 to 1.283 (see Table 2.7 and Figure 2.9, Ch. 2), it is a notable result.

The religious aspect that was found latterly not important seems still to matter. Indeed, Li *et al*. (2015) find a weak correlation of religious belonging with respect to the strong effect of social class and income in the propensity to donate, an aspect that we can consider linked and dependent on neighbourliness SC. Investigating further the ‘religious effect’ on its own, they find instead that no religious people donate less, so the Putnam intuition that religious networks are significant holds. Finally, they also show that between the groups, Muslims are those one donating more. This result has been confirmed also in this study: being Muslim increases mean score of Factor 3 by 0.14 point.

With regard to new variables resulting significant in a positive way and tested according to the hypothesis of possible significance, we can see that variables about evaluation of the level of local services such as school (primary and secondary), medical, shopping and leisure services have important coefficients. Individuals that have a good or really good evaluation of the level of local school services have a higher level of neighbourliness than the average by almost 0.28 point.

In addition, having dependent children is related to the average level of neighbourliness SC, reflecting its interaction with the local services and, probably, a higher level of commitment in the local dimension (even if for natural needs related to childhood). It increases the mean level of Factor 3 score by almost 0.05 point.
Indeed, Li *et al*. (2015) find that partnered people and people with dependent children have a higher propensity to give to others.

Expected negative correlations have been confirmed for the marital status: being single probably reduces the network available also at the local level despite of more free time being available, decreasing the average level of this SC factor by about 0.07 point. As already described, partnered people seem to be more willing also to donate and share (Li *et al.*, 2015). With regard to new variables, negative expected correlation is confirmed for the variable about moving: changing place naturally decreases the average level of neighbourliness SC of an individual by 0.052 point.

Other considerations can be made with respect to different variables taken into consideration. The first one is about gender. It seems that it does not affect the average level of SC on its own (it is not significant) but only if it interacts with the marital status, the ethnicity and the educational level (partially). This last interaction has been found especially significant in the report about the state of SC in Britain from CIS 15: while social support increases for some ethnic groups, it decreases for White men with no qualifications. Overall, women feel more supported than men (CIS 15, 2015). On the other hand, I also find that women have a higher propensity on dimensions like giving to others (possible related to this type of SC) (Li *et al.*, 2015). Therefore, unique and definitive explanations cannot be given to the gender issue.

With respect to the first one, it is interesting to notice that this interaction reports a positive sign, inverting then the initial sign for the marital status on its own. It may be then that marital status affects relatively more women in their local networks. Regarding the interaction between gender and educational level, we can see from the results that education level in itself is not significant apart for having a degree or higher educational attainments. According to CIS 8 (2015), there is not such a big difference in the availability of someone to talk with, whereas higher educational attainment seems to matter more for the level of generalized trust. Therefore, this result was partially expected: if on one side higher educational attainments make people more confident and satisfied and then sociable (according to the literature), on the other side this kind of factor - and variables considered - are more across-the board than the previous ones considered and it is not strongly related to all the dimensions (variables) upon which this factor loads. This may explain the weak significance. Finally, according to the results, the interaction suggests that people living in areas where graduated (and higher levels) are more common have lower than average level of neighbourliness by 0.045 point. This result can be explained by a possible lack of time available for profiles that may be probably intensively occupied. Indeed, the interaction between being single and employed shows a negative correlation, whereas we see that on its own, the employment status is not significant.

Indeed, CIS 8 (2015) shows that employment status has a strong effect about active participation, whereas differences between employees and unemployed people or other inactive people do not have important differences between them on the availability of someone to talk with and the percentage of assertive answers is quite high between the three typologies. The employment status matters more than how much people trust others. The overall effect also in this case can be weaker if considered on its own. The last two considerations are about two ‘unexpected’ results.

The first one regards the variables about caring. Both of the two variables about caring or not and related intensity show negative signs. If we can hypothesize that higher levels of involvement in caring activities reduces the time available invested in 'local' relations, I would expect a positive sign for the variable about caring, indicating, as previously suggested, a kind of propensity to interpersonal relations. In this case, I may suppose that caring defines a very different dimension from that which identifies from the variables about relations with neighbours. Besides, literature previously reported highlights how neighbours where relations are good show also higher levels of health wellbeing by its members, reducing the need for these kinds of activities. Last, the variety on variables on which this factor loads (social media and crime) can make this correlation weaker.

The second one is about ethnicity. I underline that according to CIS 8 and CIS 15 (2015) ethnicity has different effects, even ambiguous, on different dimensions of SC. If participation in voluntary organizations is slightly higher for the non-White, especially the active one (with differences between groups; CIS 8, 2015), trust towards neighbours is higher for White than for other ethnicities (CIS 8, 2015) and foreigners give less (Li *et al*., 2015). Other studies show how diversity can foster a higher level of SC but through complex processes. Cutts and Fieldhouse (2015), in a study comparing USA and UK using Structural Equation Modelling, find a negative direct relationship: as diversity increases, both community mindedness and community participation decrease, especially for the White majority group. However, if in more diverse communities, where they are more likely to have inter-ethnic friendships and engage in community participation, Whites have cross-groups friendships, then diversity fosters SC. So, generally speaking, diversity is negatively related to SC especially for Whites. Moreover, this aspect persists even if they control for individual and neighbourhood characteristics but it decreases with cross-groups and inter-ethnic friendships that increase generalized trust.

Therefore, keeping in mind that the relation about diversity and generalized trust and neighbourliness SC is still under analysis, for only this factor, people living in areas where diversity is higher seem to have lower than average levels of neighbourliness SC by 0.073 point. The sign of this correlation is totally inverted if interactions with the gender are taken into consideration. The estimated coefficient of 0.044 suggests that, again, being a non -White male increase the average level of SC in relation to the non-White women for the same area.

Looking at indicators for the overall model, we see from Figure B15 A & B and Figure B16 in the Appendix that the model presents a good goodness of fit. Besides, the ICC for this model says that, after the addition of individual covariates, the area effect still explains 11 percent of the phenomena, decreasing by 2 percent respect to the Null Model.

The variance explained at MSOA level is about 2.8 percent, the highest respect to the other two factors. Still higher but less than Factor 2 is the variance explained at individual level: 22.8 percent.

3.6.2 The Contextual Effects Model

After controlling for all the individual covariates, trying to investigate further the area effects, I add the contextual components created with the PCA. With respect to the individual variables, ethnicity becomes not significant in this contextual model. The possible explanation may be related to the addition of the ecological variables, as we are going to see shortly.

The ICC decreases from 11 percent to 9.3 percent, recording the highest improvement between the three CEMs from their RIMs. With it, also the variance explained at MSOA level decreases of 0.5 percent (while individual variance is constant at 22.8 percent). The hypothesis that for this kind of factor, those local characteristics are more important than the other two factors seems to be confirmed. Generally, as usual, checks for normality of residuals and other assumptions are reported in Appendix B (B17 A & B and B18).

Table 3.13: Contextual Effects Model results – Factor 3

		Contextual Effects Model (S.E.)
FIXED EFFECTS		
Constant		-0.403*** (0.004)
Gender (Ref: Female)	Male	-0.021 (0.011)
Age (Ref: Age 16-24)	Age 25-34	0.127*** (0.016)
	Age 35-45	0.231*** (0.015)
	Age 46-55	0.268*** (0.015)
	Age 56-65	0.371*** (0.015)
	Age 66 and more	0.454*** (0.016)
Health (Ref: Poor health)	Good health	0.169*** (0.015)

	Fair health	0.073***
		(0.016)
Marital Status	Single	-0.049***
(Ref: In a couple)		(0.013)
Religion	Muslim	0.173***
(Ref: Christians)		(0.017)
	Other religion	0.037
		(0.016)
Caring	Do not provide care	-0.040***
(Ref: Care someone)		(0.009)
Caring 2	Care 20 hrs or more per week	-0.072***
(Ref: Less than 20 hrs per week)		(0.017)
Ethnicity	Not White	-0.022
(Ref: White)		(0.015)
Educational Level	GSCE level	-0.012
(Ref: No qualifications)		(0.013)
	A level	0.001
		(0.012)
	Degree and higher	0.028**
		(0.009)
	Other qualifications	-0.006
		(0.014)
Employment Status	Employed	0.017
(Ref: Unemployed)		(0.010)
Local School Service	Intermediate	0.167***
(Ref: Poor)		(0.011)
	Good	0.277***
		(0.011)
Other local Services	Intermediate	0.079***
(Ref: Poor)		(0.011)
	Good	0.125***
		(0.011)
Moving	Moved	-0.053***
(Ref: Never moved in life/change address)		(0.012)
Number of Rooms in the House	3 or more rooms in the house	0.037***
(Ref: 1 or 2 Rooms)		(0.008)
Interactions	Male * Single	0.057***
		(0.015)
	Male * GSCE level	Not significant
	Male * A level	Not significant
	Male * Degree and higher	-0.045***
		(0.014)
	Male * Other qualifications	Not significant
	Male * Not White	0.044*
		(0.017)
	Employed * Single	-0.056***
		(0.015)
CONTEXTUAL EFFECTS (Level 2)		
	Heterogeneity	-0.071***
		(0.009)
	Economic Profile	0.051***
		(0.009)
	Ethnicity Index	-0.037***
		(0.009)

Interactions	Heterogeneity * Economic profile	-0.021*** (0.004)
	Heterogeneity * Ethnicity Index	0.047*** (0.005)
RANDOM EFFECTS		
MSOA level		0.023*** (0.001)
Individual level		0.228*** (0.002)
Number of cases		23,227
ICC		0.093

*** P-value <(0.001) **P-value <(0.005) *P-value <(0.010)

Ending the analysis with the comparison with the other two factors, we can see that the heterogeneity component is significant as well as for Factor 1 with a higher coefficient but the same negative sign of correlation. In addition, the ecological variable about the economic profile of the areas is significant, the same as for Factor 1 and Factor 2. Its coefficient is much higher than for Factor 1 and almost close to Factor 2.

The Ethnicity Index is significant with the smallest coefficient and a negative correlation: living in areas with high degree of ethnic diversity decreases the average level of Factor 3 by 0.037 point. However, its interaction with the degree of heterogeneity of the areas shows positive sign. Interestingly, it is also notable that the interaction between the degree of heterogeneity of areas and economic profile is negative (but with a small variance with respect to all the other ones). We can say that people living in homogeneous areas with a high economic profile (professional and tertiary) have a higher level of neighbourliness SC than the average.

3.7 Conclusions

The complex building up of the models demonstrates interesting results about which individual characteristics and areas characteristics influences the three different SC factors.

I try to summarize the wide amount of results from all the previous models for all the factors in the following tables.

Table 3.14: Variance explained by the levels of analysis - summary

	Level of Variance explained %	Factor 1 Membership	Factor 2 Citizenship and Politics	Factor 3 Neighbourliness
Random Intercept Model	MSOA	0.03	2.5	2.8
	Individual	0.4	28.8	22.8
Contextual Effects Model	MSOA	0.03	2.2	2.3
	Individual	0.4	28.8	22.8

Table 3.15: Models and factors by ICC – summary

Model	Factor 1 Membership	Factor 2 Citizenship and Politics	Factor 3 Neighbourliness
Baseline	13%	12%	13%
Random Intercept Model	8.1%	8%	11%
Contextual Effects Model	7.7%	7.1%	9.3%

In the first Table 3.14, I show the variance explained by the two levels considered in the model while in Table 3.15 I summarize the ICC for the all models.

About Factor 1, we can see that the level of variance explained at the two levels is really small: 0.03% at MSOA level and 0.4% at individual level. This last level explains variance 10 times more than the area effects. Adding contextual effects, besides, do not change the variance explained at all at both the levels. This means that the unmeasured processes generating the error terms in the random part of the model are catch already. Indeed, the variance of residuals estimated by the model is really low. It may suggest that a single level model would have been more appropriate for this factor.

But if we look at the ICC from Table 3.15, Factor 1 shows the highest change from the Baseline model to the RIM, about 5%, so it may suggest that this factor shows a mixed dependence on both individual and area effects. Indeed, in RIM the ICC is around 8%. It means that the total variance of Factor 1 is accounted by clustering.

Completely different is the situation for the other two factors. As we can from Table 3.14, the variance explained is higher than the previous factor, about 2%, at area level. Variance explained at individual level is still 10 times higher than the area effects. The change between the two models is small, 0.3% for Factor 2 and 0.5% for Factor 3, so it means that adding contextual effects do not help significantly the explanation of residual variance.

Therefore, it is useful to look at the ICC to understand better if Mixed Models were useful. Factor 2 and Factor 3 shows more similar trends about the two levels considered. Factor 3, according to hypothesis, confirms to be the factor more related to local aspects: MSOAs differences explain the highest total variance and the changes between ICC from the RIMs between the main models, showing that higher variance is explained with area effects.

Factor 2, as confirmed by the analysis and the almost null changes of ICC between models, confirms its high correlation with individual covariates, that explain most of the differences in Citizenship and Politics SC levels, like the highest percentage of variance recorded for the individual level by itself.

Focusing on the levels of analysis, I can start summarising results about individual covariates. Almost the totality of them is significant at 0.001, some at 0.005 or not significant at all. I report the final coefficient estimates at level 1 of RIMs. I can state indeed there I do not register big changes at this level during the building up of the models with contextual effects.

Table 3.16: Individual covariates by correlation and its sign at level 1– summary

Characteristic	Membership	Citizenship and Politics	Neighbourliness
Male	-	+	Not significant
Age	+	+	+
Good health	+	+	+
Fair health	+	Not significant	+
Single	-	-	-
Muslim	+	Not significant	+
Other religions	+	-	Not significant
Providing care	-	-	-
Caring more than 20 hrs per week	-	-	-
Not White	+	Not significant	-
High and Low Managerial, Administrative and Professional	+	+	Not significant
Intermediate, Small Employers, Own Account	Not significant	+	Not significant
GSCE level	+	+	Not significant
A level	+	+	Not significant
Degree and higher qualification	+	+	+
Other qualifications	+	+	Not significant
Employed	-	-	Not significant
Born out of UK	Not significant	Not significant	Not significant
Having dependent children	Not significant	Not significant	+
House with 3 or more rooms	Not significant	+	+
Moved in the life	Not significant	Not significant	-
Intermediate level of local school service	Not significant	Not significant	+
Good level of local school service	Not significant	Not significant	+

Intermediate level of other local services	Not significant	Not significant	+
Good level of other local services	Not significant	Not significant	+

Interactions not reported in the table register a positive significant correlation between being male and employed and average level of Membership SC. Being single and not White is the only interaction positively correlated with average level of Factor 2.

Finally, given the unexpected not significance of the gender variable for Factor 3, interactions with the marital status, the ethnicity and the degree level make it significant. In addition, the employment status becomes significant in its interactions with the marital status.

After checking the significance of the individual covariates, we can explore the results for contextual effects in the following table about contextual variables only.

Factor 1 model has all contextual effect as significant as well as for Factor 3. The importance of these area effects for the factor more 'local' among all of them is confirmed also from the fact that Neighbourliness' average levels increase also with interactions of contextual effects.

Factor 2, on the other side, was already shown to be more related to individual covariates than the other two factors. This may explain the non-significance of two out of three components and relative interactions.

Table 3.17: Variance explained by the contextual variables in models considered

Contextual effect	Factor 1 Membership	Factor 2 Citizenship and Politics	Factor 3 Neighbourliness
Contextual Effects Models			
Heterogeneity	-0.003	Not significant	-0.071
Economic profile	+0.005	+0.058	+0.051
Ethnicity Index	+0.003	Not significant	-0.037
Heterogeneity*Ethnicity Index	Not significant	Not significant	+0.047
Heterogeneity*Economic profile	Not significant	Not significant	-0.019

All these results show the complexity but at the same time the richness resulting from all this building up of MM. Initial hypotheses are almost confirmed, both at individual level and at area-level. Contextual effects components show to be important in the analysis, especially the economic profile: an individual living in an area with more diffused tertiary sector jobs, educated people, high profile socio-economic classification positions and healthy are expected to have higher SC score than the average level of the area.

A final thought should be done respect to the fact that Random Coefficient Models have not resulted significant in this study. It may be the case that despite MM is appropriate for studying SC and its differences due to geographical differences, individual covariates seem to be a stronger effect and they do not make SC differs between areas. Indeed, the main hypothesis of RCM is that slopes differ for all the areas. If RIM and CEM focus on the within-effects, RCM focus more on the between-effects. Mainly, according to RCM, it is possible to state that an individual with a certain amount of SC related to a particular combination of individual characteristics will have different SC from an individual with the same characteristics but in a totally different and geographically diverse area.

This result may be due to the level of the analysis chosen (MSOA level) or related statistical issues (number of observations per area and number of areas available). Nevertheless, it may be also due to the nature of this kind of capital, at the boundary of two dimensions: individual and local, local and global.

Indeed, MSOAs differences explain variance in average level of membership, citizenship, politics and neighbourliness. At the same time, SC remains related to personal characteristics: the right balance between a single person and their networks' dimensions, both in a more formal way (for the first two factors) and in an informal and local way (as for the third factor).

Appendix B

Table B1: Description of survey variables from UKHLS, year 2011 (reference category is not defined)[24]

Category and Name of Variable		Definition	Type of variable
Gender (Ref: Female)	Male	Gender	Dummy
Age (Ref: Age 16-24)	25-34	Classes of ages from the original continuous v.	Dummy
	35-45		
	46-55		
	56-65		
	66 and more		
Health (Ref: Poor health)	Fair health	Fair state of health from the original ordinal v.	Dummy
	Good health	Sum of excellent, very good and good state of health from the original ordinal v.	
Marital Status (Ref: Living in a couple)	Single	Sum of single, never married or in a legal civil partnership, divorced, widowed, separated, surviving partner, ex-civil partner from the original categorical v.	Dummy
Religion (Ref: Christians)	Muslim	Muslim from the original categorical v.	Dummy
	Other religions	Sum of Hindu, Jewish, Sikh, Buddhist and other religions from the original categorical v.	
Caring (Ref: Care someone)	Do not provide care	Sum of two original categorical vs.: looking after or giving special helps to someone who is sick, disabled or elderly living with OR not living with	Dummy
Caring 2 (Ref: Less than 20 hrs per week)	Care 20 hrs or more per week	Sum of 20-34/35-49/50-99/100 more hrs per week spent in unpaid caring someone from the original ordinal v.	Dummy
Ethnicity (Ref: White)	Not White	Sum of **Black**=Caribbean+ African+Any other Black background+ **Asian**=Indian+Pakistani+Bangladeshi+Chinese+Any other Asian background+ **Mixed**=White and Black Caribbean+White and Black African+White and Asian+Any other Mixed Background+ **Any other ethnic group**=Arabic+Any other ethnic group	Dummy
Educational Level (Ref: No qualifications)	GSCE level	Sum of GCE O Levels or Equivalent+CSE Grade 2-5 from the original ordinal v.	Dummy
	A level	GCE A Levels from the original ordinal v.	
	Degree and higher	Sum of Higher Degree+First Degree+Other Higher QF from the original ordinal v.	

[24] Definitions of variables and answered available between UKHLS and BHPS are pretty similar. More completed information will be reported when they will be different or one will result more detailed respect to the others. In any case, the recoding have been done to reflect completely the Census definitions, used for PCA and for the following work of SAE.

	Other qualifications	Sum of Teaching QF+Nursing QF+Commercial QF+No O+Apprenticeship+Other QF from the original ordinal v.	
NS-SEC of Occupations (Ref: Lower Supervisory, Technical, Semi Routine and Routine, Never Worked and Long-Term Unemployed)	High and Low Managerial, Administrative and Professional	Sum of NS-SEC category 1-2: 1. Higher Managerial, Administrative and Professional Occupations and 2. Lower Managerial, Administrative and Professional Occupations from the original ordinal v.	Dummy
	Intermediate, Small Employers, Own Account	Sum of NS-SEC category 3-4: 3. Intermediate Occupations and 4. Small Employers and Own Account Workers from the original ordinal v.	
Employment Status (Ref: Unemployed)	Employed	Sum of Paid Employed and Self Employed from the original categorical v.	Dummy
Number or Rooms in the House (Ref: 1 or 2 Rooms)	3 or more rooms in the house	Sum of classes 3, 4, 5 or more rooms derived by the original continuous v.	Dummy
Local School Service (Ref: Poor)	Intermediate	Fair evaluation of local services from two original ordinal v.: standard of local services of primary school and secondary school.	Dummy
	Good	Excellent or Very good evaluation of local services from two original ordinal v.: standard of local services of primary school and secondary school.	
Other local Services (Ref: Poor)	Intermediate	Fair evaluation of local services from three original ordinal v.: standard of local medical, shopping and leisure services.	Dummy
	Good	Excellent or Very good evaluation of local services from three original ordinal v.: standard of local medical, shopping and leisure services.	
Moving (Ref: Never moved in life/change address)	Moved	Derived from the original dummy v. asking if ever moved in his/her life or ever changed address.	Dummy
Children (Ref: Has not 1 or more dependent children under age 18)	Dependent children	Derived from the original continuous v. about the number of own dependent children in the HH crossed with the corresponding age.	Dummy

Table B2: Descriptive statistic for variables UKHLS (within between)

Variable		Mean	St. dev.	Min	Max	Observations
Male	overall	0.000	0.499	-0.462	0.538	N = 37932
	between		0.199	-0.462	0.538	n = 5255
	within		0.482	-0.857	0.889	T-bar = 7.21827
Age 25-34	overall	0.000	0.360	-0.153	0.847	N = 37932
	between		0.238	-0.153	0.847	n = 5255
	within		0.317	-0.900	0.968	T-bar = 7.21827
Age 35-45	overall	0.000	0.404	-0.205	0.795	N = 37932
	between		0.238	-0.205	0.795	n = 5255
	within		0.362	-0.889	0.964	T-bar = 7.21827
Age 46-55	overall	0.000	0.382	-0.177	0.823	N = 37932
	between		0.207	-0.177	0.823	n = 5255
	within		0.350	-0.857	0.968	T-bar = 7.21827
Age 56-65	overall	0.000	0.357	-0.150	0.850	N = 37932
	between		0.205	-0.150	0.850	n = 5255
	within		0.320	-0.857	0.968	T-bar = 7.21827
Age 66 and more	overall	0.000	0.383	-0.178	0.822	N = 37932
	between		0.246	-0.178	0.822	n = 5255
	within		0.335	-0.875	0.969	T-bar = 7.21827
Not White	overall	0.000	0.416	-0.223	0.777	N = 37448
	between		0.314	-0.223	0.777	n = 5255
	within		0.274	-0.967	0.973	T-bar = 7.12617
Good health	overall	0.000	0.408	-0.789	0.211	N = 37895
	between		0.231	-0.789	0.211	n = 5253
	within		0.371	-0.966	0.857	T-bar = 7.21397
Fair health	overall	0.000	0.355	-0.148	0.852	N = 37895
	between		0.196	-0.148	0.852	n = 5253
	within		0.326	-0.800	0.969	T-bar = 7.21397
Muslim	overall	0.000	0.282	-0.087	0.913	N = 32125
	between		0.188	-0.087	0.913	n = 5224
	within		0.178	-0.967	0.971	T-bar = 6.1495
Other religions	overall	0.000	0.240	-0.062	0.938	N = 32033
	between		0.168	-0.062	0.938	n = 5223
	within		0.191	-0.875	0.971	T-bar = 6.13307
Single	overall	0.000	0.481	-0.362	0.638	N = 37932
	between		0.305	-0.362	0.638	n = 5255
	within		0.427	-0.917	0.944	T-bar = 7.21827
Employed	overall	0.000	0.482	-0.633	0.367	N = 37912
	between		0.284	-0.633	0.367	n = 5254
	within		0.431	-0.950	0.923	T-bar = 7.21584
GSCE level	overall	0.000	0.405	-0.207	0.793	N = 34653
	between		0.225	-0.207	0.793	n = 5198

	within		0.368	-0.833	0.955	T-bar = 6.6666
A level	overall	0.000	0.405	-0.207	0.793	N = 34653
	between		0.221	-0.207	0.793	n = 5198
	within		0.371	-0.889	0.955	T-bar = 6.6666
Degree and higher	overall	0.000	0.477	-0.350	0.650	N = 34653
	between		0.303	-0.350	0.650	n = 5198
	within		0.414	-0.917	0.955	T-bar = 6.6666
Other qualifications	overall	0.000	0.307	-0.106	0.894	N = 34653
	between		0.171	-0.106	0.894	n = 5198
	within		0.281	-0.800	0.983	T-bar = 6.6666
High and Low Managerial, Administrative and Professional	overall	0.000	0.429	-0.243	0.757	N = 37751
	between		0.260	-0.243	0.757	n = 5252
	within		0.381	-0.900	0.966	T-bar = 7.18793
Intermediate, Small Employers, Own Account	overall	0.000	0.350	-0.143	0.857	N = 37751
	between		0.183	-0.143	0.857	n = 5252
	within		0.322	-0.800	0.967	T-bar = 7.18793
Do not provide care	overall	0.000	0.391	-0.811	0.189	N = 34699
	between		0.222	-0.811	0.189	n = 5255
	within		0.351	-0.966	0.857	T-bar = 6.60304
Care 20 hrs or more per week	overall	0.000	0.219	-0.050	0.950	N = 34699
	between		0.119	-0.050	0.950	n = 5255
	within		0.200	-0.667	0.969	T-bar = 6.60304
Dependent children	overall	0.000	0.403	-0.205	0.795	N = 35454
	between		0.216	-0.205	0.795	n = 5255
	within		0.374	-0.800	0.958	T-bar = 6.74672
3 or more rooms in the house	overall	0.000	0.420	-0.228	0.772	N = 37846
	between		0.294	-0.228	0.772	n = 5250
	within		0.338	-0.923	0.967	T-bar = 7.20876
Moved	overall	0.000	0.295	-0.904	0.096	N = 37070
	between		0.171	-0.904	0.096	n = 5242
	within		0.252	-0.967	0.900	T-bar = 7.07173
Local school service - Intermediate	overall	0.000	0.459	-0.302	0.698	N = 29285
	between		0.293	-0.302	0.698	n = 5032
	within		0.400	-0.875	0.967	T-bar = 5.81975
Local school service - Good	overall	0.000	0.497	-0.559	0.441	N = 29285
	between		0.327	-0.559	0.441	n = 5032
	within		0.421	-0.967	0.933	T-bar = 5.81975
Other local service - intermediate	overall	0.000	0.498	-0.454	0.546	N = 29285
	between		0.316	-0.454	0.546	n = 5032
	within		0.433	-0.929	0.947	T-bar = 5.81975
Other local service - Good	overall	0.000	0.495	-0.430	0.570	N = 29285
	between		0.331	-0.430	0.570	n = 5032
	within		0.416	-0.947	0.955	T-bar = 5.81975

Table B3: Descriptions of variables used in PCA, year 2011

Measures	Definition	Source
Born in UK	Percentage of people born in UK (England, Wales, Scotland, Northern Ireland) per MSOA (over total residents)	Census
Born in Europe	Percentage of people born in European countries per MSOA (over total residents)	Census
1 or 2 rooms house	Percentage of HH living in a house with 1 or 2 rooms per MSOA (over total number of HH)	Census
House owned	Percentage of HH with ownership of house per MSOA (over total number of HH)	Census
Terraced or flat house	Percentage of HH living in a terraced house, flat, maisonette or apartment per MSOA (over total number of HH)	Census
Resident since 5 year or more	Percentage of people resident since more than 5 years per MSOA (over total residents)	Census
Christian	Percentage of Christian people per MSOA (over total residents)	Census
Muslim	Percentage of Muslim people per MSOA (over total residents)	Census
Population density	Number of Persons per Hectare per MSOA (rate)	Neighbourhood Statistics
Single	Percentage of single, never married or in a legal civil partnership, divorced, widowed, separated, surviving partner, ex-civil partner people per MSOA (over total residents)	Census
Do not provide care	Percentage of people not providing unpaid care per MSOA (over total residents)	Census
Age 30-44	Percentage of people ageing 30 to 44 per MSOA (over total residents)	Census
Violence against the person	Number of notifiable offences recorded by the Police of: Wounding or Other Act Endangering Life, Other Wounding, Harassment Including Penalty Notices for Disorder, Common Assault, Robbery or Theft from the Person per Local Authority Districts (LAD)	Neighbourhood Statistics
Criminal damage	Number of notifiable offences recorded by the Police of: Criminal Damage Including Arson, Burglary in a Dwelling, Burglary Other than a Dwelling, Theft of a Motor Vehicle or Theft from a Motor Vehicle per Local Authority Districts (LAD)	Neighbourhood Statistics
Male	Percentage of Male people per MSOA (over total residents)	Census
High and Low Managerial, Administrative and Professional and Intermediate, Small Employers, Own Account (NS-SEC classif.)	Percentage of people employed according to NS-SEC 1, 2, 3 or 4 classification per MSOA (over total residents)	Census
Tertiary sector worker (SIC 2007 classif.)	Percentage of people employed in the Tertiary Sectors I, J, K, L, M, N, O, P, Q, R, S, T or U according	Census

	to the SIC 2007 classification per MSOA (over total residents)	
Good health	Percentage of people with very good and good health per MSOA (over total residents)	Census
A levels	Percentage of people with 2+ A Levels/VCEs, 4+ As Levels, Higher School Certificate, Progression/Advanced Diploma or Welsh Baccalaureate Advanced Diploma per MSOA (over total residents)	Census
Degree and higher qualifications	Percentage of people with Degree (For Example BA, BSc), Higher Degree (For Example MA, PhD, PGCE) NVQ Level 4-5, HNC, HND, RSA Higher Diploma or BTEC Higher Level per MSOA (over total residents)	Census
Employed	Percentage of paid employed and self-employed people per MSOA (over total residents)	Census
Benefits for disability	Percentage of Disability Living Allowance or Incapacity Benefit/Severe Disablement Allowance claimants per MSOA (over total residents)	Neighbourhood Statistics
Benefits for economic disadvantages	Percentage of Income Support, Jobseekers Allowance or Pension Credit claimants per MSOA (over total residents)	Neighbourhood Statistics
Tax credit	Percentage of Families receiving tax credit and Lone parent families in work receiving tax credit per MSOA (over total residents)	Neighbourhood Statistics
Distance travelled to work: less than 10 km (WTA classif.)	Percentage of people travelling daily to work less than 10 km per MSOA (over total residents)	Census

Table B4: Descriptive statistics of Census and Neighbourhood Statistics variables

Measures	Mean (%)	Minimum (%)	Maximum (%)	Standard Deviation
Born in UK	87.31	13.10	30.77	98.91
Born in Europe	6.80	7.45	0.45	62.18
1 or 2 rooms house	4.52	4.48	0.33	36.47
House owned	64.28	17.14	7.28	96.49
Terraced or flat house	44.22	24.38	1.75	99.00
Resident since 5 year or more	9.10	9.14	0.88	52.72
Christian	59.86	11.47	6.32	85.21
Muslim	4.42	9.39	0.00	83.37
Population density	32.18	34.31	0.10	247.20
Single	52.80	10.12	30.17	94.93
Do not provide care	89.55	1.96	83.63	97.49
Age 30-44	20.38	3.83	5.17	38.26
Violence against the person	64.60	43.72	6.17	377.33
Criminal damage	94.75	72.10	4.21	547.13
Male	49.12	1.52	44.23	61.28
High and Low Managerial, Administrative and Professional and Intermediate, Small Employers, Own Account (NS-SEC classif.)	39.61	10.62	7.04	76.79
Tertiary sector worker	34.93	7.50	15.36	68.44
Good health	81.08	4.48	61.06	95.39
A levels	18.03	10.14	2.68	72.17
Degree and higher qualifications	21.30	10.69	3.41	68.16
Employed	56.02	7.31	9.77	80.91
Benefits for disability	20.21	9.86	1.57	72.88
Benefits for economic disadvantages	22.85	11.63	2.06	94.13
Tax credit	26.36	10.52	1.79	97.68
Distance travelled to work (WTA classif.)	29.56	6.05	11.07	60.38

Figure B1: Variation of Factor 1 between MSOA

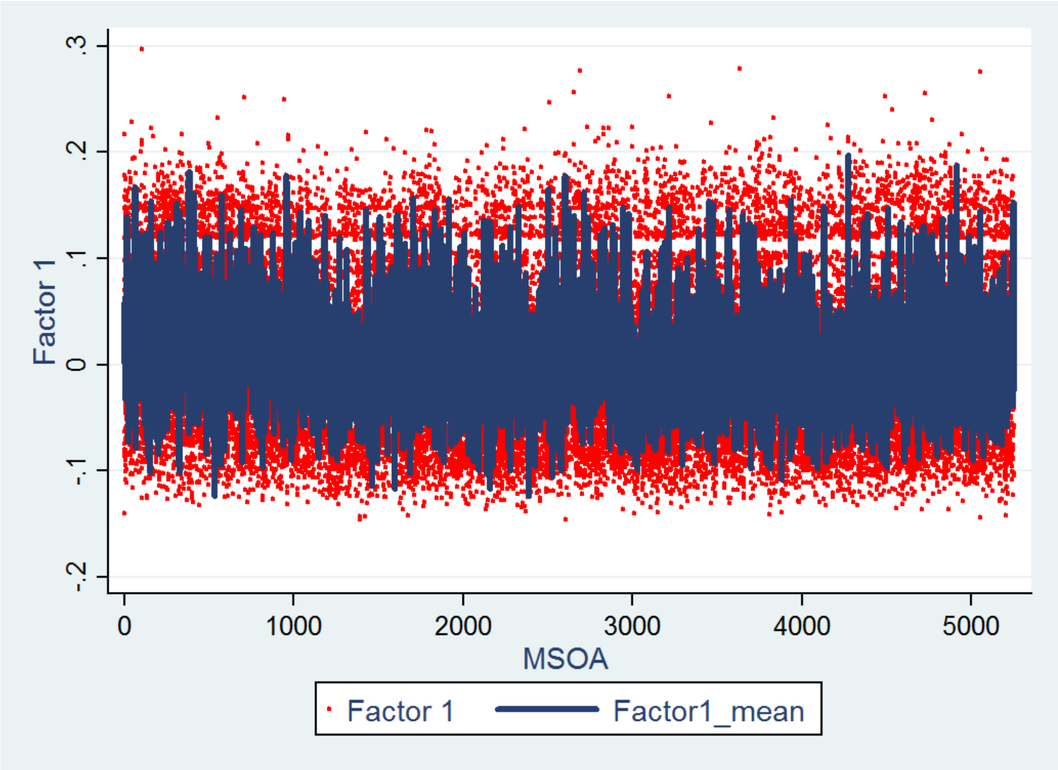


Figure B2: MSOA effects in rank for Factor 1, Null Model

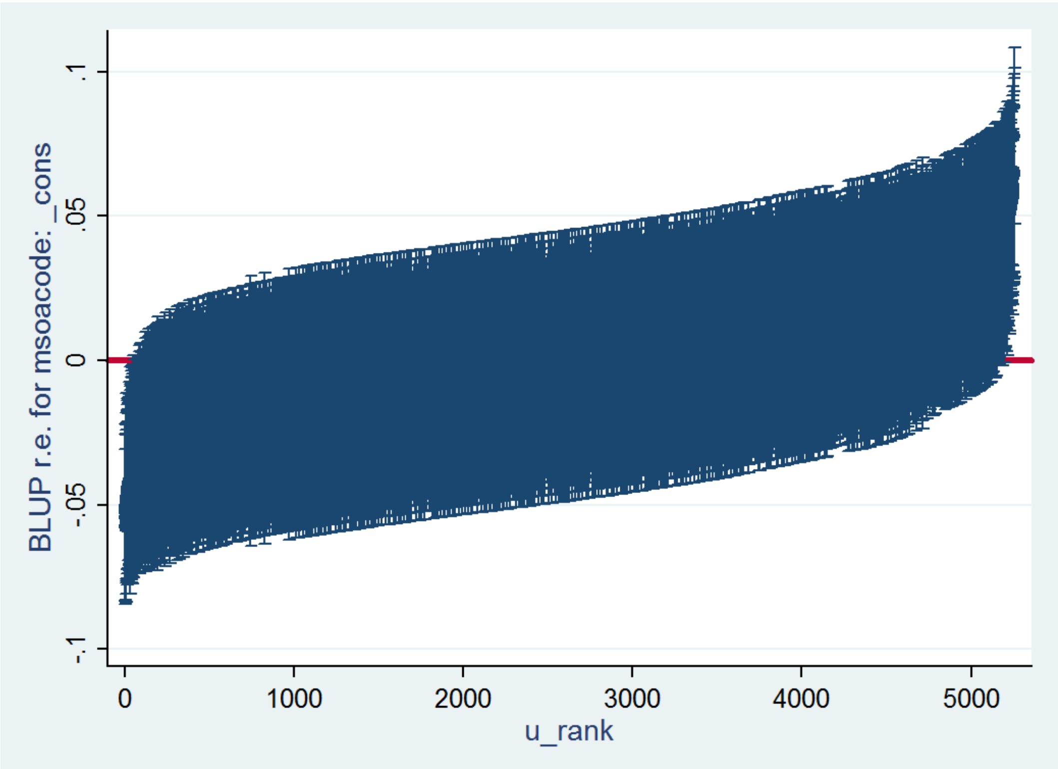


Figure B3 A & B: Factor 1 – Diagnostic plots: Residuals distributions of Level 1 for Random Intercept Model

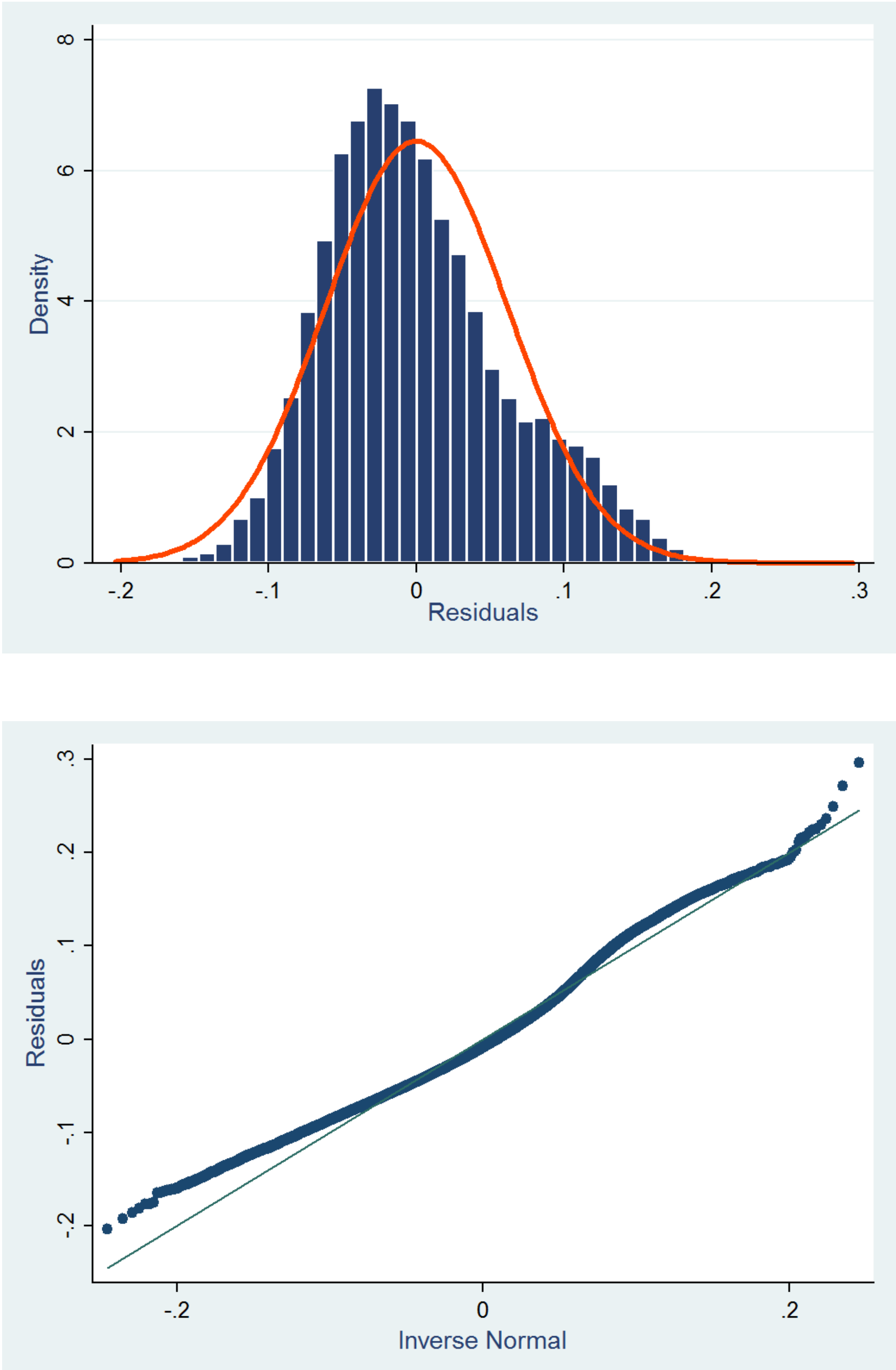


Figure B4: Factor 1 – Diagnostic plot: Residual plot of Level 1 for Random Intercept Model

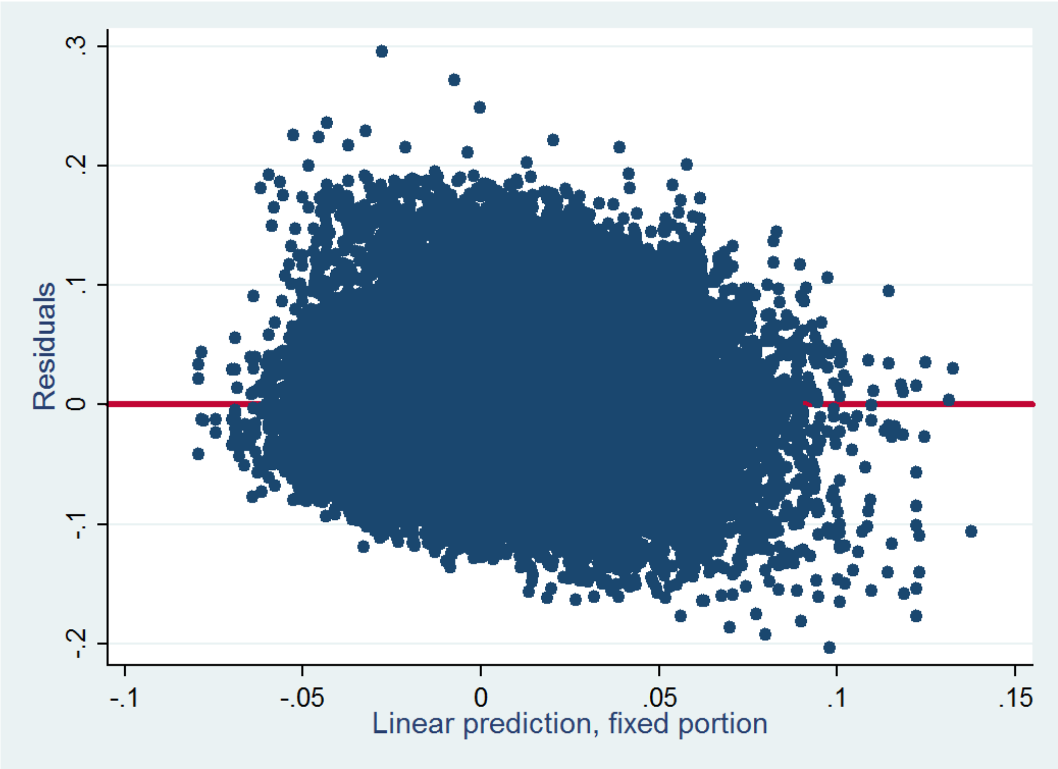
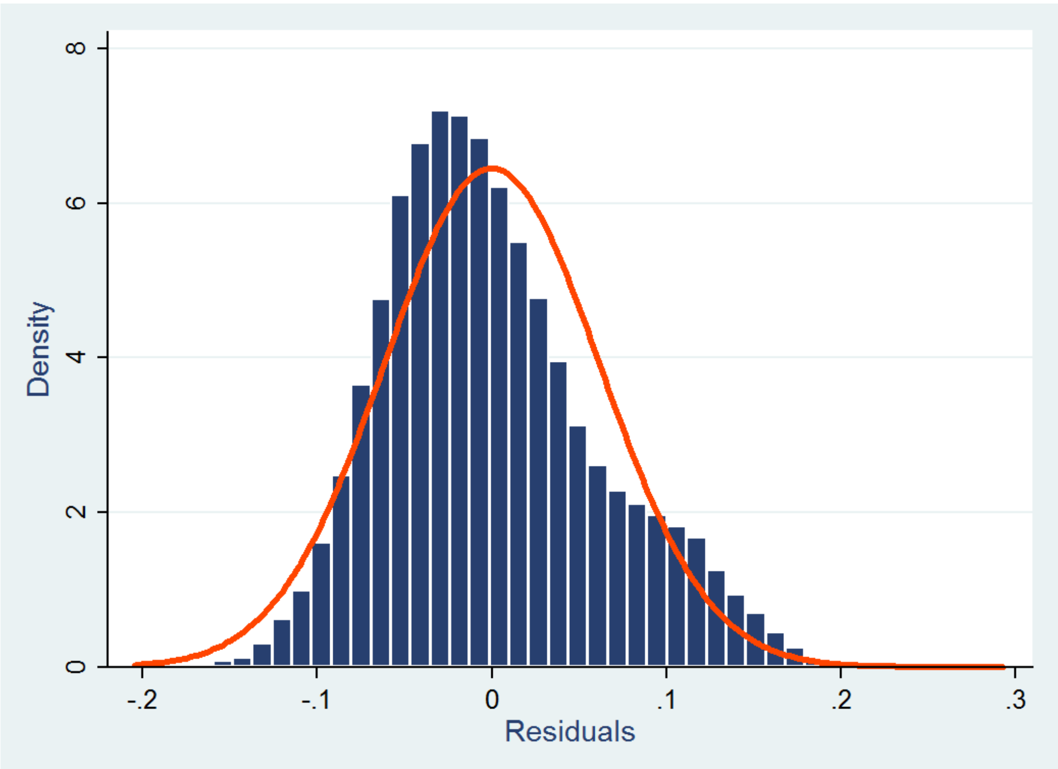


Figure B5 A & B: Factor 1 – Diagnostic plots: Residuals distributions of Level 1 for Contextual Effects Model



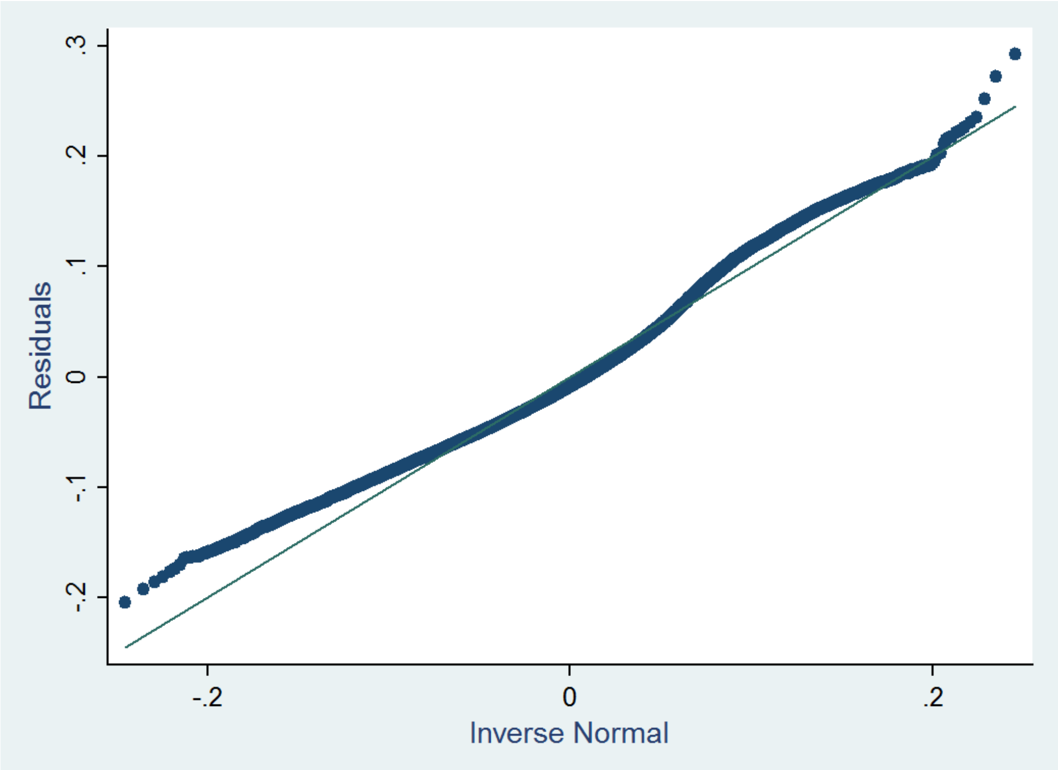


Figure B6: Factor 1 – Diagnostic plot: Residual plot of Level 1 for Contextual Effects Model

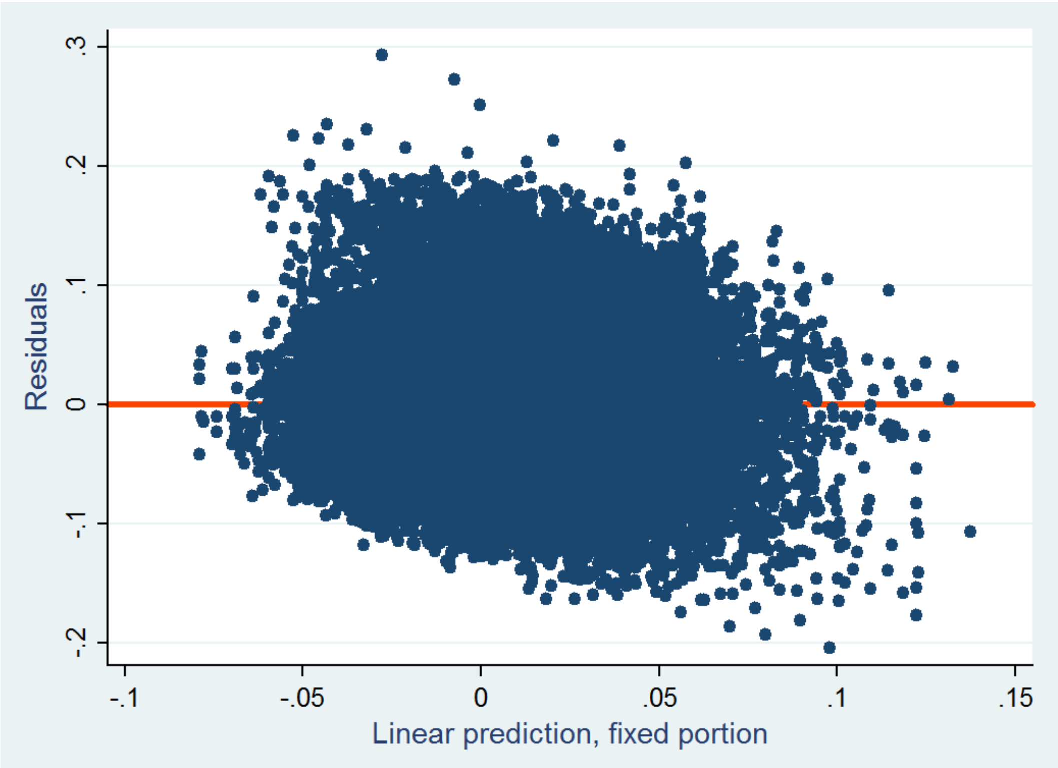


Figure B7: Variation of Factor 2 between MSOA

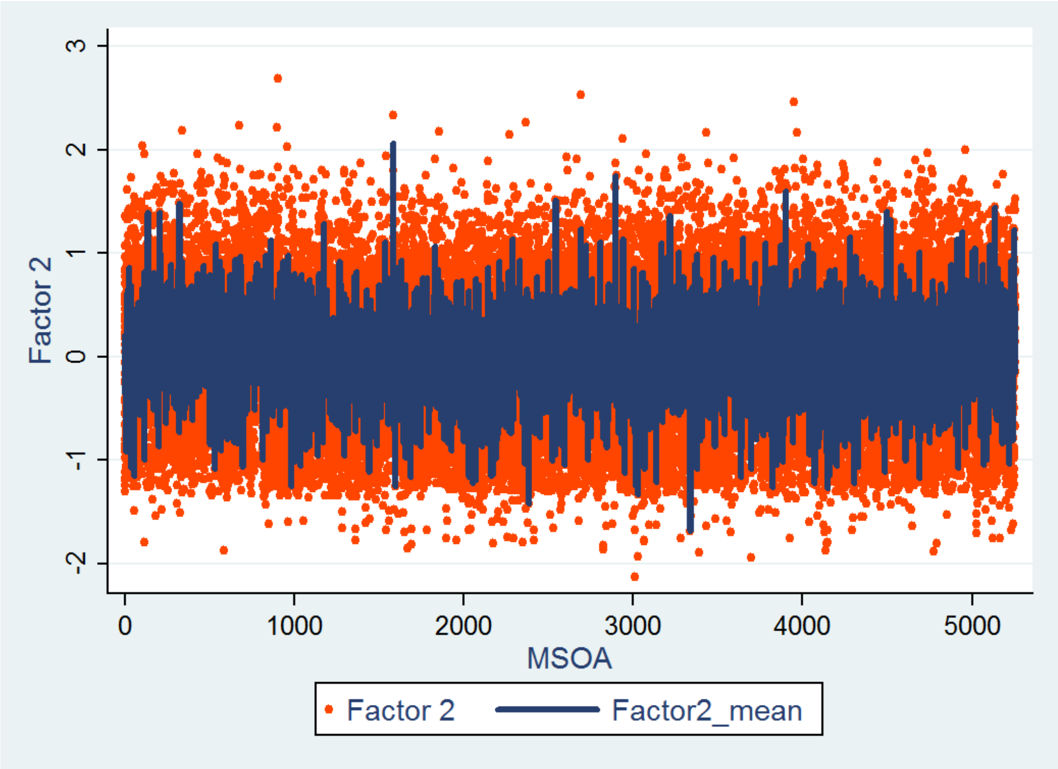


Figure B8: MSOA effects in rank for Factor 2, Null Model

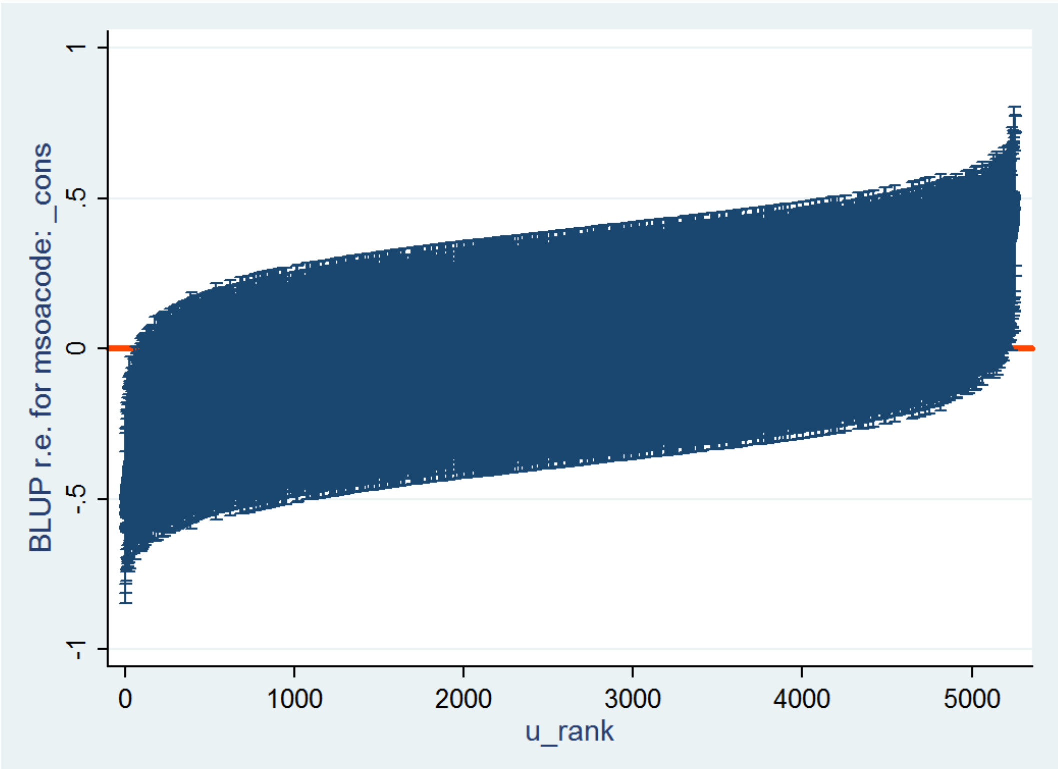
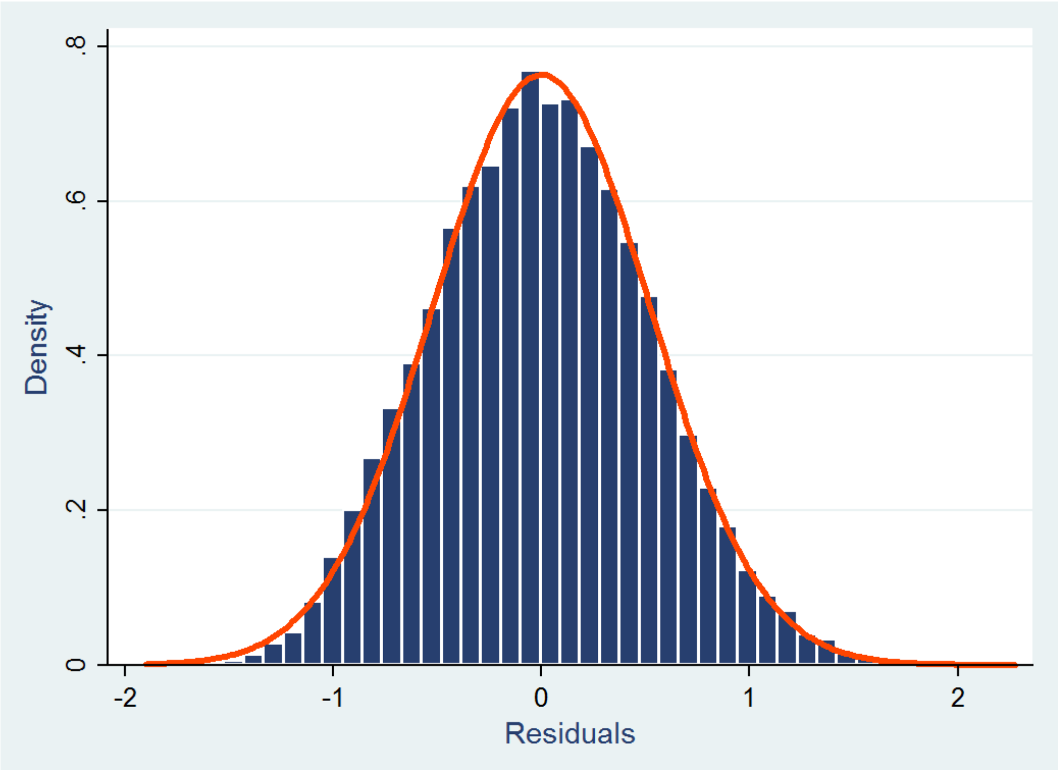


Figure B9 A & B: Factor 2 – Diagnostic plots: Residuals distributions of Level 1 for Random Intercept Model



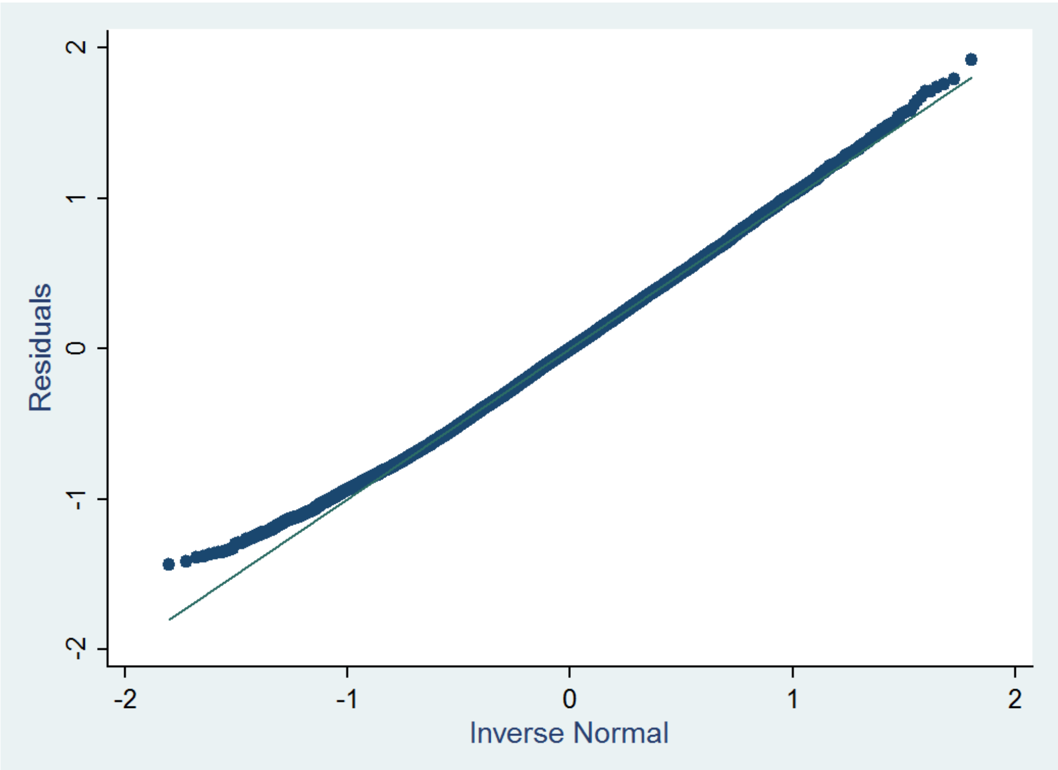


Figure B10: Factor 2 – Diagnostic plot: Residual plot of Level 1 for Random Intercept Model

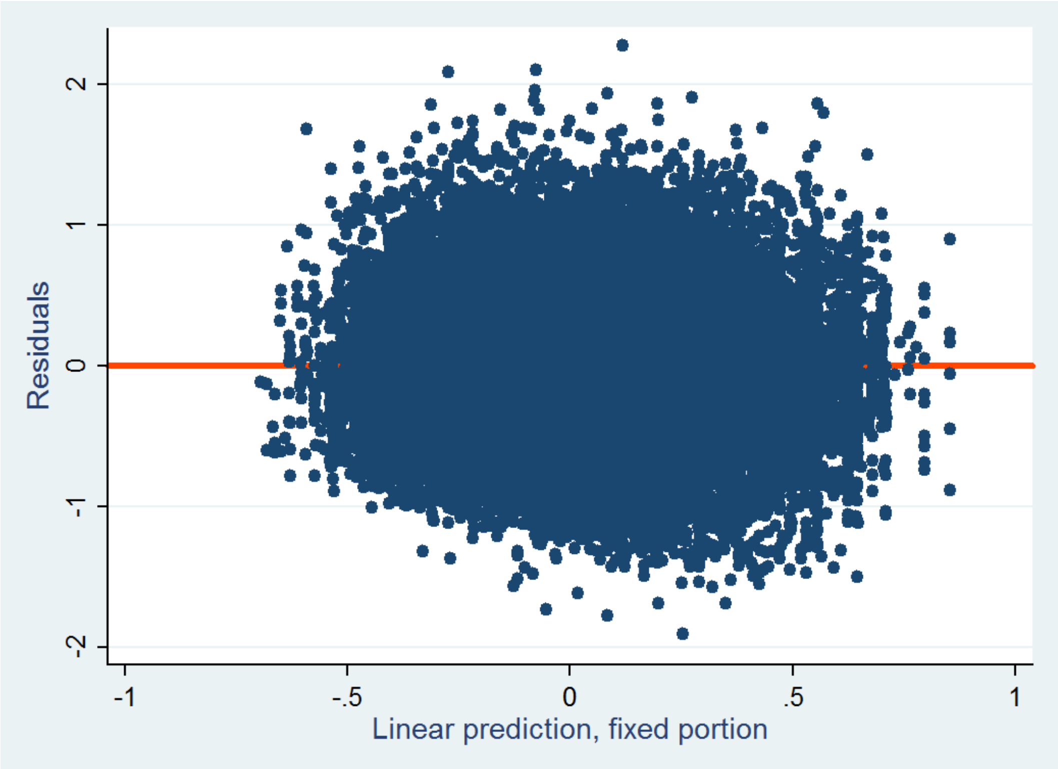
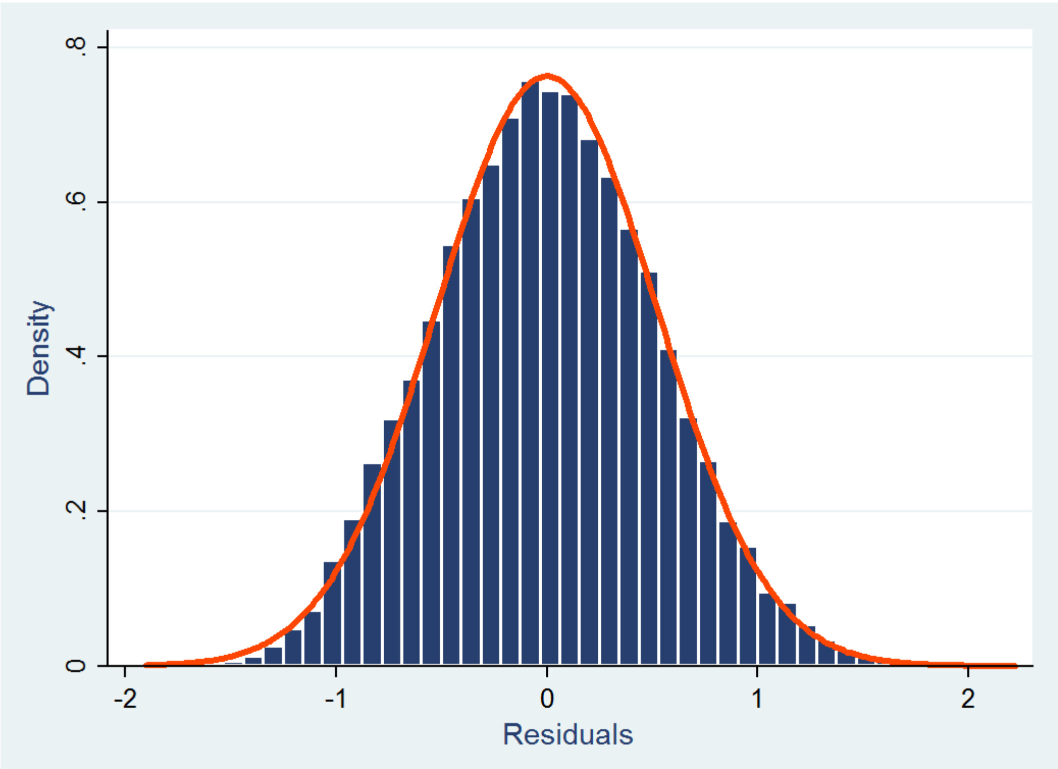


Figure B11 A & B: Factor 2 – Diagnostic plots: Residuals distributions of Level 1 for Contextual Effects Model



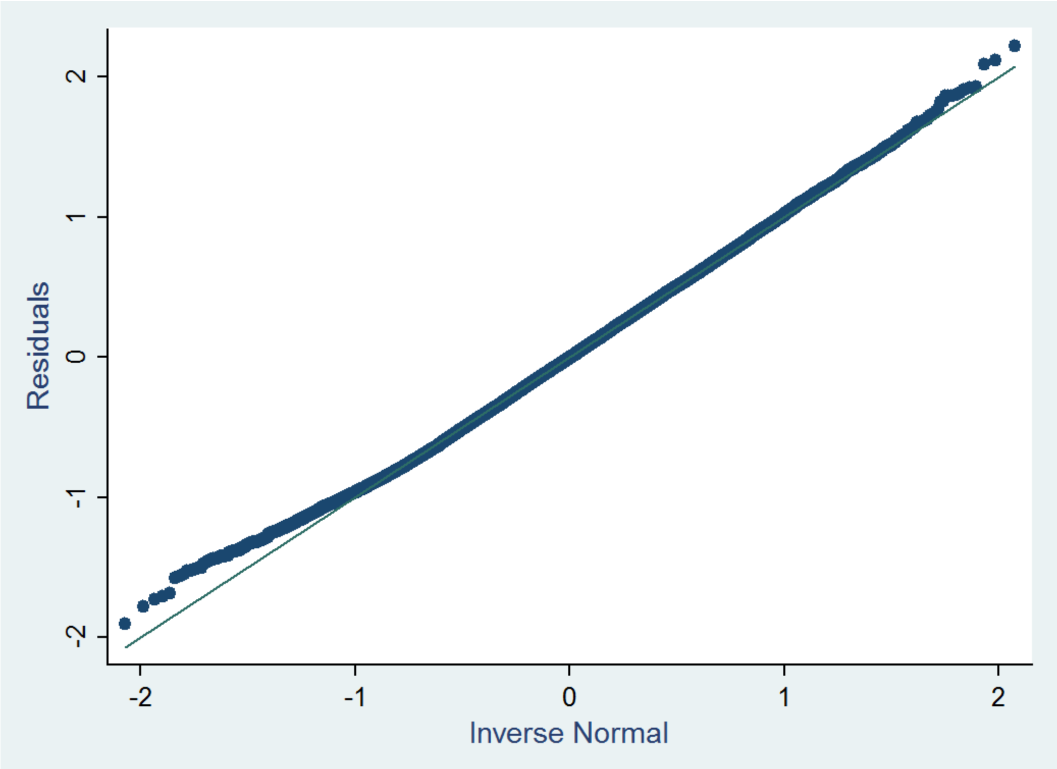


Figure B12: Factor 2 – Diagnostic plot: Residual plot of Level 1 for Contextual Effects Model

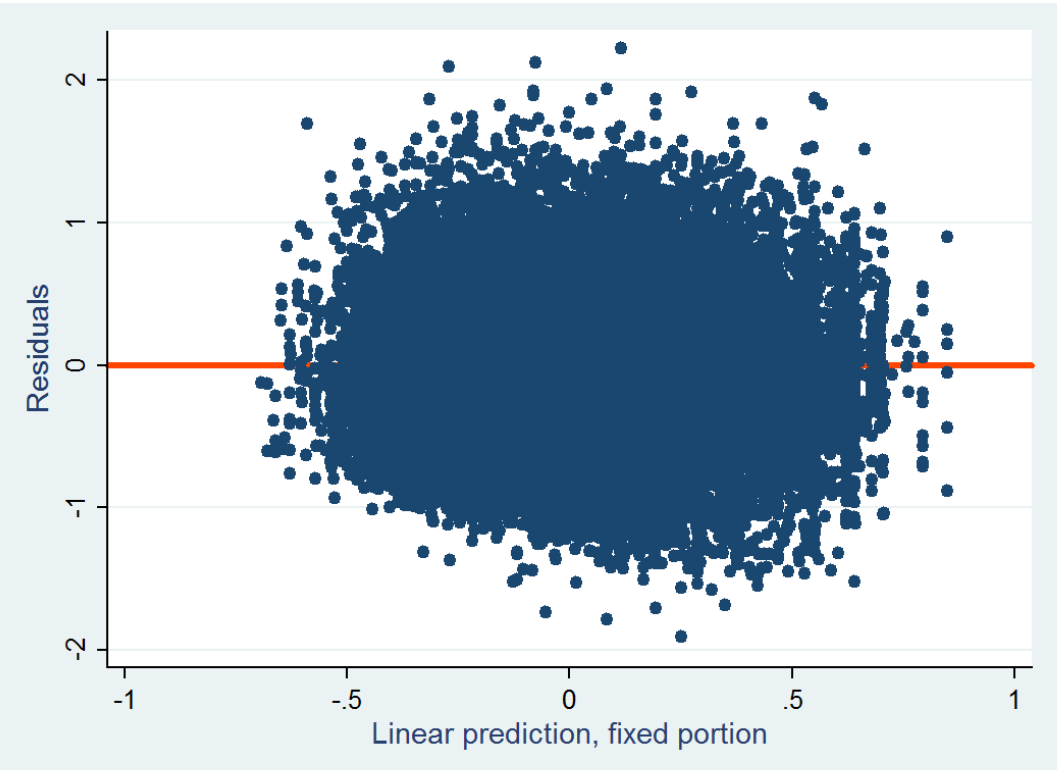


Figure B13: Variation of Factor 3 between MSOA

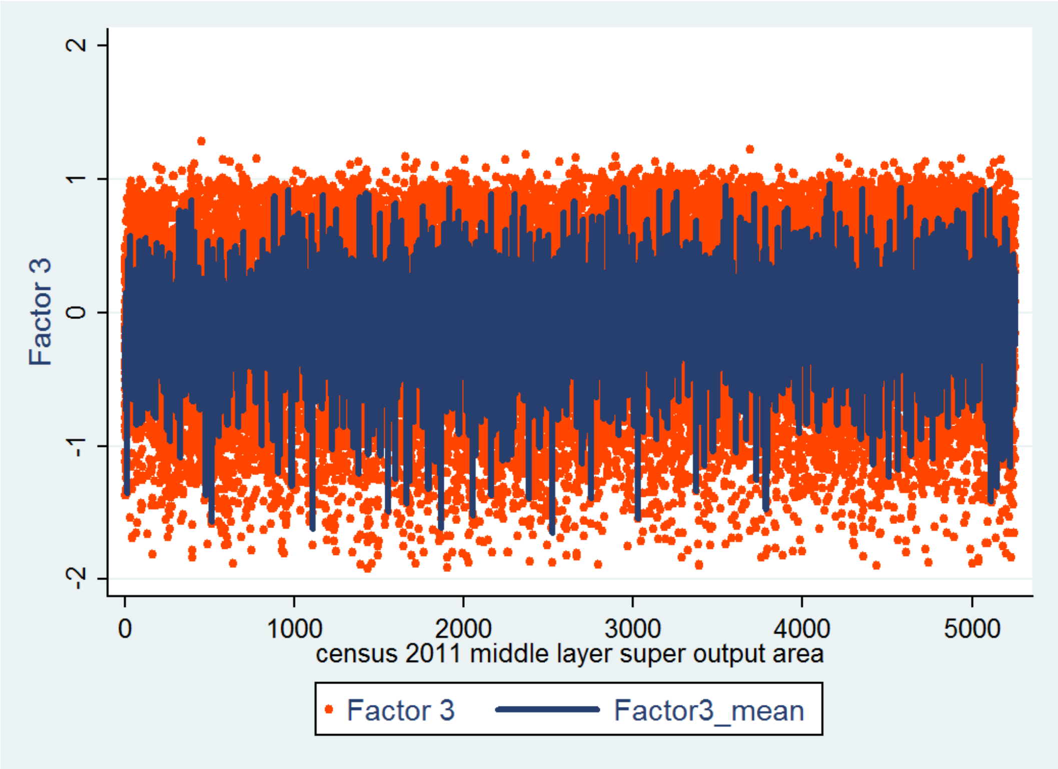


Figure B14: MSOA effects in rank for Factor 3, Null Model

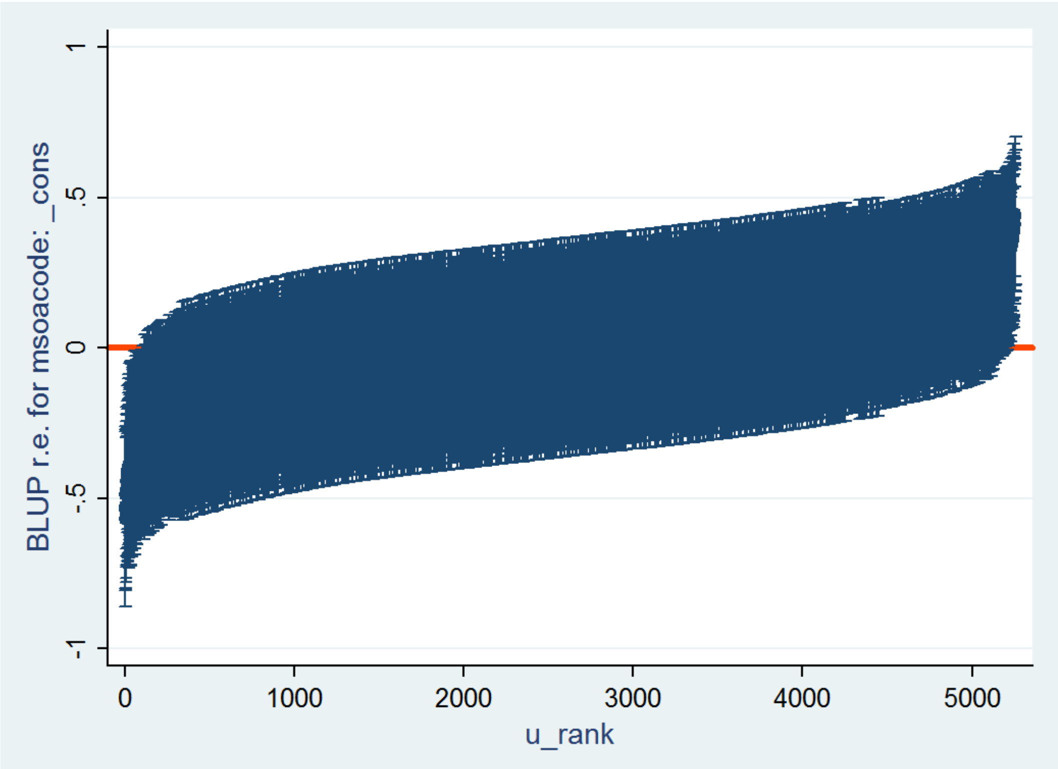


Figure B15 A & B: Factor 3 – Diagnostic plots: Residuals distributions of Level 1 for Random Intercept Model

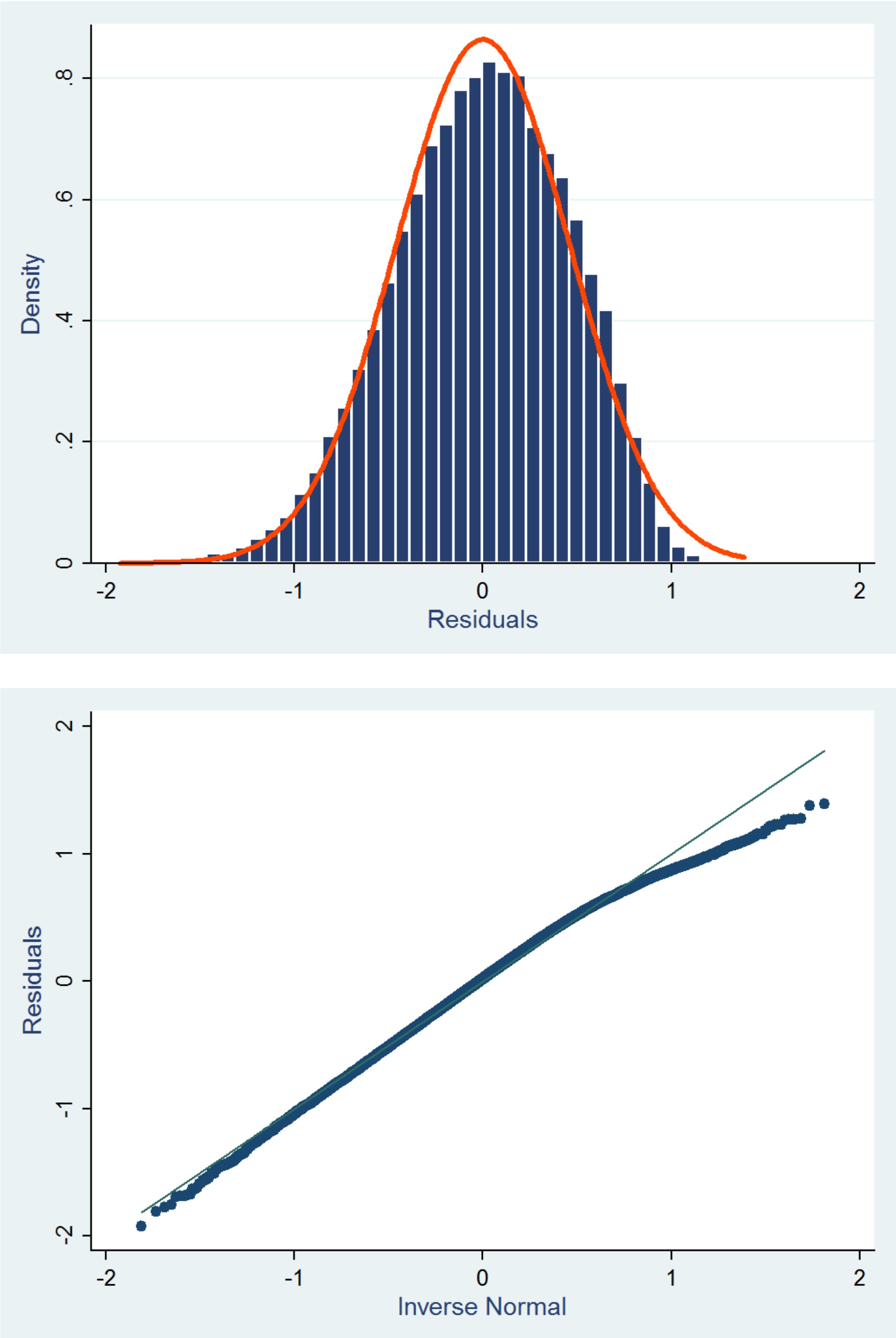


Figure B16: Factor 3 – Diagnostic plot: Residual plot of Level 1 for Random Intercept Model

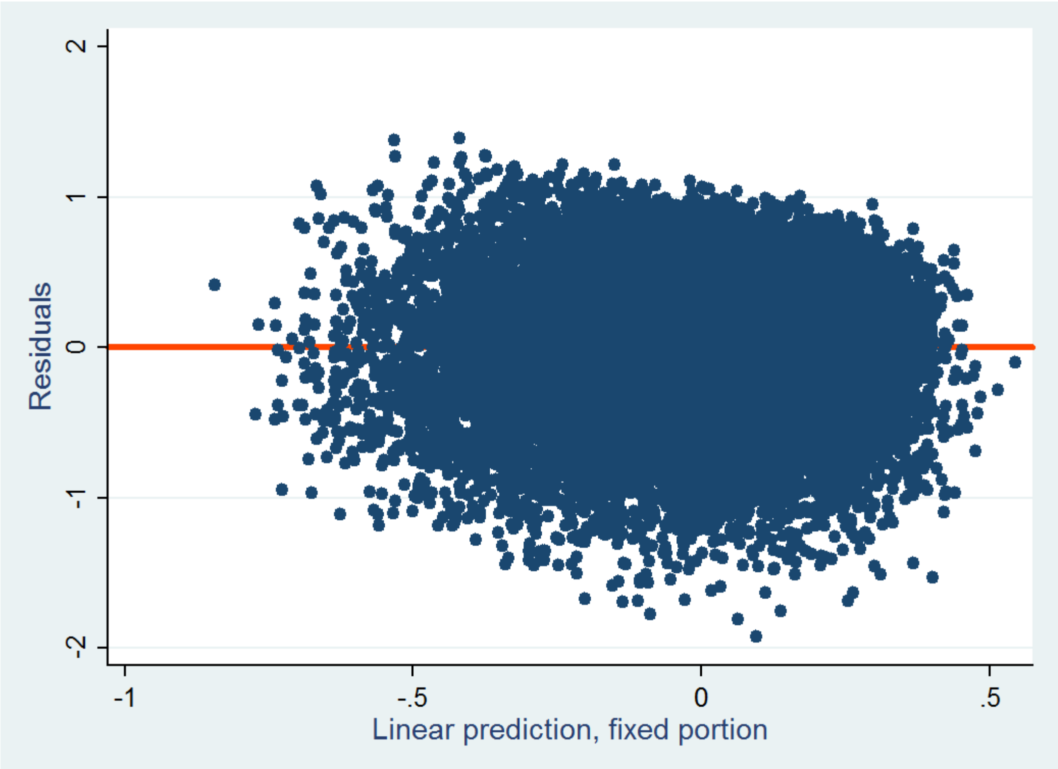
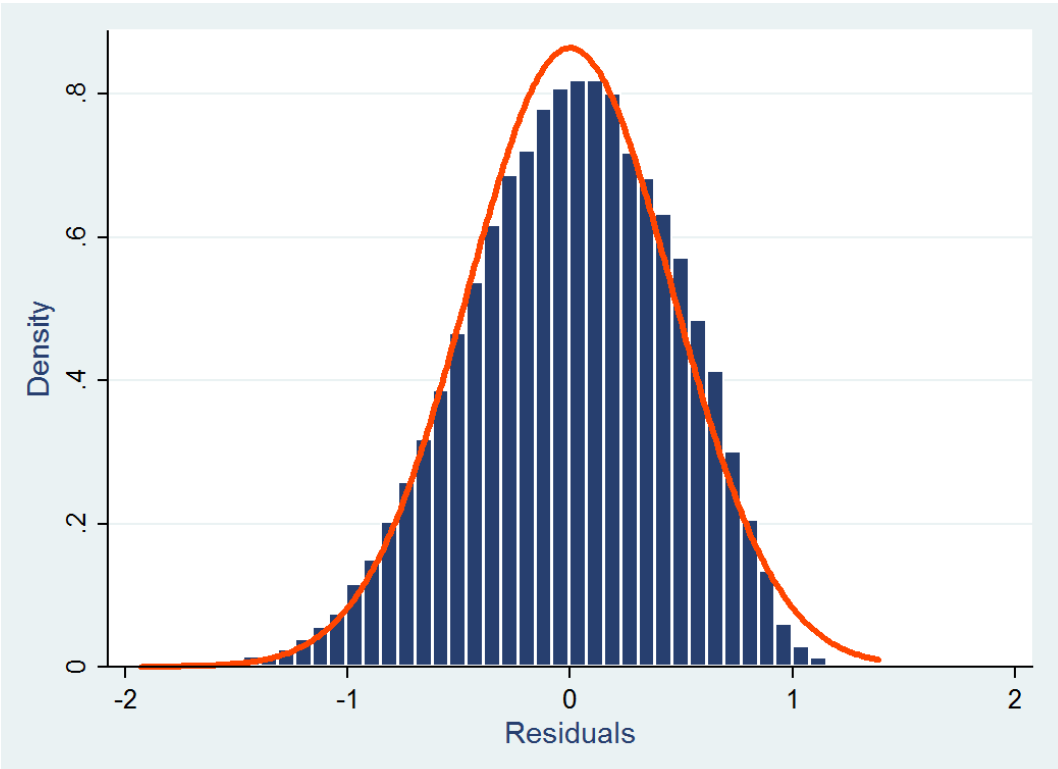


Figure B17 A & B: Factor 3 – Diagnostic plots: Residuals distributions of Level 1 for Contextual Effects Model



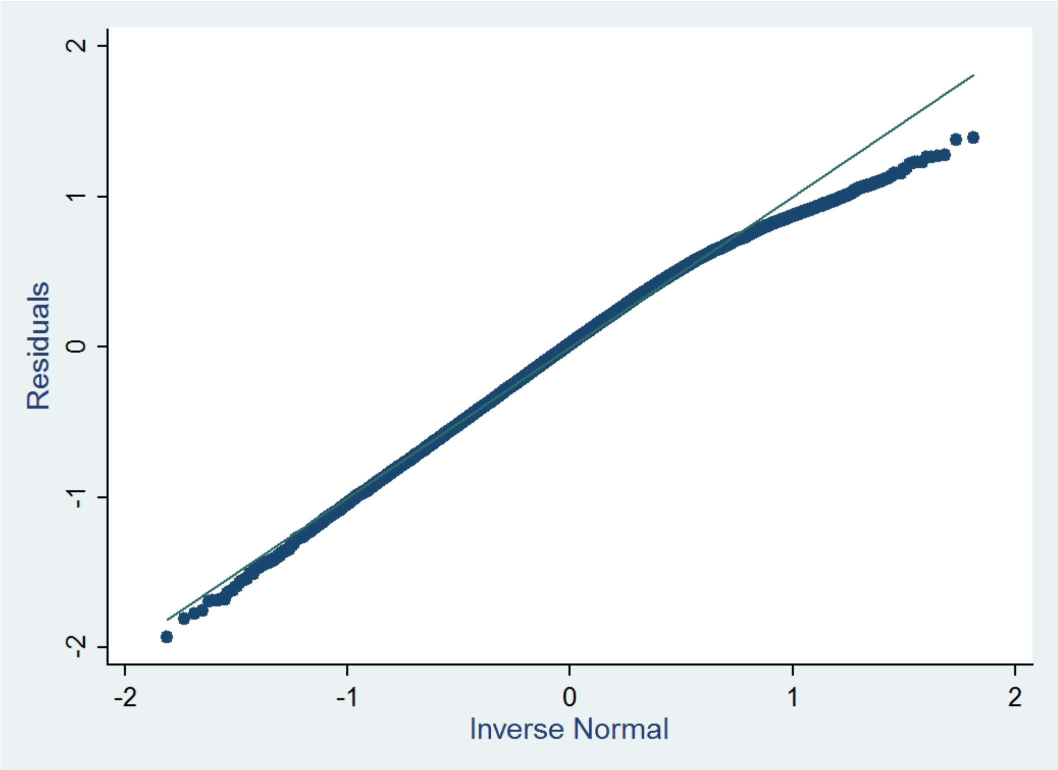
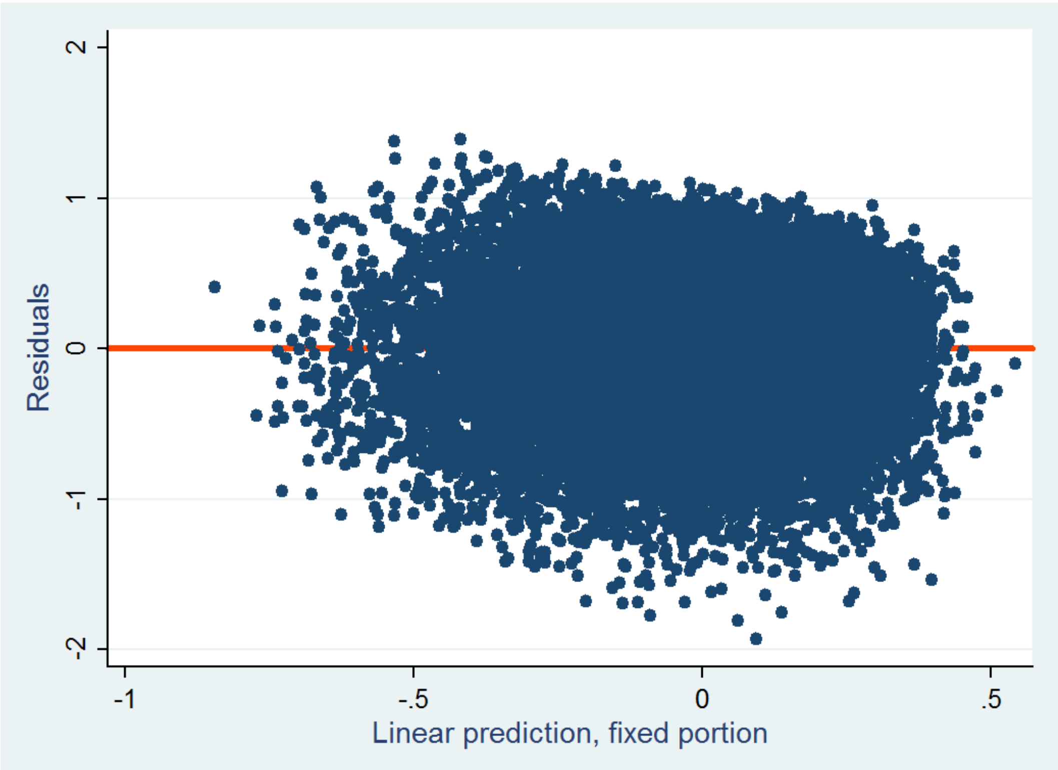


Figure B18: Factor 3 – Diagnostic plot: Residual plot of Level 1 for Contextual Effects Model



Supplementary Material

In this section I report results and relative interpretation to check that missing data are random and they do not bias results. As we saw in the previous sections, all the tables reporting results for the factors show that the number of observations of the models is lower than the full sample 37,932. More or less all the models for Factor 1 and 2 have about 10,000 observations less than the full sample available whereas models of Factor 3 have even more missing data – around 14,000 less observations than the total. Therefore, a check on the nature of these missing data is necessary (see par. 3.3.2 for further references).

SM.1 Factor 1 - Membership

The first step was to run Independent Sample T-Tests to compare means of several variables between the two groups: full sample and sample with missing data from the estimates of the models. The hypothesis is that there is no difference between the means, meaning that missing data are random and they do not bias results.

I used variables about sex, age, marital status and educational attainment from the original dataset[25]. I choose them because they do not have missing data at all and they can give information about the full sample. The hypothesis of no differences between the means of the groups has been rejected both for RIM and for RCM.

25 Age is a continuous variable; Sex is a dummy variable with 1 for Female, Education level is a categorical variable with 6 values: No qualifications, other qualifications, GSCE, A-level, Degree, Other higher degree; Marital status is a dummy variable with 1 for Married/Cohabiting/Civil partnership/Same sex partnership.

I then run logistic regressions to investigate it further. Dependent variable is a coded 1 for observations with at least a missing value and 0 otherwise. It then identifies the two samples.

Results are reported in the following table:

Table SM.1.1: Logistic regression results

	Random Intercept Model	Contextual Effects Model
Constant	-1.139*** (0.049)	-1.138*** (0.049)
Age	0.028*** (0.000)	0.028*** (0.000)
Sex	0.478*** (0.024)	0.478*** (0.024)
Educational level	0.005 (0.004)	0.006 (0.004)
Marital status	0.216*** (0.025)	0.216*** (0.025)

As we can see, apart for the educational level, all the variables are significant whereas to support the hypothesis of non-biased results due to missing data they should not result significant.

I then checked if means of variables used in the models are so different according to descriptive statistics.

In the following table, we can see the descriptive statistics for the two samples and noticing that they are not particularly different and, if any, differences are infinitesimal. This allows adjusting for MAR assumption:

Table SM.1.2: Descriptive statistics for Random Intercept Model[26]

	Full sample			Model's sample (no missing)		
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
Factor 1	10,188	0.0032	0.0597	27,744	0.0070	0.0730
Male	10,188	0.5428	0.4982	27,744	0.4326	0.4954

26 Reference dummy for each category of models is not reported.

Age 25-34	10,188	0.1778	0.3823	27,744	0.1438	0.3509
Age 35-45	10,188	0.1743	0.3794	27,744	0.2160	0.4115
Age 46-55	10,188	0.1356	0.3423	27,744	0.1922	0.3940
Age 56-65	10,188	0.0979	0.2971	27,744	0.1686	0.3744
Age 66 and more	10,188	0.1243	0.3299	27,744	0.1983	0.3987
Good health	10,151	0.7980	0.4015	27,744	0.7852	0.4107
Fair health	10,151	0.1312	0.3377	27,744	0.1540	0.3610
Single	10,188	0.4357	0.4959	27,744	0.3349	0.4720
Do not provide care	6,955	0.8479	0.3592	27,744	0.8020	0.3985
Care 20 hrs or more per week	6,955	0.0449	0.2070	27,744	0.0519	0.2218
Not White	9,704	0.3458	0.4757	27,744	0.1794	0.3837
High and Low Managerial, Administrative and Professional	10,007	0.1994	0.3995	27,744	0.2583	0.4377
Intermediate, Small Employers, Own Account	10,007	0.1340	0.3407	27,744	0.1462	0.3533
GSCE level	6,909	0.2327	0.4226	27,744	0.2011	0.4008
A level	6,909	0.2873	0.4525	27,744	0.1869	0.3898
Degree and higher	6,909	0.3018	0.4591	27,744	0.3620	0.4806
Other qualifications	6,909	0.0753	0.2638	27,744	0.1131	0.3168
Muslim	4,381	0.1719	0.3773	27,744	0.0740	0.2618
Other religion	4,289	0.0814	0.2734	27,744	0.0585	0.2347

Table SM.1.3: Descriptive statistics for Contextual Effects Model

	Full sample			Model's sample (no missing)		
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
Factor 1	10,188	0.0032	0.0597	27,744	0.0070	0.0730
Male	10,188	0.5428	0.4982	27,744	0.4326	0.4954
Age 25-34	10,188	0.1778	0.3823	27,744	0.1438	0.3509
Age 35-45	10,188	0.1743	0.3794	27,744	0.2160	0.4115
Age 46-55	10,188	0.1356	0.3423	27,744	0.1922	0.3940
Age 56-65	10,188	0.0979	0.2971	27,744	0.1686	0.3744
Age 66 and more	10,188	0.1243	0.3299	27,744	0.1983	0.3987
Good health	10,151	0.7980	0.4015	27,744	0.7852	0.4107
Fair health	10,151	0.1312	0.3377	27,744	0.1540	0.3610
Single	10,188	0.4357	0.4959	27,744	0.3349	0.4720
Do not provide care	6,955	0.8479	0.3592	27,744	0.8020	0.3985
Care 20 hrs or more per week	6,955	0.0449	0.2070	27,744	0.0519	0.2218
Not White	9,704	0.3458	0.4757	27,744	0.1794	0.3837

High and Low Managerial, Administrative and Professional	10,007	0.1994	0.3995	27,744	0.2583	0.4377
Intermediate, Small Employers, Own Account	10,007	0.1340	0.3407	27,744	0.1462	0.3533
GSCE level	6,909	0.2327	0.4226	27,744	0.2011	0.4008
A level	6,909	0.2873	0.4525	27,744	0.1869	0.3898
Degree and higher	6,909	0.3018	0.4591	27,744	0.3620	0.4806
Other qualifications	6,909	0.0753	0.2638	27,744	0.1131	0.3168
Muslim	4,381	0.1719	0.3773	27,744	0.0740	0.2618
Other religion	4,289	0.0814	0.2734	27,744	0.0585	0.2347
Heterogeneity	10,188	0.0952	1.0280	27,744	-0.0350	0.9872
Economic Profile	10,188	-0.0238	1.0330	27,744	0.0087	0.9875
Ethnicity Index	10,188	0.0799	1.0251	27,744	-0.0293	0.9890

SM.2 Factor 2 – Citizenship and Politics

The same procedure has been applied also for Factor 2. After T-tests have shown that differences between samples are different from zero both for RIM and for CEM, I investigated furtherly running logistic regressions on a similar dependent variable coded 1 for observations with at least a missing value and 0 otherwise and same predictors of the previous case: age, sex, educational level and marital status. Results are shown in the following table:

Table SM.2.1: Logistic regression results

	Random Intercept Model	**Contextual Effects Model**
Constant	-1.146*** (0.049)	-1.146*** (0.049)
Age	0.028*** (0.000)	0.028*** (0.000)
Sex	0.476*** (0.024)	0.475*** (0.024)
Educational level	0.006 (0.004)	0.006 (0.004)
Marital status	0.221*** (0.025)	0.221*** (0.025)

As we can see, they do not differ and, as for Factor 1, all variables are significant apart for the educational level. To see if differences in the means are important and MAR assumption consequently cannot hold I checked those using descriptive statistics in both the samples, as in the previous case:

Table SM.2.2: Descriptive statistics for Random Intercept Model

	Full sample			Model's sample (no missing)		
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
Factor 2	10,232	0.0223	0.6101	27,700	0.0407	0.6193
Male	10,232	0.5422	0.4982	27,700	0.4326	0.4955
Age 25-34	10,232	0.1775	0.3821	27,700	0.1439	0.3510
Age 35-45	10,232	0.1746	0.3797	27,700	0.2159	0.4115
Age 46-55	10,232	0.1361	0.3430	27,700	0.1921	0.3939
Age 56-65	10,232	0.0976	0.2968	27,700	0.1688	0.3746
Age 66 and more	10,232	0.1239	0.3295	27,700	0.1985	0.3989
Good health	10,195	0.7981	0.4014	27,700	0.7851	0.4107
Fair health	10,195	0.1312	0.3377	27,700	0.1540	0.3610
Single	10,232	0.4364	0.4960	27,700	0.3345	0.4718
Employed	10,212	0.6112	0.4875	27,700	0.6417	0.4795
GSCE level	6,953	0.2327	0.4226	27,700	0.2011	0.4008
A level	6,953	0.2865	0.4522	27,700	0.1869	0.3898
Degree and higher	6,953	0.3026	0.4594	27,700	0.3618	0.4805
Other qualifications	6,953	0.0755	0.2642	27,700	0.1131	0.3168
Do not provide care	6,999	0.8475	0.3595	27,700	0.8021	0.3985
Care 20 hrs or more per week	6,999	0.0449	0.2070	27,700	0.0519	0.2219
Not White	9,748	0.3457	0.4756	27,700	0.1792	0.3835
Muslim	4,425	0.1715	0.3770	27,700	0.0739	0.2616
Other religion	4,333	0.0812	0.2732	27,700	0.0585	0.2347
High and Low Managerial, Administrative and Professional	10,051	0.1998	0.3999	27,700	0.2582	0.4376
Intermediate, Small Employers, Own Account	10,051	0.1339	0.3406	27,700	0.1463	0.3534
3 or more rooms in the house	10,146	0.2246	0.4174	27,700	0.2295	0.4205

Table SM.2.3: Descriptive statistic for Contextual Effects Model

	Full sample			Model's sample (no missing)		
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
Factor 2	10,233	0.0224	0.6102	27,699	0.0407	0.6193
Male	10,233	0.5422	0.4982	27,699	0.4327	0.4955

Age 25-34	10,233	0.1775	0.3821	27,699	0.1439	0.3510
Age 35-45	10,233	0.1746	0.3797	27,699	0.2159	0.4115
Age 46-55	10,233	0.1361	0.3429	27,699	0.1921	0.3939
Age 56-65	10,233	0.0976	0.2968	27,699	0.1688	0.3746
Age 66 and more	10,233	0.1240	0.3296	27,699	0.1985	0.3989
Good health	10,196	0.7982	0.4014	27,699	0.7851	0.4107
Fair health	10,196	0.1312	0.3377	27,699	0.1540	0.3610
Single	10,233	0.4364	0.4960	27,699	0.3345	0.4718
Employed	10,213	0.6113	0.4875	27,699	0.6417	0.4795
GSCE level	6,954	0.2327	0.4226	27,699	0.2011	0.4008
A level	6,954	0.2865	0.4521	27,699	0.1869	0.3898
Degree and higher	6,954	0.3027	0.4595	27,699	0.3618	0.4805
Other qualifications	6,954	0.0755	0.2642	27,699	0.1131	0.3168
Do not provide care	7,000	0.8476	0.3595	27,699	0.8021	0.3985
Care 20 hrs or more per week	7,000	0.0449	0.2070	27,699	0.0519	0.2219
Not White	9,749	0.3457	0.4756	27,699	0.1792	0.3835
Muslim	4,426	0.1715	0.3770	27,699	0.0739	0.2616
Other religion	4,334	0.0812	0.2732	27,699	0.0585	0.2347
High and Low Managerial, Administrative and Professional	10,052	0.1998	0.3998	27,699	0.2582	0.4377
Intermediate, Small Employers, Own Account	10,052	0.1339	0.3406	27,699	0.1463	0.3534
3 or more rooms in the house	10,147	0.2246	0.4173	27,699	0.2295	0.4205
Heterogeneity	10,232	0.0963	1.0283	27,699	-0.0356	0.9870
Economic Profile	10,232	-0.0251	1.0334	27,699	0.0093	0.9872
Ethnicity Index	10,233	0.0813	1.0254	27,699	-0.0300	0.9888

In addition, in this case, the differences in means between the two samples for most of variables used are small.

SM.3 Factor 3 – Neighbourliness

Finally, the same check of missing data has been done also for the last factor about Neighbourliness.

Independent T-tests have been carried out and, as for previous factors; for all the tested variables, the hypothesis of no differences between means have been rejected.

Progressively, logistic regressions have been run, with the usual dummy dependent variable about the two samples and the predictors about sample: age, sex, marital status and educational level.

In the following table, we can see the results for both the RIM and the CEM models of this factor:

Table SM.3.1: Logistic regression results

	Random Intercept Model	**Contextual Effects Model**
Constant	-1.025*** (0.045)	-1.024*** (0.000)
Age	0.010*** (0.000)	0.010*** (0.000)
Sex	0.472*** (0.021)	0.472*** (0.021)
Educational level	-0.002 (0.003)	-0.002 (0.003)
Marital status	0.435*** (0.022)	0.436*** (0.022)

As for the first two factors, all variables, except for the ordinal about educational level, are significant.

Therefore, to check if differences in means are so important to reject the MAR assumptions, I used descriptive statistics. Summarized in the following tables, we can see than also in this case differences are minimal, so MAR assumption can hold for all the models:

Table SM.3.2: Descriptive statistics for Random Intercept Model

	Full sample			**Model's sample (no missing)**		
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
Factor 3	14,704	-0.1027	0.5050	23,228	0.0017	0.5481
Male	14,704	0.5280	0.4992	23,228	0.4205	0.4937
Age 25-34	14,704	0.1706	0.3762	23,228	0.1417	0.3488
Age 35-45	14,704	0.1616	0.3681	23,228	0.2321	0.4222
Age 46-55	14,704	0.1453	0.3524	23,228	0.1970	0.3978
Age 56-65	14,704	0.1284	0.3345	23,228	0.1630	0.3694

Age 66 and more	14,704	0.1761	0.3809	23,228	0.1799	0.3841
Not White	14,220	0.2802	0.4491	23,228	0.1872	0.3901
Good health	14,667	0.7795	0.4146	23,228	0.7944	0.4042
Fair health	14,667	0.1452	0.3523	23,228	0.1496	0.3567
Muslim	8,897	0.1007	0.3010	23,228	0.0822	0.2747
Other religion	8,805	0.0705	0.2560	23,228	0.0582	0.2341
Single	14,704	0.4298	0.4951	23,228	0.3191	0.4661
Dependent children	12,226	0.1179	0.3225	23,228	0.2502	0.4331
3 or more rooms in the house	14,618	0.2110	0.4081	23,228	0.2390	0.4265
Do not provide care	11,471	0.8422	0.3646	23,228	0.7959	0.4030
Care 20 hrs or more per week	11,471	0.0470	0.2116	23,228	0.0522	0.2225
Intermediate local school services	6,112	0.3109	0.4629	23,228	0.2994	0.4580
Good local school services	6,112	0.5299	0.4991	23,228	0.5663	0.4956
Intermediate other local services	11,325	0.4852	0.4998	23,228	0.4547	0.4980
Good other local services	11,325	0.4047	0.4909	23,228	0.4319	0.4954
Moved	13,842	0.8985	0.3020	23,228	0.9070	0.2905

Table SM.3.3: Descriptive statistic for Contextual Effects Model

	Full sample			Model's sample (no missing)		
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Obs.**	**Mean**	**Std. Dev.**
Factor 3	14,705	-0.1027	0.5050	23,227	0.0017	0.5481
Male	14,705	0.5280	0.4992	23,227	0.4205	0.4937
Age 25-34	14,705	0.1706	0.3762	23,227	0.1417	0.3488
Age 35-45	14,705	0.1616	0.3681	23,227	0.2321	0.4222
Age 46-55	14,705	0.1453	0.3524	23,227	0.1971	0.3978
Age 56-65	14,705	0.1284	0.3345	23,227	0.1630	0.3694
Age 66 and more	14,705	0.1761	0.3809	23,227	0.1798	0.3841
Not White	14,221	0.2801	0.4491	23,227	0.1872	0.3901
Good health	14,668	0.7795	0.4146	23,227	0.7944	0.4042
Fair health	14,668	0.1452	0.3523	23,227	0.1496	0.3567
Muslim	8,898	0.1007	0.3009	23,227	0.0822	0.2747
Other religion	8,806	0.0705	0.2560	23,227	0.0582	0.2341
Single	14,705	0.4299	0.4951	23,227	0.3190	0.4661
Dependent children	12,227	0.1179	0.3224	23,227	0.2502	0.4331
3 or more rooms in the house	14,619	0.2110	0.4081	23,227	0.2390	0.4265
Do not provide care	11,472	0.8422	0.3645	23,227	0.7959	0.4030
Care 20 hrs or more per week	11,472	0.0470	0.2116	23,227	0.0522	0.2225

Intermediate local school services	6,113	0.3108	0.4629	23,227	0.2994	0.4580
Good local school services	6,113	0.5300	0.4991	23,227	0.5663	0.4956
Intermediate other local services	11,326	0.4852	0.4998	23,227	0.4547	0.4980
Good other local services	11,326	0.4047	0.4909	23,227	0.4319	0.4953
Moved	13,843	0.8985	0.3020	23,227	0.9070	0.2905
Heterogeneity	14,704	0.0770	1.0131	23,227	-0.0488	0.9886
Economic Profile	14,704	-0.0050	1.0236	23,227	0.0032	0.9848
Ethnicity Index	14,705	0.0536	1.0085	23,227	-0.0340	0.9931

CHAPTER 4 – SMALL AREA ESTIMATES OF SOCIAL CAPITAL

4.1 Introduction

As claimed from the outset of this work, SC is a complex concept. Several problems have been identified in measuring it one of the main being the lack of indicators, especially the at small area level. Being a type of capital strongly dependent on an individual's networks and on local communities may affect the search for indicators. Indeed, we can notice a propensity to find indicators at regional or national levels (for example the European Social Survey) or to keep the analysis at survey level (therefore at sample level). However, due to the development of all Small Area Estimates (SAE hereafter) methods, it is possible to start thinking about estimating SC at an intermediate level between the individual and the regional. Estimating this, geographically speaking, at areas wide as LSOA or MSOA could help to capture the right dimension for its study. Therefore, after the development of the three factors with CFA (chapter 2) and the test of these factors with MM to see which individual and area effects influence it (chapter 3), I lastly estimate it for all the MSOA in England and Wales.

4.2 Small Area Estimates methods: an introduction

We can trace the first studies using SAE in two famous studies from Fay and Herriot (1979) and Battese *et al*. (1988). In these seminal studies, the authors estimate their dependent variable for all areas of interest using other data able to fill the lack of information. The main idea of SAE is, indeed, to use other sources of data to obtain synthetic estimation for the areas where we do not have information. This procedure overcomes the classical problem of using surveys: under a certain level, not all the areas are covered (no observations) and some of them have really low number of observations.

The choice of these complementary data is based on precise theoretical assumptions: they should be correlated to the dependent variable as well as the survey covariates. Empirical assumptions further specify that they should have observations for all areas and be recoded in the same way as the survey covariates. From these two studies, two main kinds of SAE methods have been identified: *unit-level* studies and *area-level* study. According to Namazi-Rad and Steel (2015) statistical models for SAE can be structured at individual or aggregate levels. If we have sufficient information about the geographical indicators for target areas available for all individuals in the sample then we can estimate a unit-level model. It is also possible to aggregate the data to area-level and estimate these parameters based on model for the area means.

If the unit-level model is properly specified, we can expect less variance. However, the use of different levels of data analysis often differs because of some model misspecifications. If our targets of inference are the area-level, we should ask ourselves when it is preferable to use an area-level analysis and under which conditions it may be better. Therefore, if we include contextual effects of the area-level means, the area-level analysis should produce less biased estimates of the regression coefficients. Indirect techniques for SAE rely on statistical models that, as just described, borrow strength from other auxiliary data resources. These data are used to include random effects to explain better variations between target areas within the population as well as several covariates for available auxiliary variables (Chambers and Tzavidis, 2006; Tzavidis *et al*., 2008).

After this main and important distinction, even more complex models have been studied (see Rao 2003 for a full and proper revision of all SAE methods). We can summarize them in this useful scheme (NatCen, 2004b) and give brief details of how they work and their main limitations:

1) Simple (non-modelled) methods using indirect standardization
2) Models using individual level covariates only
3) Models combining individual and area-level covariates
4) Models using areal level covariates only
5) Other approaches for larger areas of geography

The procedure of these models is almost the same: they use two datasets: a first dataset with individual level data (usually from survey) and a second dataset with population information per areas (usually Census or other administrative datasets). Differences lie in the choice of variables and covariates, as we will see in the brief description following.

1) Indirect standardization

This method applies national estimates derived from survey data to area-level population counts to generate *expected* area estimates. It mainly needs two datasets: a first dataset containing data at individual level, usually a survey, and a second dataset, usually Census or very detailed administrative dataset, containing population information. The second dataset must contain full information for all the areas. On the base of a common variable, the two datasets are merged.

From the first dataset, we would estimate proportions at national level for classes or band or cohorts (according to variables) and then these national estimates would have been applying to the census counts of the same class or band, obtaining the estimates of the number with those characteristic for each area. Dividing by the total census count in the area finally would give us the proportion by ward. Essentially is like obtaining the national prevalence rates for sub-groups weighted by the proportions of sub-groups of persons in the small area.

It has been considered appealing because in an intuitive way it seems likely that the mean level of many variables in a population is related to the distribution in the population of such kinds of demographic variables. It is also easy to implement because the cell proportions for all the areas are available from the Census and estimates from national surveys. The major drawback is the assumption that all the national rates for each sub-group apply uniformly across all areas. So, even if theoretically we can assume that differences are due to demographic composition and therefore different areas with the same characteristics should expect similar rates, several studies have shown that even within the same social groups there are differences due to ‘contextual effects’ operating at area level. Just to cite some examples, Macintyre et al (2002) find that both material infrastructure and collective social functioning may affect health. On the path of ecological analysis for health studies, Duncan et al. (1993) used aggregate levels of deprivation in the area of residence to see how it can influence health-related behaviours and psychiatric mobility. Even when area effects resulted not significant, like in the study of Sloggett and Joshi (1998) on adverse fertility events, they highlighted that some observed area or regional variations are not explained only by individual deprivation. In some way, population characteristics and area-level effects should be taken into account.

2) Models using individual level covariates only

We can consider this as an extension of the previous model: the relationship is modelled between measures obtained from the survey against predictor variables acquired from the same survey for the same individuals. All these data belong then to the same dataset, the survey dataset. Only in a second stage, covariates from a second dataset (e.g. counts for all small areas from the Census) are chosen and used.

The models estimate the probability that a person has the characteristic of the dependent variables with specific known characteristics. The model-based estimates probabilities are then converted into estimated proportions in each sub-group defined by the covariates who fall into the relevant category. These proportions are finally applied to covariate counts available from the Census to derive estimates for all areas. Some studies use also covariates from the Sample of Anonymised Records (SAR) from the Census. Respect to the previous model, here the first probability estimates are done using individual level covariates from survey and only in a second step ‘weighted’ by Census information.

The main limitation related to data requirements: the exact correspondence between covariates used in the model and covariates from the Census. Besides, the cross-tabulations available from the Census are limited (because of sensitiveness of information) and, therefore, reduce the choice of predictors in these models.

3) Models combining individual and area level covariates

This model is the first of an alternative set to the first two models presented, the *multilevel models*, used in the previous chapter. Indeed, until now, we examined two type of models mainly focused on estimates (of proportions) at individual level that are then 'weighted' by population counts. Multilevel models instead incorporate area effects already in the first step of estimation with individual data, as widely described in the previous chapter. These area effects are the random effects. Their importance lies in the fact that they explain the significant systematic variation between small areas after the part accounted for from the covariates in the model, that are considered as fixed effects. Other main advantages are the higher suitability to the clustered nature of social survey, use of covariates at different levels that can allow exploration of the degree to which differences between geographical areas are associated with individual or household or area characteristics. In this case, one of the first seminal studies is Moura and Holt's (1999) paper *Small area estimation using multilevel models*.

Despite the use of both individual and area level covariates, the previous limitations remain valid: necessity of exact correspondence between individual and area covariates from the Census and limited cross-tabulations available from the Census at small area. Finally, as NatCen (2004b) suggests, the estimation of standard errors for the synthetic estimates based on MM is really complex.

4) Models using area level covariates only

These are case of more restricted multilevel models than the previous version. Using only area-level covariates, the regression estimates relate to between-area variation. It gives a constant predicted value for all individuals within an area that can be interpreted as a predicted mean for that small area. Then coefficient estimates are attached to the known means or proportion of the covariates for all areas, taken from the Census and other administrative sources, to obtain synthetic estimates. In this case, therefore, we still use two datasets: a first dataset with individual data. It is used to obtain coefficient estimates from a model where dependent variable comes from the survey and covariates from the population dataset with full information for all the areas (e.g. Census).

While advantages can be identified in less stringent data requirements about covariates and the decrease of potential redundancy of information from individual level covariates[27], the main drawback is the possibility of disaggregation of estimates for sub-groups within each small area.

5) Other approaches for larger geographical areas

There are three main methods for this class:

- *Generalised Regression Synthetic Estimator (GREG)*: this adjusts the survey predictions by taking into account any numerical differences between the

[27] According to several studies, it is reasonable to assume that if individual characteristics affect the dependent variable, in the same way the average level of the dependent in all the areas will differ only if there will be differences in these characteristics. So controlling for these variables at area level is considered enough.

survey and the Census area means. Its main drawback is that it cannot be used for those areas which do not contain any survey respondents. Therefore, it cannot be used for producing area-level estimates at small level but it may be suitable for larger areas;

- *Composite estimators*: these combine estimators from direct survey-based estimates (that are design-unbiased meaning that the expected value estimated for a small area is equal to its true value) and model-based estimates. It naturally balances the potential bias of a model based estimator against the large variability of an unbiased direct estimator, taking a weighted average of the two and obtaining an estimator that may be more accurate than either of its components. In practice, however, it requires at least one survey respondent in each estimation area of interest. Therefore, it is better to be used for larger areas estimates;
- *Fay-Herriot estimator*: already described, can be classified as an area-level model that relates the area means of the dependent variable from survey to area-specific covariates values and to random effects. Rao (2003) describes that the best predictor can be expressed as a weighted average of the survey-based estimator and a regression-synthetic estimator that uses the fixed effects only.

Other more recent approaches have been tested and continuously revised and improved: spatial microsimulation approaches, Iterative Proportional Fitting, Generalised Regression Reweighting, Combinatorial Optimisation, Agent-Based approaches, Bayesian Models and so on. An interesting and complete revision of the state of the art has been carried out by the National Centre for Research Methods (2013) under the supervision of Whitworth but including many authors involved in SAE following a conference held in order to build a network and a multi-method project for SAE.

I will now focus on the third type of SAE that are the models combining individual and area level covariates starting from multilevel models. I chose this way to reach the estimates of SC at small area and MM in the previous chapters have been set, as already declared, to this purpose.

4.2.1 Small area estimates and Multilevel Models

In the last year, this approach has been used increasingly in different fields of study; the most common and more developed being applications of SAE to health topics (Tranmer *et al*., 2005). All the studies originate from MM using a range of national surveys on health condition and Census information to estimate at small area[28] several critical aspects (from diseases, to obesity, drinking, smoking, health risk, mental illness, heart problems and so on): Hindmarsh (2013), Li and Zaslavsky (2010), Zhang *et al* (2014), Zhang *et al*. (2013), Twigg and Moon (2002), Twigg *et al*. (2000), Allaga and Muhuri (1994), Twigg *et al*. (2004).

Chambers and Tzavidis (2005) used MM and the Labour Force Survey to estimate unemployment and inactive force at British Local Authority District level and Unitary Districts.

[28] Some of them use other levels of clustering such as households or schools but the main trend is the estimation at small geographical levels in order also to support policies.

However, other topics also see the application of this method. Nguyen *et al*. (2010) use this method to estimate living standards starting from household information on income and other characteristics at small area level in Vietnam. Reder (1994) applies the methodology to produce synthetic estimates of literacy proficiency at small Census areas. Whitworth (2012) estimates the fear of crime at MSOA level using the British Crime Survey in a MM framework and then, adding area level covariates of different types, produces synthetic estimates of it. In the same way, Haughton and Nguyen (2010) estimate inequality in Vietnam.

4.2.2 Small Area Estimation using Multilevel Models and social capital studies

Concerning the SAE of Social Capital, there are few studies that actually analyse it. They mainly do so using SC variables more as explanatory rather than a dependent variable in itself. Sampson *et al*. (1999) collect survey data in Chicago and clustering by neighbourhoods in the MM completing the information with Census data, they focus on how spatial interdependence produces collective efficacy for children. They define the spatial interdependence according to neighbourliness SC: intergenerational closure, reciprocal local exchange and shared expectations for informal social control. They find several results. First, residential stability and concentrated affluence predict intergenerational closure and reciprocal exchange more than poverty and ethnic composition. Second, concentrated disadvantages are associated with sharply lower expectations for shared child control. Third, spatial dynamics (proximity to areas high in closure, exchange and control), generally, affect collective efficacy for children even more than neighbourhoods' characteristics.

Lastly, spatial advantages are naturally higher in white neighbourhoods more than black ones.

Mohan *et al*. (2005) want to test how much SC influences health outcomes. Using this method in order to overcome the classical problem of a lack of indicators for small areas, they produce 'synthetic estimates' of aspects of SC linking coefficients from MM in the national survey to Census data. They use the Health and Lifestyle Survey of England for data on health and the General Household Survey (GHS), BHPS and English Housing Survey (EHS) for data about SC. They incorporate individual attributes, health-related behaviours, area measures of deprivation, and area measures of SC. They find little support, at this small level, for the hypothesis that SC exerts a beneficial effect on health outcomes. I have suggested that few studies report proper small area estimates of SC and its dimensions. I will now explain how I attempt to carry out this synthetic estimation.

4.3 Multilevel Models and synthetic Small Area Estimation

There are several main empirical works to which I refer for this stage in terms of quantity and topics involved, as we can see from the literature discussed so far. Indeed, many use the applications of the chosen methodology – *models combining individual and area level covariates, starting from multilevel models.* These works are also useful as sources about the choice of the area level covariates. Beyond those listed in the previous section, given the involvement of this thesis in the ONS project, I naturally look at two of the first works produced by their Small Area Estimation Project Report: Heady *et al*. (2003) and Goldring *et al*. (2005). In this project, several ranges of measures are estimated using a method based only on area level covariates.

One of the most widely used and important is the estimation of average gross weekly HH income at Ward level. This work has been updated through the years and its completeness, due to a wide use of area covariates at different geographical levels, makes it an important example to look at.

4.3.1 Description of the method used

One of the main studies, applied since then in many works and mainly also in this study, is Twigg *et al*. (2000). They estimate smoking and drinking indicators at ward level for England. Starting with survey data nested in a 3 levels MM, they produce the synthetic estimates using Census data. Their method has been used subsequently in a wider project from the National Centre for Social Research for the Department of Health (NatCen 2004a, 2004b, 2004c, 2004d), that is the main reference for this study. They produced four reports: a user guide plus three reports[29] on the work of estimation at small area of healthy lifestyle indicators: smoking, obesity in adults, fruit and vegetable consumption by children and adults and binge drinking. They start by fitting a MM using survey data from the Health Survey for England on health information at individual level and area-level covariates from the Census, several administrative data, Office of the Deputy Minister data, and Neighbourhood Statistics at different geographical levels according to the necessity. They then compare the results of synthetic estimates at Wards level via three methods: indirect standardisation, model combining individual and area-level covariates (Twigg estimator), model using area-level covariates only (ONS estimator).

[29] Each report addresses a stage of the empirical work. The first discusses the literature, the scoping, the setting up of the datasets and the test of the software. The second is about the estimation process, the results, the internal and external validation and the last is a summary of all whole process.

I will focus only on the model combining individual and area-level covariates, which is the method I chose to apply. I indeed consider it the most complete model given the availability of the data I have.

Taking like example the estimation of synthetic area-level predictors that mostly increased the odds of smoking, according to this method, they firstly run a multilevel logistic regression where dependent variable is about whether or not adults in the 2002 to 2002 Health Survey for England (HSfE) data. The MM has like predictors individual-level data from the survey: 56 categories defined by sex, age and marital status and predictors at area-level from the different data sources listed above. The geographical level chosen was the ward, for an easier match between the two datasets and the relative availability of data. Only covariates that were significantly associated with smoking were retained in the model. The resulting model is the optima model.

The parameter estimates from the multilevel model were applied to corresponding measures for each ward. Within each ward, the estimates of smoking prevalence for each of the 56 sub-group ere multiplied by the corresponding census counts to estimate the total number of smokers in each sub-group. Summing up the categories by group estimates gave the total number of smokers in each ward. Dividing the estimated total number of smokers by the total population of the ward finally give the synthetic estimates of smoking prevalence for each ward.

Following these works, especially this last procedure, I develop the following procedure.

I start the small area estimation process fitting MM.

The previous MM estimated in Chapter 3 has been useful in showing the most significant individual covariates for each of the three factors. Respect to these variables and their cross-tabulations, I check the same availability for the Census 2011 data (usually not more than three variables together) and the necessary geographical level available (MSOA). I decide to opt for these cross-tabulations, trying to combine the most significant individual covariates:

- Factor 1 – Membership: Age – Ethnicity – Educational level
- Factor 2 – Citizenship and Politics: Age – Gender – Educational level
- Factor 3 – Neighbourliness: Age – Health – Hours spent caring someone

Given the necessary condition that survey covariates and Census covariates have to be defined in the same way, I recode all with exactly the same combinations of sub-groups. In Appendix C, Table C1, it is possible to see all the covariates. One particular specification has to be done: unfortunately, given the availability of the Census data, the estimations at small areas for Neighbourliness - Factor 3 of SC - have been reduced for the over twenty-five population whereas the remaining two factors have been carried out on population over sixteen, according to the survey sample. Besides, with regard to the same problem of availability, the cohorts of age are wider than the other two cases.

I then rerun a MM with survey and Census covariates hypothesising only a Random Intercept Model form at MSOA level. According to Twigg *et al*. (2000), simple random intercept models are preferable. More sophisticated full-random models would have introduced terms that would not have helped the final generation of predictions 'the net result of this modelling strategy is that the models are fitted for purpose but in other senses limited' (Twigg *et al*., 2000, pg.1118). Besides, in light of how the MM in the previous chapter tested, none of the individual covariates shows to have significant random coefficients which indeed have been not taken into consideration. Contextual effects, instead, show to have important random coefficients. I test then for the ecological variables found significant at level 2 but they are found to be not significant. Finally, the procedure in itself, as we can see from the following table, guarantees in some way to take into account for between-areas differences with the use of all the Census cross-tabulations. Consequently, random coefficients are not necessary.

Following these specifications, individual level covariates and area-level covariates have been tested and retained only if significant in the model. Results are showed in Appendix C, Table C2, C3 and C4[30]. Once MM have been fitted, the synthetic procedure of small area estimates of three SC factors have been done according to Twigg *et al*. (2000), Mohan *et al*. (2005) and based on the 'Twigg estimator', as it has been named by the National Centre for Social Research (NatCen, 2004c).

After fitting the MM with individual survey data covariates (by sub-groups) and area-level covariates from different sources (retained only if significant) for each SC factor, I use the model parameters to estimate the average scores of factors for each of the

[30] I standardized area-level covariates in order to make the results readable and more comparable, given the different measurements.

sub-groups in each MSOA. After the merge with the Census dataset by MSOA codes, I multiply the average score of factor for each sub-group of survey data by the corresponding count of all MSOA for each sub-group of Census cross-tabulated variables. Following, I add them over the groups to obtain the overall estimates of the number of individuals that, for each MSOA, has the average score of SC factor. Dividing this number by the population of the MSOA give me the final estimates average score of the factor for each MSOA.

The procedure is summarized in the following scheme with reference to NatCen (2004c):

Figure 4.1: Predicting process for synthetic estimates generated by a model combining individual and area-level covariates for the three SC factors

Modelling stage – UKHLS survey data		
Group the respondents according to age/ethnicity/educational level (60 groups) for Factor 1 Membership	Group the respondents according to age/gender/educational level (50 groups) for Factor 2 Citizenship and Politics	Group the respondents according to age/health/hours spent caring someone (27 groups) for Factor 3 Neighbourliness
⇩	⇩	⇩
Fit model of individual Membership SC level using area-level variables and a dummy variable representing the 60 age/ethnicity/educational level groups	Fit model of individual Citizenship and Politics SC level using area-level variables and a dummy variable representing the 50 age/gender/educational level groups	Fit model of individual Neighbourliness SC level using area-level variables and a dummy variable representing the 27 age/health/hours spent caring someone groups
⇩	⇩	⇩
Implementation stage (external data for 7,201 MSOAs)		
Use MM model parameters to estimate average scores of Membership SC in each of the 60 sub-groups in each MSOA	Use MM model parameters to estimate average scores of Citizenship and Politics SC in each of the 50 sub-groups in each MSOA	Use MM model parameters to estimate average scores of Neighbourliness SC in each of the 27 sub-groups in each MSOA
⇩	⇩	⇩
Multiply the appropriate sub-group average score by the corresponding MSOA count to estimate the total number of individual with that average score in each sub-group		
⇩	⇩	⇩
Add these over the age/ethnicity/educational level groups to obtain an overall estimate of the number of individuals with average score of Membership SC in the MSOA	Add these over the age/gender/educational level groups to obtain an overall estimate of the number of individuals with average score of Citizenship and Politics SC in the MSOA	Add these over the age/health/hours spent caring someone groups to obtain an overall estimate of the number of individuals with average score of Neighbourliness SC in the MSOA
⇩	⇩	⇩
To estimate the average score of Membership SC in the MSOA divide the estimated number in the MSOA by the corresponding MSOA population count	To estimate the average score of Citizenship and Politics SC in the MSOA divide the estimated number in the MSOA by the corresponding MSOA population count	To estimate the average score of Neighbourliness SC in the MSOA divide the estimated number in the MSOA by the corresponding MSOA population count

Source: personal elaboration on NatCen (2004c)

4.3.2 Factor 1 – Membership: final synthetic small area estimates

Once the procedure described in the previous section has been followed, synthetic estimates of average Membership SC factor for each MSOA have been obtained. In Appendix C, Table C2, estimates of the first stage with survey data are reported. As we can see, a general trend is a negative correlation with the lowest levels of qualifications and the White ethnicity and youngest cohorts of age, similar to the results found for the MM tested in the previous chapter. Original individual covariates from the survey still resulting significant are only Not White and Being religious with the same sign found in the MM models.

Census area-level covariates referred to the same categories used in the MM (defined according to the Table B4, Appendix B) have been tested as well but not one of them resulted significant. Other area-level covariates have been added and the main resulted significant is the general Index of Deprivation Score and some of its sub-dimensions. Despite a general positive sign but with a very low coefficient of the general Index as well as the Environmental attitudes sub-dimension, the other sub-dimensions show expected results: people living in areas with higher scores on deprivation in dimensions like Income, Employment and Education have a lower level of Membership SC than the average.

The ICC for this model is 7 percent (similar to the ICC for the final Random Coefficient Model for this factor). According to the model, the variance explained at individual level is around 0.3 percent whereas the variance explained at MSOA level is around 0.02 percent.

For a fast check of the goodness of the model, histogram of residuals of level 1 is reported in the Appendix C, Figure C1. As was the case for the MM, the distribution approximates to a normal one but, as expected, the skewness is more noticeable. This may be caused by the complexity of the model.

After having described the results and having carried out the procedure, it is possible to obtain a map of the predicted synthetic estimates of average levels of Membership SC factor per MSOAs for England and Wales. These are shown in the following maps – Figure 4.2 and 4.3. For a better visualisation, the second map is a zoom on Greater London's MSOAs, given its importance and its complexity.

Some specifications necessary to interpret them are due and are valid for all the three factors. The original final average scores show a distribution that goes from negative to positive values. This does not mean that MSOAs showing negative scores have negative levels of Membership SC. In the same way, MSOAs showing scores approaching 0 do not have almost null SC. The scores indicate that there are MSOAs with lower levels of SC factor than others.

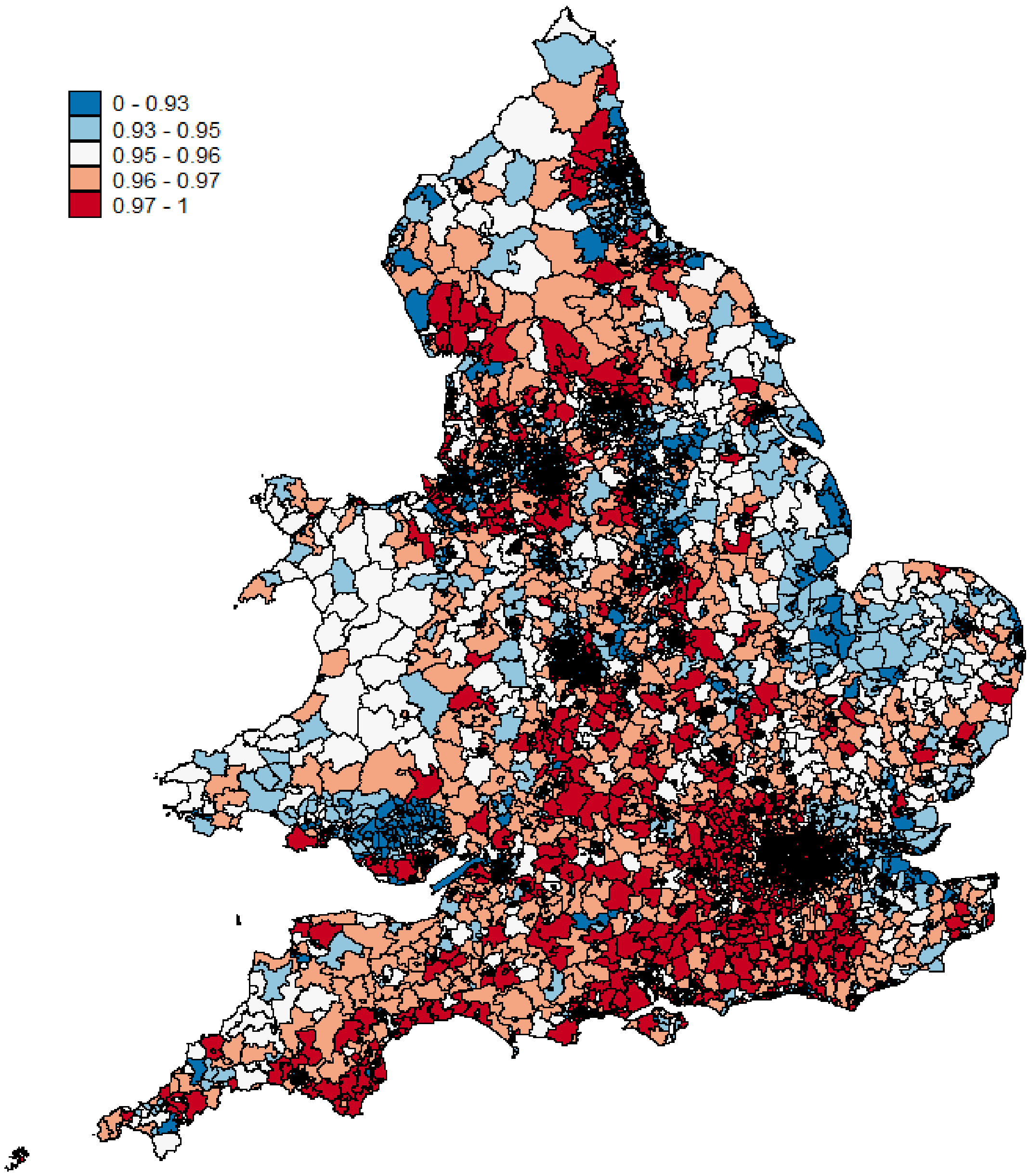
Second, to make the maps more readable I normalize the three factors, reporting quintiles[31].

Given that these estimates are synthetic and, as I will specify later, validity controls have still to be done, this is not an attempt to in some way order the MSOAs. Besides,

[31] As we can see, the range of classes of quantiles for normalized results differs because of the different final distributions.

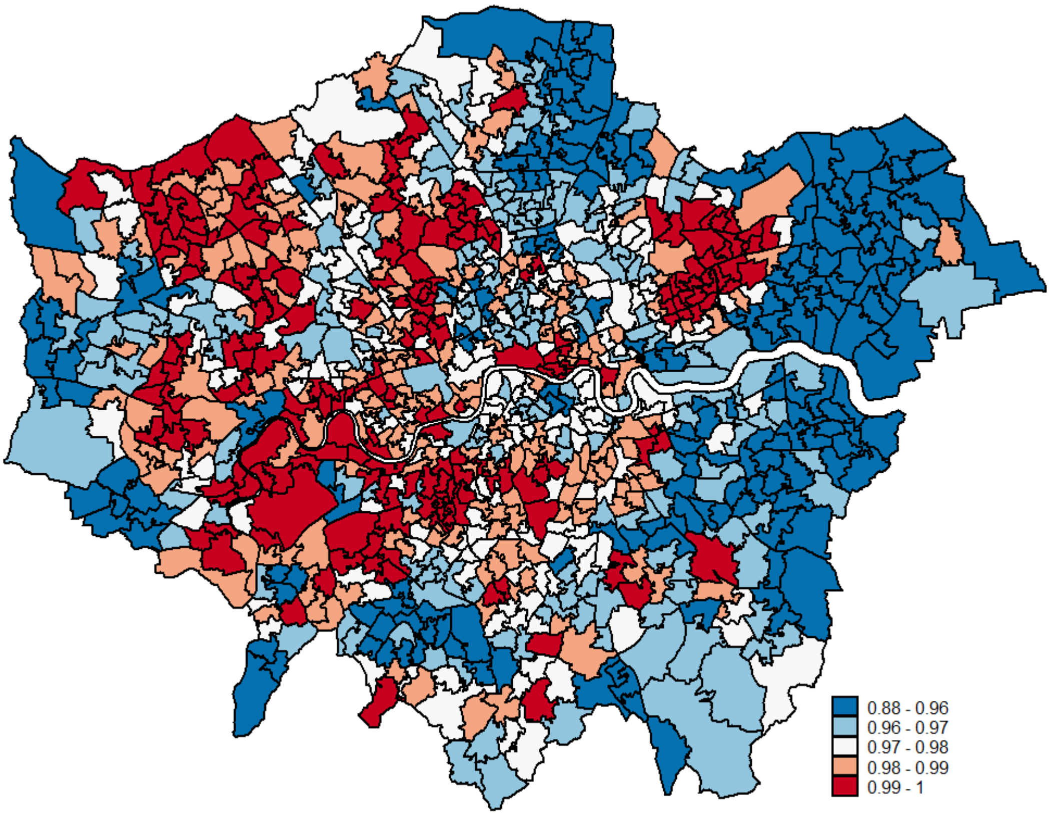
I also present a rank of the fifty areas highest and lowest in SC average for the three factors. It is not an official rank but is simply an easier way of interpreting the final small area estimates produced. Rank has to be interpreted in the same way that the maps are. Therefore, areas in the first positions will be those with the lowest levels of Membership SC until the last that are the areas with highest average levels. They are presented in Appendix C, Table C5 and C6.

Figure 4.2: MSOA synthetic estimates of Membership SC for England and Wales



As we can see, the highest average levels of Membership SC are found in the South of England (East and West) and in the region of West Midlands (especially around Birmingham) and are more dispersed. The other areas higher in the ranks are concentrated around the big cities of the northern areas: Manchester and Liverpool (North West region) and Leeds (Yorkshire and the Humber region). Looking at the map for Greater London, we can see that higher levels are concentrated around the central and western boroughs: Harrow, Brent, Camden, Richmond, Wandsworth, Lambeth and, exceptionally, the more peripheral and suburban boroughs of Hillingdon and part of Ealing.

Figure 4.3: MSOA synthetic estimates of Membership SC for Greater London

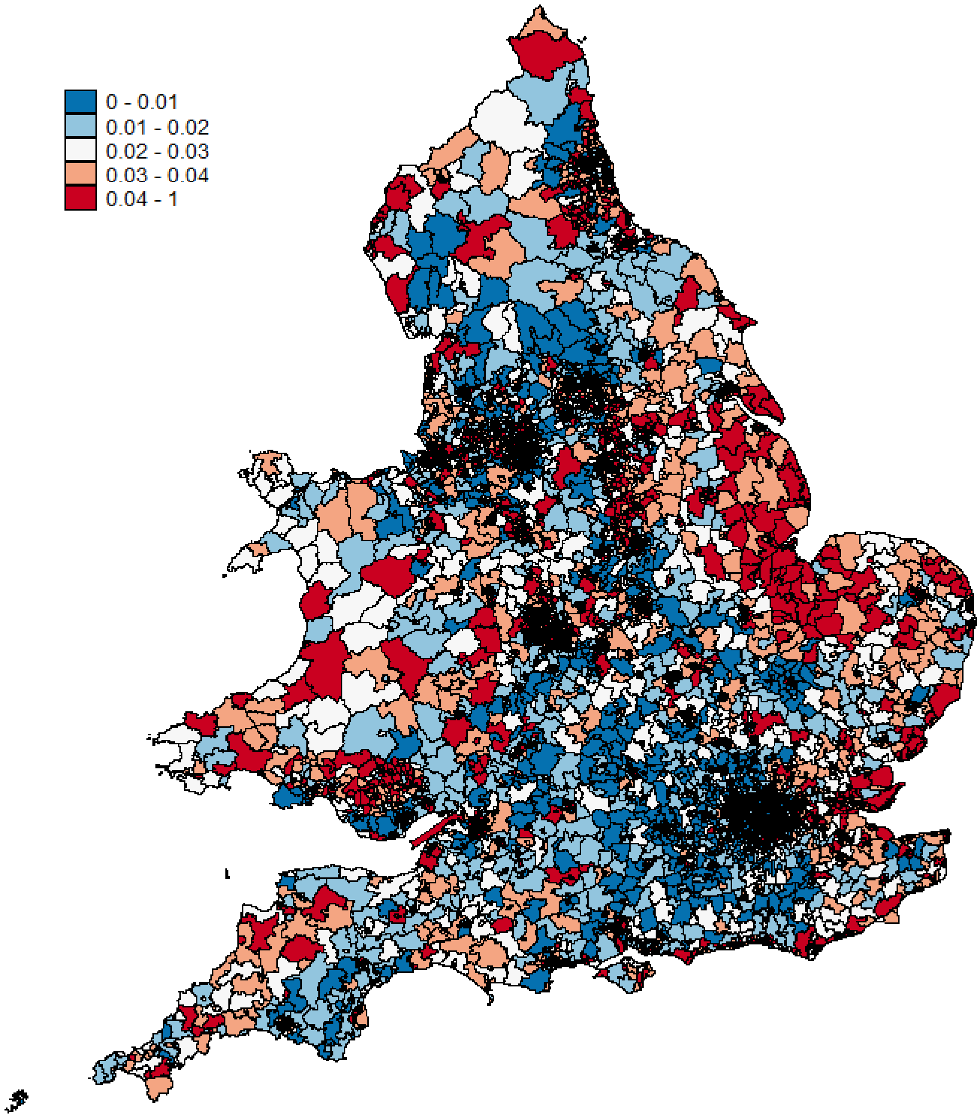


4.3.3 Factor 2 – Citizenship and Politics: final synthetic small area estimates

With the same procedure described, I obtain the estimates at small areas of average levels of Citizenship and Politics of SC. From MM, the most significant individual covariates are age, gender and qualification. The results of the MM with individual and area covariates showed in Table C3 in the Appendix highlight a general positive correlation with the increase of age and educational levels together with a higher coefficient for females. Other survey covariates still significant are: being employed, caring intensively and being not white. The change that are notable with regard to the previous MM are being not white, that turns out to be significant in this case, and being employed, which is now positively correlated.

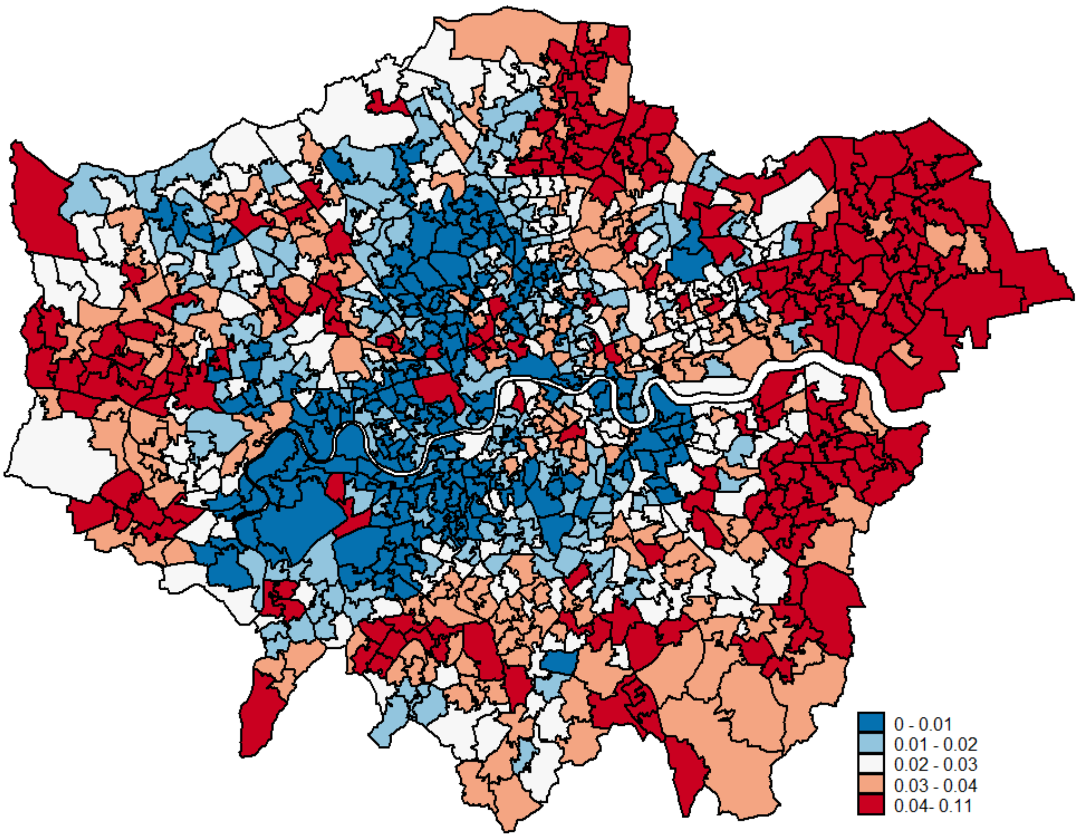
Between the area level covariates, only one result is significant and it is the sub-dimension of the Index of Deprivation regarding Housing, positively correlated but with a small coefficient. Again, Census area-level covariates referred to the same categories used in the MM have been tested too but not one of them resulted significant. Looking at the general model, the ICC for this model is around 7.5 percent, on average with the ICC of the MM of previous chapter. In the same way, the variance explained at the two levels considered is also considerably higher than Factor 1: the individual level explains the 29.5 percent while the variance explained at MSOA level is around 2 percent. Focusing on the maps of the estimates we can see that the first about England and Wales shows a different trend in relation to the previous case.

Figure 4.4: MSOA synthetic estimates of Citizenship and Politics SC for England and Wales



The areas with higher levels of SC are more concentrated in the East Midlands and East of England on one side and in Wales and the West Midlands. A few are also registered in the Somerset area, in the South West region. Looking at the Greater London map, in the following figure we can also see a different trend in this case, almost specular to the previous case: the central and western boroughs are those showing lower levels of SC than the average whereas suburban areas around the centre seem to have higher levels: parts of Hillingdon, Enfield and Waltham Forest, Redbridge and Havering, Bexley and Greenwich and few areas in the boroughs of Croydon, Sutton, Hounslow and the more central Brent.

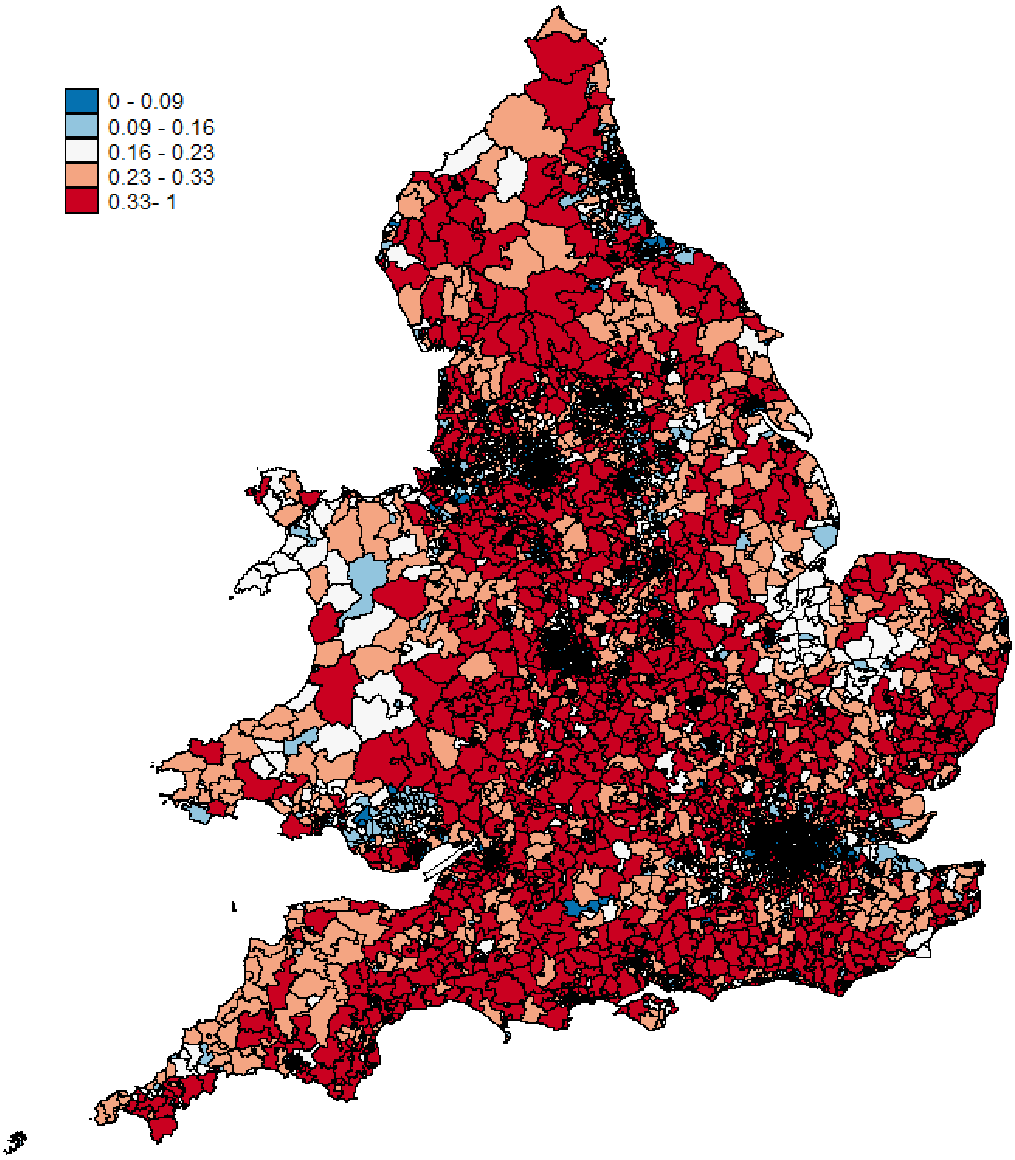
Figure 4.5: MSOA synthetic estimates of Citizenship and politics SC for Greater London



4.3.4 Factor 3 – Neighbourliness: final synthetic small area estimates

For the last factor, as expected, the trend is strongly different from the other two, as we can see from the following Figure 4.6:

Figure 4.6: MSOA synthetic estimates of Neighbourliness SC for England and Wales



This can be due to several aspects: the slightly different reference age population (due to Census availability, age cohorts start from twenty-five instead from sixteen as in the other two factors) or, more probably, the different types of SC's dimensions involved. As previously described, this SC factor loads on variables regarding more informal, private, common and spread aspects of an individual's daily life. Being a member, active or not, or being involved or interested in politics can become secondary or even optional aspects for an individual (therefore less spread or even absent). However, Neighbourliness SC refers to relations with neighbours, generalized trust, caring, fear of crime and use of social media and we can consider these aspects common dimensions of daily life, with a certain amount of it taken for granted by all individuals.

From Table C4 in the Appendix, we can see that it depends positively on the increase of the age and negatively with a poor state of health and hours spent per week caring. The only exception is for the last cohorts of age, after sixty-five, that constantly shows positive correlations and the highest coefficients no matter the intensity of caring and the state of general health.

Confirming the idea and the previous MM results that suggest that this SC factor is more linked with local aspects, more survey and area-level covariates result significant than for the previous factors. Neighbourliness SC seems to be negatively correlated with areas showing higher ethnic diversity. The contextual variable, indeed, turns to be significant in this model.

Gender is significant and in a positive way whereas negative correlation is confirmed with the single status and being employed. The correlations are various with area-level covariates from other statistics than the Census, for which covariates are again not significant on their own. While Factor 3 SC is positively correlated for extended areas but negatively with their residential density, it is negative correlated, as expected, with two covariates suggesting poorest household conditions: families with tax credits and claiming income support.

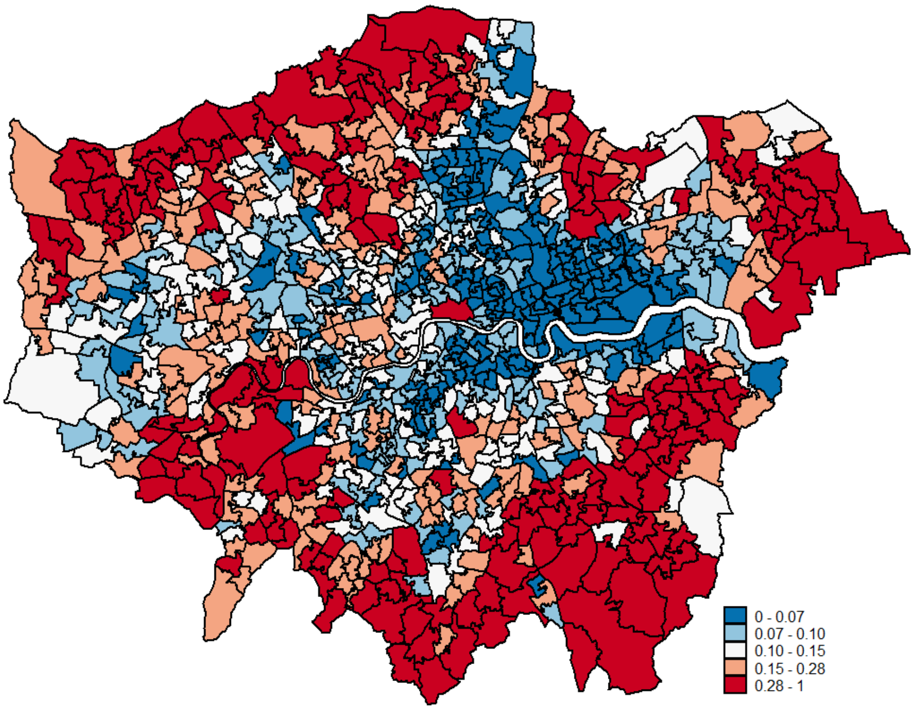
Covariates for consumption patterns show that SC is positively linked with the more 'environmental' or 'cheaper' type of consumption. Therefore, we cannot assume safely that it can be considered as an environmentally friendly attitude, usually linked with higher levels of SC. Last, only two sub-dimensions of the general Index of Deprivation result significant: Education, with an unexpected negative sign, and Income with a positive sign and a coefficient almost double than Education.

Focusing on the general model, the ICC for this factor is the highest: 8.2 percent. The variance explained at the two levels of analysis mirror the previous evidence: despite the deeper link with an area's aspects, the individual level explains the 24.5 percent of the variance whereas MSOA level the 2.2 percent. The overall fit of the model, checkable with the graph for the distribution of level 1 residuals, is the best between the three (Figure C3 in the Appendix).

Returning to the map, we can see that on average this SC is overall higher in England and more spread. Indeed, many areas are in the highest quintile of the distribution. On the other side, the lowest levels are registered in Wales and the MSOAs on the boundaries between East Midlands and East of England. Looking at the map of Greater London, the trend is less uniform than in the rest of England. We can see that there is difference between boroughs. It is more similar to the Citizenship and Politics SC factor: higher levels are registered in the suburban areas than in the centre. More specifically, Factor 3 SC is higher for boroughs in the North-West: Enfield, Barnet, Harrow and north Hillingdon. Then we have Redbridge and the North-East Borough of

Havering. Southern boroughs with higher levels than the average are Bexley, Bromley, parts of Croydon, Sutton, north of Kingston and Richmond.

Figure 4.7: MSOA synthetic estimates of Neighbourliness SC for Greater London



4.4 Conclusions

The small area estimates just presented are the final results of a complex process of modelling. In the current state of art, works on small area estimates of SC for England and Wales cannot be found. As Twigg *et al*. (2000) state, the feasibility of the approach of MM from national survey data calibrated with Census and area-level covariates has been confirmed. The three factors show differentials between them – in the trends and the correlations with the individual-level and area-level covariates – and between areas. While Membership and Citizenship and Politics SC factors behave differently throughout England and Wales with a various range, Neighbourliness SC factor seems to be higher on average between the MSOAs of the two countries and more spread. Once more, the factors are confirmed as complementary and deeply linked but they show peculiar and autonomous characteristics by themselves. Other considerations can be made about the many improvements and further research that has to be carried out. The first issue is about the prediction of the confidence intervals of these estimates and the check for validity. Goldring *et al*. (2005) report diagnostic checks using a plot of residuals against estimates. This kind of check has been used in the previous chapter for the MM with survey variables from which I derive these estimates. In this chapter I also used a plot of residuals of level 1, one of the basic ways to check the goodness of MM. NatCen (2004b) states that any calculus of standard errors for such complex models can be considered as a proper work to be completed. Besides, different methods of calculation of confidence intervals can be chosen, derived from the estimates or by simulation. One of the most used is by Markov Chain Monte Carlo simulation.

Other ways may be linked to the use of additional significant area-level covariates according to the availability of their cross-tabulations in the Census or other administrative sources. To check the validity, and besides, to pursue internal and external validation, other survey data with availability of information about SC are necessary. There are many surveys available including some measurement of SC (ONS, 2001): the Cultural Capital and Social Exclusion Survey, Home Office Citizenship Survey, Community Life Survey, Taking Part Survey, General Household Survey, English Housing Survey, British Crime Survey, Health Education Monitoring Survey, British Election Study, Health and Lifestyle Survey, Health Survey for England, British Social Attitudes Survey, Home Office Citizenship Survey, Citizen Audit Questionnaire, National Adult Learning Survey, English Longitudinal Study of Ageing, English Housing Condition Survey, Poverty and Social Exclusion Survey, and the UK Time Use Survey but there is a common problem underpinning their use. In contrast to other topics, given its complexity and multidimensionality, SC is measured in these surveys in different ways and definitions, for samples and years that cannot always be compared. Consequently, it can be difficult identifying alternative comparable sources. In any case, all the modelling work on validation and confidence of intervals calculus can be an interesting open question for further works.

Appendix C

Table C1: Description of individual-level covariates and area-level covariates

Variable	Definition	Level	Source
Being religious	Defined like belonging to a religion from the original categorical v.	Individual	UKHLS 2011
Care 50 hrs and more per week	Sum of 20-34/35-49/50-99/100 more hrs per week spent in unpaid caring someone from the original ordinal v.	Individual	UKHLS 2011
Male	Gender	Individual	UKHLS 2011
Not White	Sum of Black=Caribbean+ African+Any other Black background+ Asian=Indian+Pakistani+Bangladeshi+Chinese+Any other Asian background+ Mixed=White and Black Caribbean+White and Black African+White and Asian+Any other Mixed Background+ Any other ethnic group=Arabic+Any other ethnic group	Individual	UKHLS 2011
Single	Sum of single, never married or in a legal civil partnership, divorced, widowed, separated, surviving partner, ex-civil partner from the original categorical v.	Individual	UKHLS 2011
Employed	Sum of Paid Employed and Self Employed from the original categorical v.	Individual	UKHLS 2011
Ethnicity Index	Built according to literature - see par. 3.3.4, chapter 3	MSOA	Census 2011
Consumption Domestic Electricity_Economy 7	Average Consumption of Economy 7 Domestic Electricity	MSOA	Office for the Deputy Prime Minister
Consumption Domestic Electricity_ordinary	Average Consumption of Ordinary Domestic Electricity	MSOA	Office for the Deputy Prime Minister
Density	Density (Number of Persons per Hectare)		Office for the Deputy Prime Minister
Family with tax credits	Families Receiving; Tax Credit	MSOA	Office for the Deputy Prime Minister
Income support claimed	Income support claimants	MSOA	Office for the Deputy Prime Minister
IoD Education	Education skills and training score	MSOA	Office for the Deputy Prime Minister
IoD Employment	Employment Score	MSOA	Office for the Deputy Prime Minister
IoD Environmental	Living Environment Score	MSOA	Office for the Deputy Prime Minister

IoD General	Index of Multiple Deprivation Score	MSOA	Office for the Deputy Prime Minister
IoD Housing	Barriers to Housing and Services Score	MSOA	Office for the Deputy Prime Minister
IoD Income	Income Score	MSOA	Office for the Deputy Prime Minister
Area – hectars	Area (Hectares)	MSOA	Office for the Deputy Prime Minister

Table C2: Factor 1 – Membership parameter estimates

Variable	Estimate	Std. Err.	P-Value
Constant	-0.031	0.002	0.000
16-24 White No qualifications	-0.030	0.010	0.002
25-34 White No qualifications	-0.025	0.006	0.000
35-49 White No qualifications	-0.030	0.004	0.000
50-64 White No qualifications	-0.024	0.002	0.000
65-74 White No qualifications	-0.016	0.002	0.000
75more White No qualifications	-0.005	0.002	0.024
16-24 White GSCE level	-0.023	0.003	0.000
25-34 White GSCE level	-0.014	0.002	0.000
35-49 White GSCE level	-0.012	0.002	0.000
50-64 White GSCE level	-0.007	0.002	0.000
65-74 White GSCE level	0.020	0.003	0.000
75more White GSCE level	0.025	0.004	0.000
16-24 White A level	-0.010	0.002	0.000
25-34 White A level	-0.008	0.002	0.001
35-49 White A level	-0.002	0.002	0.284
50-64 White A level	-0.002	0.002	0.242
65-74 White A level	0.015	0.003	0.000
75more White A level	0.005	0.004	0.206
16-24 White Degree and higher	-0.005	0.003	0.129
25-34 White Degree and higher	0.004	0.002	0.021
35-49 White Degree and higher	0.017	0.002	0.000
50-64 White Degree and higher	0.025	0.002	0.000
65-74 White Degree and higher	0.039	0.002	0.000
75more White Degree and higher	0.047	0.003	0.000
16-24 White Other qualifications	-0.005	0.008	0.510
25-34 White Other qualifications	-0.005	0.006	0.365
35-49 White Other qualifications	-0.018	0.003	0.000
50-64 White Other qualifications	-0.011	0.002	0.000
65-74 White Other qualifications	-0.003	0.003	0.348
75more White Other qualifications	0.012	0.003	0.000
16-24 not White No qualifications	-0.023	0.010	0.024

25-34 not White No qualifications	-0.027	0.007	0.000
35-49 not White No qualifications	-0.008	0.004	0.054
50-64 not White No qualifications	-0.010	0.004	0.024
65-74 not White No qualifications	0.002	0.006	0.797
75more not White No qualifications	0.011	0.007	0.118
16-24 not White GSCE level	-0.024	0.004	0.000
25-34 not White GSCE level	-0.026	0.005	0.000
35-49 not White GSCE level	-0.011	0.004	0.004
50-64 not White GSCE level	0.009	0.005	0.096
65-74 not White GSCE level	0.010	0.010	0.302
75more not White GSCE level	0.008	0.018	0.661
16-24 not White A level	-0.016	0.003	0.000
25-34 not White A level	-0.010	0.005	0.034
35-49 not White A level	-0.006	0.004	0.149
50-64 not White A level	0.009	0.006	0.105
65-74 not White A level	-0.015	0.012	0.203
75more not White A level	0.027	0.018	0.135
16-24 not White Degree and higher	-0.013	0.004	0.002
25-34 not White Degree and higher	-0.009	0.003	0.005
35-49 not White Degree and higher	0.007	0.003	0.009
50-64 not White Degree and higher	0.010	0.004	0.008
65-74 not White Degree and higher	0.010	0.006	0.109
75more not White Degree and higher	0.034	0.011	0.001
16-24 not White Other qualifications	-0.023	0.010	0.017
25-34 not White Other qualifications	-0.009	0.007	0.205
35-49 not White Other qualifications	-0.007	0.005	0.118
50-64 not White Other qualifications	0.000	0.006	0.950
65-74 not White Other qualifications	0.007	0.009	0.442
75more not White Other qualifications	0.026	0.014	0.061
Not White	0.028	0.002	0.000
Being religious	0.073	0.001	0.000
IoD General	0.017	0.003	0.000
IoD Income	-0.007	0.002	0.000
IoD Employment	-0.247	0.048	0.000
IoD Education	-0.005	0.000	0.000
IoD Environmental	0.002	0.000	0.010

Figure C1: Factor 1 – Diagnostics plots: Residuals distributions of Level 1 for Random Intercept Model

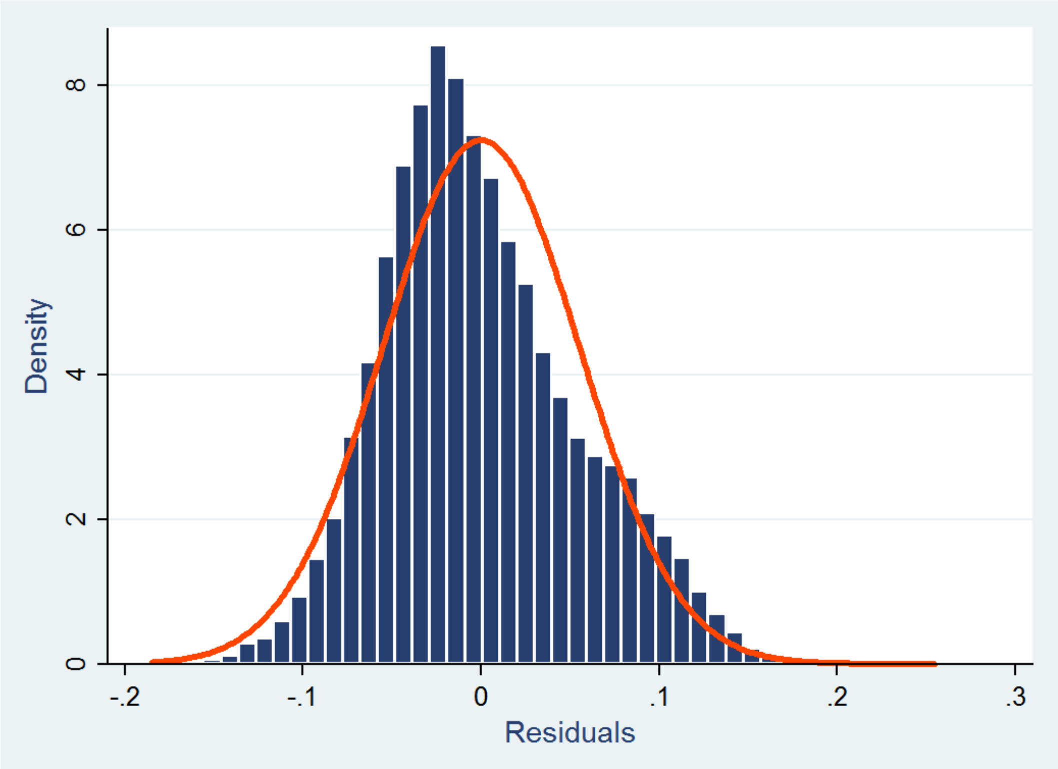


Table C3: Factor 2 – Citizenship and Politics parameter estimates

Variable	Estimate	Std. Err.	P-Value
Constant	-0.439	0.228	0.054
16-24 No qualifications Male	0.389	0.235	0.098
25-34 No qualifications Male	0.354	0.231	0.124
35-49 No qualifications Male	0.456	0.228	0.045
50-64 No qualifications Male	0.542	0.229	0.018
65more No qualifications Male	0.727	0.228	0.001
16-24 GSCE level Male	0.369	0.229	0.106
25-34 GSCE level Male	0.487	0.228	0.033
35-49 GSCE level Male	0.554	0.228	0.015
50-64 GSCE level Male	0.722	0.229	0.002
65more GSCE level Male	0.897	0.229	0.000
16-24 A level Male	0.519	0.228	0.023
25-34 A level Male	0.581	0.229	0.011
35-49 A level Male	0.669	0.228	0.003
50-64 A level Male	0.775	0.228	0.001
65more A level Male	0.880	0.229	0.000
16-24 Degree and higher Male	0.684	0.229	0.003
25-34 Degree and higher Male	0.761	0.228	0.001
35-49 Degree and higher Male	0.872	0.228	0.000

50-64 Degree and higher Male	0.971	0.228	0.000
65more Degree and higher Male	1.085	0.228	0.000
16-24 Other qualifications Male	0.292	0.236	0.216
25-34 Other qualifications Male	0.239	0.233	0.306
35-49 Other qualifications Male	0.437	0.229	0.056
50-64 Other qualifications Male	0.636	0.229	0.005
65more Other qualifications Male	0.876	0.229	0.000
16-24 No qualifications Female	0.315	0.238	0.185
25-34 No qualifications Female	0.227	0.232	0.327
35-49 No qualifications Female	0.304	0.229	0.184
50-64 No qualifications Female	0.377	0.229	0.099
65more No qualifications Female	0.627	0.228	0.006
16-24 GSCE level Female	0.147	0.229	0.520
25-34 GSCE level Female	0.241	0.229	0.293
35-49 GSCE level Female	0.418	0.228	0.067
50-64 GSCE level Female	0.580	0.228	0.011
65more GSCE level Female	0.893	0.229	0.000
16-24 A level Female	0.428	0.228	0.061
25-34 A level Female	0.410	0.229	0.073
35-49 A level Female	0.516	0.228	0.024
50-64 A level Female	0.691	0.229	0.003
65more A level Female	0.836	0.231	0.000
16-24 Degree and higher Female	0.533	0.229	0.020
25-34 Degree and higher Female	0.642	0.228	0.005
35-49 Degree and higher Female	0.748	0.228	0.001
50-64 Degree and higher Female	0.912	0.228	0.000
65more Degree and higher Female	0.980	0.229	0.000
16-24 Other qualifications Female	0.080	0.237	0.735
25-34 Other qualifications Female	0.231	0.233	0.321
35-49 Other qualifications Female	0.268	0.229	0.241
50-64 Other qualifications Female	0.550	0.229	0.016
65more Other qualifications Female	0.779	0.229	0.001
Employed	0.040	0.008	0.000
Care 50 hrs and more per week	-0.001	0.000	0.000
Not White	0.121	0.008	0.000
IoD Housing	0.026	0.006	0.000

Figure C2: Factor 2 – Diagnostics plots: Residuals distributions of Level 1 for Random Intercept Model

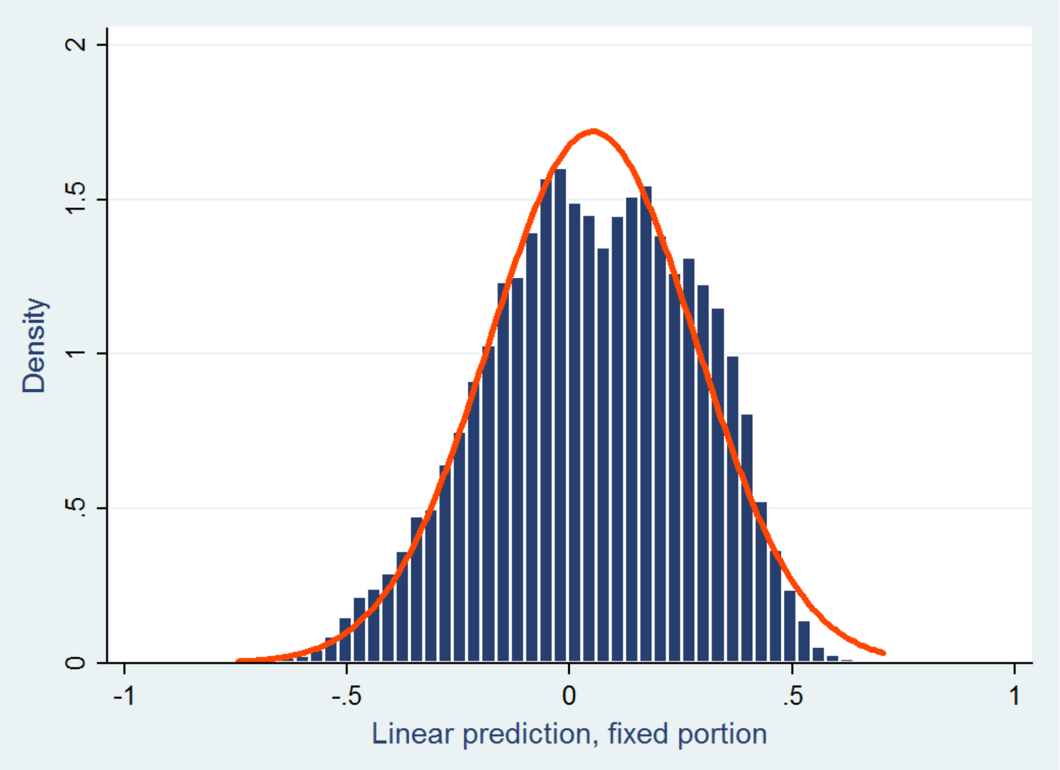


Table C4: Factor 3 – Neighbouring parameter estimates

Variable	Estimate	Std. Err.	P-Value
Constant	-0.062	0.013	0.000
25-49 Good health 1-19 hrs per week	0.026	0.014	0.057
50-64 Good health 1-19 hrs per week	0.137	0.014	0.000
65more Good health 1-19 hrs per week	0.270	0.019	0.000
25-49 Fair health 1-19 hrs per week	-0.076	0.033	0.020
50-64 Fair health 1-19 hrs per week	0.049	0.032	0.125
65more Fair health 1-19 hrs per week	0.178	0.036	0.000
25-49 Poor health 1-19 hrs per week	-0.131	0.064	0.041
50-64 Poor health 1-19 hrs per week	-0.076	0.057	0.179
65more Poor health 1-19 hrs per week	-0.040	0.070	0.566
25-49 Good health 20-49 hrs per week	-0.067	0.032	0.034
50-64 Good health 20-49 hrs per week	0.092	0.033	0.005
65more Good health 20-49 hrs per week	0.219	0.040	0.000
25-49 Fair health 20-49 hrs per week	-0.060	0.064	0.352
50-64 Fair health 20-49 hrs per week	0.030	0.061	0.626
65more Fair health 20-49 hrs per week	0.024	0.065	0.711
25-49 Poor health 20-49 hrs per week	-0.426	0.116	0.000
50-64 Poor health 20-49 hrs per week	-0.040	0.093	0.668

65more Poor health 20-49 hrs per week	-0.130	0.102	0.202
25-49 Good health 50 more hrs per week	-0.096	0.037	0.010
50-64 Good health 50 more hrs per week	0.080	0.051	0.112
65more Good health 50 more hrs per week	0.185	0.043	0.000
25-49 Fair health 50 more hrs per week	-0.169	0.060	0.005
50-64 Fair health 50 more hrs per week	0.076	0.066	0.253
65more Fair health 50 more hrs per week	0.061	0.060	0.309
25-49 Poor health 50 more hrs per week	-0.206	0.081	0.011
50-64 Poor health 50 more hrs per week	-0.094	0.082	0.252
65more Poor health 50 more hrs per week	0.160	0.113	0.157
Ethnicity Index	-0.120	0.026	0.000
Male	0.108	0.009	0.000
Single	-0.094	0.008	0.000
Employed	-0.026	0.006	0.000
Area – hectares	0.020	0.004	0.000
Density	-0.018	0.006	0.003
Family with tax credits	-0.051	0.009	0.000
Income support claimed	-0.036	0.009	0.000
Consumption Domestic Electricity_ordinary	-0.031	0.006	0.000
Consumption Domestic Electricity_Economy 7	0.025	0.005	0.000
IoD Income	0.043	0.107	0.000
IoD Education	-0.027	0.001	0.000

Figure C3: Factor 3 – Diagnostics plots: Residuals distributions of Level 1 for Random Intercept Model

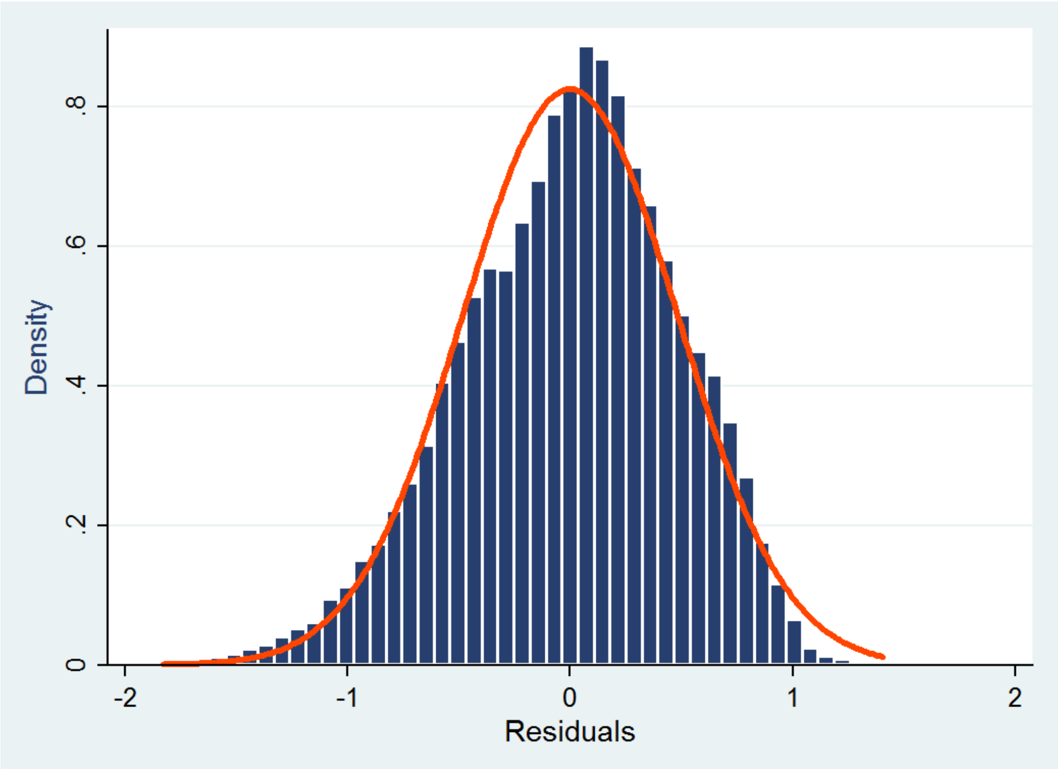


Table C5: MSOAs' Rank – 50 areas with highest levels of SC by factors

Rank	Factor 1 - Membership	Factor 2 - Citizenship and Politics	Factor 3 - Neighbouring
1	North Somerset 002	Canterbury 013	East Dorset 011
2	Cheltenham 014	Bristol 032	South Lakeland 010
3	Three Rivers 004	Sheffield 073	Arun 018
4	Derby 002	Coventry 042	West Dorset 004
5	Guildford 003	Nottingham 028	New Forest 001
6	Bournemouth 020	Manchester 022	Cotswold 004
7	Hart 007	York 023	Chichester 013
8	Harrow 016	Cardiff 025	Horsham 012
9	Merton 002	Canterbury 012	Cheshire West and Chester 015
10	Three Rivers 011	Sheffield 028	Wakefield 034
11	Cheshire East 010	Brighton and Hove 002	Havant 013
12	Bradford 001	Sheffield 038	Suffolk Coastal 002
13	Isle of Wight 010	Plymouth 023	Derbyshire Dales 002
14	Bristol 006	Exeter 004	Sedgemoor 012
15	Waverley 004	Swansea 026	South Cambridgeshire 016
16	Eastleigh 001	Sheffield 042	East Dorset 012
17	South Hams 012	Leicester 040	Bradford 002
18	Sefton 009	Birmingham 079	Chiltern 001
19	North Tyneside 001	Manchester 026	Solihull 026

20	Woking 002	Charnwood 003	New Forest 017
21	Worthing 013	Lincoln 005	Wychavon 019
22	Cheshire East 003	Newcastle upon Tyne 017	Wrexham 020
23	Stroud 002	Birmingham 050	York 024
24	Kensington and Chelsea 018	Lancaster 019	Wirral 023
25	Bristol 011	Nottingham 026	Bath and North East Somerset 016
26	Darlington 011	Nottingham 022	Pendle 008
27	Cheshire East 012	Sheffield 074	Stroud 002
28	Cardiff 004	Cambridge 005	Cheshire East 020
29	Sefton 012	Newcastle upon Tyne 023	Darlington 011
30	Arun 018	Sheffield 036	Sheffield 041
31	Shepway 008	Sheffield 030	Bath and North East Somerset 010
32	Wirral 040	Liverpool 038	South Staffordshire 013
33	Bristol 009	Manchester 036	East Hampshire 011
34	Swansea 031	Gwynedd 001	Rochford 003
35	Christchurch 007	Liverpool 060	Poole 003
36	Horsham 012	Cardiff 028	Rother 009
37	Harrow 001	Cambridge 007	Lewes 002
38	Mid Sussex 008	Nottingham 032	Guildford 003
39	North Somerset 005	Ceredigion 002	New Forest 023
40	Arun 008	Charnwood 007	Chesterfield 011
41	Rother 010	Leeds 063	Suffolk Coastal 013
42	Swansea 028	Nottingham 031	Lichfield 005
43	East Devon 017	Newcastle upon Tyne 013	East Hampshire 007
44	Eastbourne 012	Leeds 054	Sunderland 022
45	Rother 009	Leeds 044	Herefordshire 023
46	East Devon 012	Cardiff 032	Sheffield 068
47	Christchurch 003	Leeds 110	Hambleton 002
48	New Forest 023	Manchester 032	Solihull 028
49	New Forest 022	Oxford 008	Sheffield 055
50	Poole 018	County Durham 030	Stafford 014

Table C6: MSOAs' Rank – 50 areas with lowest levels of SC by factors

Rank	Factor 1 - Membership	Factor 2 - Citizenship and Politics	Factor 3 - Neighbouring
1	County Durham 030	Camden 011	Birmingham 051
2	Birmingham 096	Haringey 035	Birmingham 070
3	Leeds 110	Tower Hamlets 027	Oldham 016
4	Manchester 032	Tower Hamlets 033	Birmingham 139
5	Oxford 008	Camden 016	Leicester 018
6	Cardiff 032	Richmond upon Thames 008	Birmingham 082
7	Leeds 044	Lambeth 019	Nottingham 011
8	Leeds 054	Haringey 033	Bradford 051

9	Newcastle upon Tyne 013	Richmond upon Thames 007	Birmingham 077
10	Ceredigion 002	Wandsworth 009	Middlesbrough 001
11	Leeds 063	City of London 001	Sheffield 022
12	Gwynedd 001	Camden 002	Leicester 026
13	Nottingham 031	Wandsworth 015	Birmingham 054
14	Cardiff 028	Camden 008	Birmingham 083
15	Liverpool 038	Richmond upon Thames 006	Bradford 044
16	Newcastle upon Tyne 023	Haringey 022	Coventry 015
17	Sheffield 036	Oxford 002	Birmingham 063
18	Lincoln 005	Merton 004	Bradford 042
19	Charnwood 007	Wandsworth 017	Hyndburn 006
20	Manchester 036	Haringey 021	Birmingham 055
21	Sheffield 030	Richmond upon Thames 009	Bradford 048
22	Newcastle upon Tyne 017	Hammersmith and Fulham 022	Leicester 035
23	Liverpool 060	Hammersmith and Fulham 024	Newham 019
24	Lancaster 019	Wandsworth 006	Redcar and Cleveland 009
25	Cambridge 007	Wandsworth 021	Birmingham 052
26	Nottingham 032	Haringey 034	Birmingham 048
27	Nottingham 022	Lambeth 013	Luton 015
28	Swansea 026	Southwark 031	Birmingham 037
29	Exeter 004	Haringey 036	Leeds 048
30	Charnwood 003	Oxford 012	Manchester 012
31	Sheffield 042	Ealing 034	Bradford 038
32	Plymouth 023	Oxford 001	Wirral 011
33	Birmingham 079	Wandsworth 010	Birmingham 040
34	Sheffield 074	Merton 009	Manchester 058
35	Sheffield 038	Greenwich 021	Medway 022
36	Sheffield 028	Camden 014	Bradford 039
37	Cardiff 025	Guildford 016	Wigan 010
38	Cambridge 005	Hammersmith and Fulham 014	Newcastle upon Tyne 028
39	Plymouth 027	St Albans 011	Manchester 015
40	Brighton and Hove 002	Ealing 013	Bradford 041
41	Nottingham 028	Merton 003	Rochdale 015
42	York 023	Haringey 009	Calderdale 012
43	Cheshire West and Chester 034	Wandsworth 029	Coventry 009
44	Manchester 026	Merton 002	Manchester 018
45	Southampton 023	Haringey 030	Rushmoor 008
46	Canterbury 013	Westminster 016	Bradford 034
47	Southampton 017	Hammersmith and Fulham 011	Swansea 011
48	Nottingham 026	Lewisham 009	Southwark 019
49	Bristol 032	Wandsworth 030	Salford 017
50	Richmondshire 004	Westminster 008	Newham 030

CONCLUSIONS

In recent years, opposing trends have been registered about SC in Britain. The famous and controversial study of Hall (1999) addressing the state of SC in Britain argues that the apparent decline in social trust sees equivalent erosion of social participation. Trying to take a counter-example to Putnam's version of declining SC in USA, the author refers to other aspects as possible explanations: change of values, governmental policies, changes in social integration and so on. A subsequent study by Grenier and Wright (2003) attempts to update the results with new data and further check Hall's results. They find similar trends to the past but also state that other aspects now have to be considered to explain differences in SC levels: rise in income inequality, distributional issues and class divisions.

CIS 8 (2015) find that supportive informal relationships and social support have not in fact declined over the last two decades but that having 'someone to discuss personal matters with' is much less likely among the over seventy-fives, those with less education, and those outside the labour market. Besides, activity in voluntary organizations seems to slowly decline in the long-term and the activity related specifically to the local community or neighbourhood is declining as well (about 3.7 percent, from 11.5 percent to 7.8 percent over the last decade).

Looking at civic participation, the ethnic gap may be widening while the age gap has remained the same over time. Gaps related to education and economic activity are still confirmed as contributing to a socially unequal distribution of SC. The related social trust, not in decline, seems to be weakest in London and among some ethnic minorities and less educated people.

CIS 15 (2015) confirm that SC is ‘multidimensional’ and comes in different types and combinations according to which individuals can show higher levels on some and lower on others. Isolation seems to be rare, confirming the idea that a certain amount of SC has been accumulated in Britain. The attention to SC therefore is remains high and has been increasing since the beginning of this century when we saw the main governmental bodies at national and international level starting to focus on it: the World Bank, OECD, European Commission and ONS. In November 2015, in London, I attended a conference on the State of Social Capital in Britain. Speakers and participants from academia but also from the private and public sectors, the government and the third sector took part. The general final agreement on the topic after a day of discussion sounded more like a programmatic intent: SC is one of the most important resources that must be fed and continuously pursued in an integrated approach from all the agents at different levels. It is, indeed, one of the answers to local development and all related (and listed) benefits especially during times of scarce resources (Wilding, 2015).

Deeply linked to the centrality of SC is the issue of its measurement and the necessity for more precise and local estimates. The small area estimates methodology overcomes classical problems of having disaggregated estimates. This level of analysis, indeed, seems to allow more precise analysis in relation to previous studies where the highest levels of analysis (regional or national) might have not allowed for the capture of differentials. Just to cite an example, Van Oorschot *et al*. (2006) state that SC is highly accumulated between European countries as well as physical and human capital but that there are no substantial differences between countries and regions (except for Northern Europe and Scandinavia where it is higher) and between categories of European citizens. However, there is no detail on lower levels of aggregation.

Inference about aspects and problems for which we do not have the adequate data or coverage is probably the main strength and advantage of these methods. Moreover, the variety of approaches and methodologies available easily accommodate different types of data, increase precision, allow the derivation of 'optimal' estimates and associated measures of variability under the assumed model, and the validation of models from the sample data (Rao, 2003).

Mohan and Mohan (2002) explore the necessity of a geographical, more disaggregated, analysis of SC. In a critical study, they begin to address criticism of the concept as defined in the classical way (Putnam's definitions and so on). They then report on studies that attempt to estimate different SC dimensions at different geographical levels. They notice how these studies focus on one dimension per time and how it is difficult to find disaggregated data on SC. They then conclude that overcoming these problems would help in a practical way to structure better policies on development, poverty alleviation, social cohesion, industrial policies, educational development, health, housing, social security at different levels. They then confirm in this way, the importance of SC as real tool for governments and not only as a theoretical concept.

The many important advantages just highlighted about the application of small area estimates to SC have to be balanced with the weaknesses of these methodologies. As widely described in the last chapter of this thesis, the main problems are: the reliability and validity of such approaches, the risks around confidentiality, anonymity and disclosure, the use of different outcomes produced and the linkages between different datasets and different type of data. Indeed, on the one hand, the validation process, both internal and external, can be eased by the use of sample data, on the other hand the availability of these data and the degree of comparability between them and the synthetic estimates obtained is one of the main limitations. A weak validation could make these estimates less reliable, and less generalizable, with respect to other methods of estimation.

After all these considerations, I can state that this PhD work has been an attempt to merge and deal with all these aspects. Theoretically, the complexity and multidimensionality of this concept finds fulfilment in the several methodologies used and the factors built with the Factorial Analysis. The three factors identified perfectly fit together in a model but maintain an independent and clear structure by themselves, becoming proper variables. All the original variables representing membership and active membership to different types of organizations, institutional trust, voting attitude and behaviours, interest in politics, relations with neighbours, generalised and social trust and caring have resulted significant. The three SC dimensions identified – Membership, Citizenship and Politics and Neighbouring – perfectly load on them, therefore demonstrating a common pattern of variance between dimensions for individuals. Other variables resulted significant as well for use of social media and fear of crime.

Once identified, they resulted as correlated with all the main individual characteristics, showing interesting and, sometimes, different trends between the three factors. Age, gender, ethnicity, religion and education are the variables resulting more frequently significant, both in a positive and in a negative way. Variables representing caring, state of health, marital status, employment and its socio-economic classification, dependent children, moving and housing conditions have resulted significant in a weaker way.

Multilevel models, tested in their main forms of Random Intercept and Random Coefficient Models with Contextual Effect, also showed that the SC factors are related to area characteristics regarding mainly the ethnicity diversity and the economic profile. This last aspect has been identified by higher levels of employment, a more developed service sector and higher levels of educational attainments.

Comparing the factors, we saw how the factors of Membership and Citizenship and Politics seem to be more related (even in a contrasting way) and to have higher differentials between MSOAs in England and Wales. The third factor about Neighbouring, that I intend to be more informal and private, seems to have a different trend and to be, in any case, spread more uniformly around the countries and, on average, higher than the other two.

The analysis carried out is wide and deep but of course, there are limitations and improvements that can be achieved. Alternative data can be used as well as alternative methodologies and alternative cross-tabulations from the Census. Even the cross-sectional approach constrained by the Census availability can be revised for a longitudinal study. Validity control of estimates have still to be carried out and compared with other possible estimates. Besides, we can imagine that further dimensions of SC can still be investigated and integrated to the current study.

The aim of 'Beyond 2011' has been reached: it is possible to produce estimates at small area from multivariate and model-based approaches. I used Census covariates as well as other administrative sources. The statistical significance reached by the covariates confirms that we can use other sources than the Census for the same information. An interesting attempt would be the use of the Sample Anonymised Records (SAR) from the Census that can be easily and more frequently updated, allowing a more longitudinal perspective. This last aspect especially deserves further investigation to confirm or not the general idea that SC is long-term capital: it needs time to accumulate and even more time to change because of the dimensions included.

Expanding the analysis, SC factors and estimates can then be used as predictors on variables representing all the dimensions we saw as related to SC, from the more economic to those that are distinctly social. Also, a more qualitative or mixed approach can be used to investigate it.

I can conclude then by asserting one more time that social capital proves to be an interesting and useful concept both as an analytical tool and for understanding social trends. More than twenty years after the initial conceptualizations, it is still actual and central to the debate both at scientific and political levels. It is deeply related to a wide range of other topics across the various fields of social sciences and therefore, it is to still be considered as a fundamental asset for policies aiming to foster better local and global worlds.

BIBLIOGRAPHY

Adler P. S., Kwon S. W., (2002), Social capital: prospects for a new concept, *Academy of Management Review*, 27, [1], 17-40.

Akçomak I. S., (2009), Bridges in social capital: a review of the definitions and the social capital of social capital researchers, *UNU-MERIT Working Series*, 002.

Akçomak I. S., ter Weel B., (2012), The impact of social capital on crime: Evidence from the Netherlands, *Regional Science and Urban Economics*, 42 [1-2], 323-340.

Akçomak, I. S., ter Weel, B., (2009), Social capital, innovation and growth: Evidence from Europe, *European Economic Review*, 53 (5), 544-567.

Akçomak, I. S., ter Weel, B., (2005), How do social capital and government support affect innovation and growth? Evidence from the EU regional support programs, *UNU-MERIT Working Paper Series*, 2007-009.

Allaga A., Muhuri P. K., (1994), Methods of Estimating Contraceptive Prevalence Rates for Small Areas: Applications in The Dominican Republic and Kenya, *Demographic and Health Surveys – Methodological Report N. 3*.

Allison P. D., (2005), Imputation of Categorical Variables with PROC MI, *Sugi 30 – Focus Session*, Paper 113-30, 1-14.

Allison P. D., (2000), Multiple Imputation for Missing Data – A Cautionary Tale, *Sociological Methods & Research*, 28 (3), 301-309.

Australian Bureau of Statistics, (2002), Social Capital and Social Wellbeing, *Discussion Paper*.

Baron R. A., Markman G. D., (2003), Beyond social capital: the role of entrepreneur's social competence in their financial success, *Journal of Business Venturing*, 18 (1), 41-60.

Barrett P., Hale B., Butler M., (2014), Family care and Social Capital: Transitions in Informal Care, *Springer Eds*.

Battese G. E., Harter R. M., Fuller H. W., (1988), An Error-Components Model for Prediction of Country Crop Areas Using Survey and Satellite Data, *Journal of the American Statistical Association*, 83 [401], 28-36.

Bauböck R., (2005), Expansive Citizenship – Voting beyond Territory and Membership, *Political Science and Politics*, 38 [4], 683-687.

Bell A., Jones K., (2015), Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data, *Political Science Research and Methods*, 3 [01], 133-153.

Bentler P. M., Chou C-P., (1987), Practical issues in structural modelling, *Sociological Methods & Research*, 16, 78-117.

Beugelsdijk S., van Schaik T., (2005), Social capital and growth in European regions: an empirical test, *European Journal of Political Economy*, 21 (2), 301-324.

Bourdieu P., (1986), The forms of capital in Richardson J., *Handbook of theory and research for the sociology of education*, Greenwood, 241-258.

Bourdieu P., (1979), La distinction: critique sociale du jugement, *Paperback*.

Brehm J, Rahn W, (1997), Individual – Level Evidence for the Causes and Consequences of Social Capital, *American Journal of Political Science*, 41 (3), 999-1023.

Brunton-Smith I., (2011), Untangling the Relationship Between Fear of Crime and Perceptions of Disorder – Evidence from a Longitudinal Study of Young People in England and Wales, *The British Journal of Criminology*, 51 [6], 885-899.

Brunton-Smith I., Sturgis P., (2011), Do neighbourhoods generate fear of crime? An empirical test using the British Crime Survey, *Criminology*, 49 [2], 331-369.

Buonanno P., Montolio D., Vanin P., (2009), Does Social Capital Reduce Crime?, *Journal of Law and Economics*, 52 (1), 145-170.

Burke M., Kraut R., Marlow C., (2011), Social capital on Facebook: Differentiating users and users, *Proceedings of the 2010 Annual Conference on Human Factors in Computing Systems*, 1909-1921, New York.

Burt R. S., (2000), The network structure in social capital, *Research in Organizational Behavior*.

Burt R. S., (1992), Structural holes: the social structure of competition, *Harvard University Press*.

Byrne B. M., (2012), Structural Equation Modeling With Mplus, *Routledge Press*.

Caplan S. E., (2003), Preference for Online Social Interaction - A Theory of Problematic Internet Use and Psychosocial Well-Being, *Communication Research*, 30 (6), 625-648.

Carpiano R. M., Kimbro R. T., (2012), Neighbourhood Social Capital, Parenting Strain, and Personal Mastery among Female Primary Caregivers of Children, *Journal of Health and Social Behaviour*, 53 (2), 232-247.

Carpiano R. M., (2007), Neighborhood social capital and adult health: An empirical test of a Bourdieu-based model, *Health & Place*, 13, 639-655.

Castiglione D., Van Deth J. W., Wolleb G., (2008), The Handbook of Social Capital, *Oxford University Press*.

Cattell V., (2001), Poor people, poor places, and poor health: the mediating role of social networks and social capital, *Social Science & Medicine*, 52 (10), 1501-1516.

Chambers R., Tzavidis N., (2006), M-quantile models for small area estimation, *Biometrika*, 93, 255-268.

Chambers R., Tzavidi N., (2005), Using Multilevel Models for Small Area Estimation, *NCRM Summer School*.

Chou Y. K., (2003), Modelling the impact of network social capital on business and technological innovations, *University of Melbourne – Department of Economics*, Research Paper n. 890.

Claibourn M. P., Martin P. S., (2000), Trusting and Joining? An empirical test of the reciprocal nature of social capital, *Political Behaviour*, 22 [4], 267-291.

Coleman J. S. (1990), Foundations of social theory, *Belknap Press of Harvard University*.

Coleman J. S., (1988), Social capital in the creation of human-capital, *American Journal of Sociology*, n. 94, S95-S120.

Costa D. L., Kahn M. E., (2001), Understanding the decline in social capital, 1952-1998, *NBER Working Paper*, 8295.

Criscuolo C., Haskel J. E., Slaughter M. J., (2010), Global engagement and the innovation activities of the firm, *International Journal of Industrial Organization*, 28, [2], 191-202.

CSI 15, (2015), The uneven distribution and decline of social capital in Britain, *Centre for Social Investigation Nuffield College Report*, Oxford.

CSI 8, (2015), Social Capital – Are we becoming lonelier and less civic?, *Centre for Social Investigation Nuffield College Report*, Oxford.

Cutts D., Fieldhouse E., (2015), Diversity and social capital in the US and UK: the role of bridging friendships, in Li Y., (2015), Handbook of Research Methods and Applications in Social Capital, *Elgar Edition*, chapter 7.

Dakhli, M., De Clercq D., (2004), Human capital, social capital, and innovation: a multicountry study, *Entrepreneurship & Regional Development*, 16, 107-128.

Dasgupta P. (2005), Economics of social capital, *Economic Record*, 81, S2-S21.

David Q., Janiak A., Wasmer E., (2010), Local social capital and geographical mobility, *Journal of Urban Economics*, 68 (2), 191-204.

De Souza Briggs X., (1998), Brown Kids in White Suburbs: Housing Mobility and the Many Faces of Social Capital, *Housing Policy Debate - U.S. Department of Housing and Urban Development and Harvard University*, 9 (1), 177-221.

Dedrick R. F., Ferron J. M., Hess M. R., Hogarty K. Y., Kromrey J. D., Lang T. R., Niles J. D., Lee R. S., (2009), Multilevel Modelling: A Review of Methodological Issues and Applications, *Review of Educational Research*, 79 [1], 69-102.

Di Stefano C., Hess B., (2005), Using Confirmatory Factor Analysis for Construct Validation: An Empirical Review, *Journal of Psychological Assessment*, 23, 225-241.

Di Stefano C., Zhu M., Mîndrilă D, (2009), Understanding and Using Factor Scores: Considerations for the Applied Researchers, *Practical Assessment, Research and Evaluation*, 14 (20).

Douglas A., (2004), Chapter 2 in Multilevel Modeling*, Sage*.

Duke N. N., Skay C. L., Pettingell S. L., Borowsky I. W., (2009), From Adolescent Connections to Social Capital: Predictors of Civic Engagement in Young Adulthood, *Journal of Adolescent Health*, 44 (2), 161-168.

Duncan C., Jones K., Moon G., (1993), Health related behaviour in context – a multi level modelling approach, *Social Science and Medicine*, 42 (6), 817-830.

Ebstyne King P., Furrow J. L., (2004), Religion as a Resource for Positive Youth Development: Religion, Social Capital and Moral Outcomes, *Developmental Psychology*, 40 (5), 703-713.

Ellison N. B., Steinfield C., Lampe C., (2007), The benefits of Facebook friends: Social Capital and college students' use of online social network sites, *Journal of Computer – Mediated Communication*, 12, 1143–1168.

Ellison N. B., Vitak J., Gray R., Lampe C., (2014), Cultivating Social Resources on Social Network Sites: Facebook Relationship Maintenance Behaviours and Their Role in Social Capital Processes, *Journal of Computer – Mediated Communication*, 19 (4), 855-870.

Erickson B. H., (2004), The distribution of gendered social capital in Canada, in Creation and Returns of Social Capital – A new research program, edited by Flap H. and Völker B, *Routledge*.

Fay R. E., Herriot R. A., (1979), Estimates of income for small places: an application of James-Stein procedure to Census data, *Journal of the American Statistical Association*, 74 [366a], 269-277.

Fielding A., (2004), The Role of the Hausman Test and wheter Higher Level Effects should be treated as Random or Fixed, *Multilevel Modelling Newsletter*, 16 [2], 3-9.

Fine B. (2010), Theories of Social Capital: Researchers Behaving Badly, *Pluto Press*.

Geiser C. (2013), Data Analysis with Mplus, *The Guildford Press*, London.

Glaeser E. L., Laibson D., Sacerdote B., (2002), An Economic Approach to Social Capital, The Economic *Journal*, 112, F437-F458.

Goldin C., Katz L. F., (1999), Human capital and social capital: the rise of secondary schooling in America, *Journal of Interdisciplinary History*, 29, [4], 683-723.

Goldring S., Longhurst J., Cruddas M., (2005), Model-Based Estimates of Income for Wards, 2001/02 – Technical Report, *Office for National Statistics*.

Goldstein H., (2010), Multilevel Statistical Models - 4th Edition, *Wiley Series in Probability and Statistics*, London.

Grafton R. Q., Kompas T., Owen P. D., (2007), Bridging the barriers: knowledge connections, productivity and capital accumulation, *Journal of Productivity Analysis*, 8, [3], 219-231.

Granovetter M., (1973), The strength of weak ties, *American Journal of Sociology*, 78, 1360-1380.

Green S. B., Akey T. M., Fleming K. K. Hershberger S. L., Marquis J. G., (1997), Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis, *Structural Equation Modeling*, 4, 108-120.

Grenier P., Wright K., (2003), Social capital in Britain: an update and critique of Hall's analysis, *International Working Paper Series*, 14, Centre for Civil Society, London School of Economics and Political Science, London.

Grootaert C., van Bastelaer T., (2001), Understanding and measuring social capital. A synthesis of findings and recommendations from the social capital initiative, *World Bank Social Capital Initiative Working Paper*, n. 24.

Hall P., (1999), Social Capital in Britain, *British Journal of Politics*, 29, 417-461.

Harper R., Kelly M., (2003), Measuring Social Capital in the United Kingdom, *Office for National Statistics Report*.

Harrington D., (2009), Confirmatory factor analysis, *Oxford University Press*.

Haughton D., Nguyen P., (2010), Multilevel Models and Inequality in Vietnam, *Journal of Data Science*, 8, 289-306.

Hauser C., Tappeneir G., Walde J., (2007), The learning region: the impact of social capital and weak ties on innovation, *Regional Studies*, 41, 75-88.

Hayes A. F., (2006), A Primer on Multilevel Modeling, *Human Communication Research*, 32 [4], 385-410.

Heady P., Clarke P., Brown G., Ellis K., Heasman D., Hennell S., Longhurst J., Mitchell B., (2003), Model-based small area estimation series No. 2, *Small Area Estimation Project Report: Office for National Statistics*.

Helliwell J. F., (2006), Well-Being, Social Capital and Public Policy: What's New? *The Economic Journal*, 116 [510], C34-C45.

Hendryx M. S., Ahern M. M., Lovrich N. P., Mc Curdy A. H., (2002), Access to Health Care and Community Social Capital, *Health Services Research*, 37 [1], 85-101.

Hindmarsh D. M., (2013), Small area estimation for health surveys, *University of Wollongong Research Online*.

Hofstede G. (1980), Culture's Consequences: International Differences in Work Related Values*, Sage Publications*.

Hox J. J., Moerbeek M., van de Schoot R., (2010), Multilevel Analysis: Techniques and Applications, *Quantitative Methodology Series – Second Edition.*

Hu L-T., Bentler P. M., (1999), Cutoof criteria for fit indexes in covariances structure analysis: Conventional criteria versus new alternatives, *Phsycological Methods*, 3, 424-453.

Hu L-T., Bentler P. M., (1995), Evaluation model fit in R. H. Hoyle (Eds), Structural equation modelling: Concepts, issues and applications, 76-99, *Thousand Oaks*, Sage.

Jacobs J., (1961), The life and death of great American cities, *Random House*.

Kaasa A., Vadi M. (2010), How does culture contribute to innovation? Evidence from European countries, *Economics of Innovation and New Technology*, 19 [7], 583-604.

Kaasa, A., Kaldaru, H., Parts, E. (2007), Social capital and institutional quality as factors of innovation: evidence from Europe, *Tartu University Press*, Order n. 292.

Kaufmann V., Bergman M. M., Joye D., (2004), Motility: Mobility as Capital, *International Journal of Urban and Regional Research*, 28.4, 745-756.

Kawachi I., Berkman L. F., (2003), Neighborhoods and Health, *Oxford University Press*, Chapter 4.

Kawachi I., Kennedy B. P., Lochner K., Prothrow-Stith D., (1997), Social Capital, income inequality, and mortality, *American Journal of Public Health*, 87 [9], 1491-1498.

Kish L., (1967), Survey Sampling – 2nd Edition, *Wiley and Sons*, London.

Knack S., Keefer P., (1997), Does social capital have an economic payoff? A crosscountry investigation, *The Quarterly Journal of Economics*, 21, 1251-1288.

La Due Lake R. and Huckfeldt R, (1998), Social Capital, Social Networks, and Political Participation, *Political Phsycology*, 19 (3), 567-584.

Laursen K., Masciarelli F., (2008), The effect of regional social capital and external knowledge acquisition on process and product innovation, *ROCK Working Paper*.

Leckie G., (2010), Module 5: Introducing to Multilevel Modelling – Stata Practical, *Centre for Multilevel Modelling*, adapted from the corresponding MLwiN practical: Steele, F. (2008) Module 5: Introduction to Multilevel Modelling. LEMMA VLE, Centre for Multilevel Modelling. Accessed at http://www.cmm.bris.ac.uk/lemma/course/view.php?id=13.

Li F., Zaslavsky A. M., (2010), Using a Short Screening Scale for Small-Area Estimation of Mental Illness Prevalence for Schools, *Journal of the American Statistical Association*, 105 [492], 1323-1332.

Li Y., (2015), Social capital in sociological research: conceptual rigour and empirical application in Li Y., (2015), Handbook of Research Methods and Applications in Social Capital, *Elgar Edition*, chapter 1.

Li Y., (2015b), The flow of soul: a sociological study of generosity in England and Wales (2001-2011), in Li Y., (2015), Handbook of Research Methods and Applications in Social Capital, *Elgar Edition*, chapter 3.

Li Y., Pickles A., Savage M., (2005), Social capital and social trust in Britain, *European Sociological Review*, 21, [2], 109-123.

Li Y., Savage M., Warde A., (2015), Social stratification, social capital and cultural practices in the UK, in Li Y., (2015), Handbook of Research Methods and Applications in Social Capital, *Elgar Edition*, chapter 2.

Lin N., (2001), Social capital: a theory of social structure and action, *Cambridge University Press*.

Lochner K. A., Kawachi I., Brennan R. T., Buka S. L., (2003), Social capital and neighborhood mortality rates in Chicago, *Social Science & Medicine*, 56 [8], 1797-1805.

Lomas J., (1998), Social capital and health implications for public health and epidemiology, *Social Science and Medicine*,47 [9], 1181-1188.

Lorenc T., Petticrew M., Whitehead M., Neary D., Clayton S., Wright K., Thomson H., Cummins S., Sowden A., Renton A., (2013), Fear of crime and the environment: systematic review of UK qualitative evidence, *BMC Public Health*, 13:496.

Loury G., (1977), A dynamic theory of racial income difference in *Women, Minorities and Employment Discrimination*, by Wallace P. A., La Mond A. M., *Lexington* Books, 153-186.

Luthans F., Avolio B. J., Avey J. B., Norman S. M., (2007), Positive Psychological Capital: Measurement and Relationship with Performance and Satisfaction, *Personnel Psychology*, 60, 541-572.

Macintyre S., Ellaway A., Cummins S., (2002), Place effects on health: how can we conceptualise, operationalise and measure them?, *Social Science and Medicine*, 55 (2002), 125-139.

Matsumoto D., Van De Vijver F. J. R., (2010), Cross-Cultural Research Methods in Psychology, *Cambridge Edition*.

Milbrath L.W., Goel M.L., (1977), Political participation: How and why do people get involved in politics, *Chicago: Rand McNally*.

Modena F., (2009), Under the social capital umbrella: definition and measurement, *Openloc Working Paper Series*, 11.

Mohan G., Mohan J., (2002), Placing Social Capital, *Progess in Human Geography*, 26 [2], 191-210.

Mohan J., Twigg L., Barnard S., Jones K., (2005), Social capital, geography and health: a small-area analysis for England, *Social Science & Medicine*, 60 [6], 1267-1283.

Mohnen S. M., Völker B., Flap H., Groenewegen P. P., (2011), Neighbourhood social capital and individual health, *Social Science & Medicine*, 72 (5), 660-667.

Morrin N. J., Blane D. B., White I. R., (1996), Levels of mortality, education, and social conditions in the 107 local education authority areas of England, *Journal of Epidemiology and Community Health*, 50 [1], 15-7.

Moura F. A. S., Holt D., (1999), Small area estimation using multilevel models, *Survey Methodology*.

Murray, C. (2005), Social capital and cooperation in Central and Eastern Europe– A theoretical perspective, *ICAR Discussion Paper*, n. 9.

Muthén L. & Muthén B., (1998), MPlus (Version 2.01) [Computer Software], Los Angeles: *Muthén & Muthén*.

Nahapiet J., Ghoshal S., (1998), Social capital, intellectual capital and the organizational advantage, *Academy of Management Review*, [23], 242-266.

Namazi-Rad M.-R., Steel D. G., (2015), What Level of Statistical Model Should We Use in Small Domain Estimation, *Australian and New Zealand Journal of Statistics*, 57 [2], 275-298.

Narayan D., Cassidy M. F., (2001), A Dimensional Approach to Measuring Social Capital: Development and Validation of a Social Capital Inventory, Current Sociology, 49(2), 59–102 SAGE Publications.

NatCen, (2004a), Synthetic estimation of healthy lifestyle indicators: User guide, Edited by Scholes S., Bajekal M., Pickering K., *National Centre for Social Research*.

NatCen, (2004b), Synthetic estimation of healthy lifestyle indicators: Stage 1 report, Edited by Bajekal M., Scholes S., Pickering K., Purdon S., *National Centre for Social Research*.

NatCen, (2004c), Synthetic estimation of healthy lifestyle indicators: Stage 2 report, Edited by Pickering K., Scholes S., Bajekal M., *National Centre for Social Research*.

NatCen, (2004d), Synthetic estimation of healthy lifestyle indicators: Stage 3 report, Edited by Pickering K., Scholes S., Bajekal M., *National Centre for Social Research*.

Newton K., (2001), Trust, Social Capital, Civil Society, and Democracy, *International Political Science Review*, 22 [2], 201-2014.

Nguyen P., Haughton D., Hudson I., Boland J., (2010), Multilevel models and small area estimation in the context of Vietnam living standards surveys, *42èmes Journées de* Statistique – Conference Papers.

Nie N. H., Hillygus D. S., Erbring L., (2002), Internet Use, Interpersonal Relations, and Sociability in B. Wellman and C. Haythorntwaite, The Internet in everyday life - A Time Diary Study, 215-244, *Blackwell Publishing*.

Norris P., Inglehart R., (2003), Gendering Social Capital: Bowling in Women's Leagues?, *Harvard University Paper*.

O'Neill B., Gidengil E., (2013), Gender and Social Capital, *Routledge*.

OECD (2013), The OECD measurement of social capital project and question databank, http://www.oecd.org/std/social-capital-project-and-question-databank.htm, accessed from January 2013.

OECD, (2001), The OECD measurement of social capital project and question databank, http://www.oecd.org/std/social-capital-project-and-question-databank.htm, accessed from January 2011.

Olson M., (1982), The Rise and Decline of Nations: Economic Growth, Stagflation, and Social Rigidities, *New Haven, CT: Yale University Press*.

ONS, (2011), Beyond 2011 – PhD Specifications, http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/index.html, accessed from January 2011.

ONS (2003), The Social Capital Project, http://www.ons.gov.uk/ons/guide-method/user-guidance/social-capital-guide/index.html, accessed from January 2012.

ONS, (2001), Social Capital A review of the literature, *Social Analysis and Reporting Division Office for National Statistics*.

Onyx J., Bullen P., (2000), Measuring social capital in five communities, *The Journal of Applied Behavioral Science*, n. 36, 23-42.

Paldam M., (2000), Social capital: one or many? Definition and measurement, *Journal of Economic Survey*, 14, (5), 629-653.

Paxton P., (2000), Social Capital and Democracy: An Interdependent Relationship, *American Sociological Review*, 67 (2), 254-277.

Paxton P., (1999), Is Social Capital Declining in the United States? A Multiple Indicator Assessment, *American Sociological Review*, 105 (1), 88-127.

Perna L. W., Titus M. A., (2005), The Relationship between Parental Involvement as Social Capital and College Enrolment: An Examination of Racial/Ethnic Group Differences, *The Journal of Higher Education*, 76 [5], 485-518.

Perry M., Williams R. L., Wallerstein N., Waitzkin H., (2008), Social Capital and Health Care Experiences Among Low-Income Individuals, *American Journal of Public Health*, 98 [2], 330-336.

Phongsavan P., Chey T., Bauman A., Brooks R., Silove D., (2006), Social capital, socio-economic status and psychological distress among Australian adults, *Social Science and Medicine*, 63 (10), 2546-2561.

Pohlmann J. T., (2004), *Use and interpretation of factor analysis*, *The Journal of Educational Research*, 98, 14-23.

Portes A., (1998), Social Capital: Its Origins and Applications in Modern Sociology, *Annual Review of Sociology*, 24, 1-24.

Portes A., (1995), The economic sociology of immigration: essays on networks, ethnicity and entrepreneurship, *Russel Sage Foundation*.

Putnam R. D., (2001), Bowling alone: The collapse and revival of American community, *S&S Paperbacks*.

Putnam R. D., (1995), Bowling alone: America's declining social capital, *Journal of Democracy*, 6 (1), 5-78.

Putnam R. D., Leonardi R., Nanetti R. Y., (1993), Making democracy work: civic traditions in modern Italy, *Princeton University Press*.

Rabe-Hesketh S., Skrondal A., (2012), Multilevel and Longitudinal Modeling Using Stata, Third Edition – Volume I: Continuous Responses, *Stata Press*.

Rao J. N. K., (2003), Small Area Estimation, *Wiley New York*.

Rasbash, J., Browne, W. J., Goldstein, H., Yang, M., Plewis, I., Healy, M. et al. (2002). A User's Guide to Mlwin. (2 ed.), *London: Centre for Multlevel Modelling: Institute of Education*.

Reder S., (1997), Synthetic estimates of literacy proficiency for Small Census Areas, *report prepared for Division of Adult Education and Literacy - Office of Vocational and Adult Education, U.S. Department of Education*.

Reesken T., Wright M., (2013), Cohesive Society: A Multilevel Analysis of the Interplay Among Diversity, National Identity, and Social Capital Across 27 European Societies, *Comparative Political Studies*, 46 [2], 153-181.

Rose R., (2000), How much does social capital add to individual health? A survey study of Russians, *Social Science & Medicine*, 51 (9), 1421-1435.

Rostila M., (2007), Social capital and health in European welfare regimes: a multilevel approach, *Journal of European Social Policy*, 17 [3], 223-239.

Rousseau D. M., Sitkin S. B., Burt R. S., Camerer C., (1998), Not so Different After All: A Cross-Discipline View of Trust, *Academy of Management*, 23 [3], 393-404.

Rubin A. M., Perse E. M., Powell R. A., (1985), Loneliness, Parasocial Interaction, and Local Televison News Viewing, *Human Communication Research*, 12 (2), 155-180.

Ruston D., Akinrodoye L., (2002), Social Capital Question Bank June 2002 – Questions from Social Capital surveys included in the Social Capital Survey Matrix 2002, *Social Analysis and Reporting Division – ONS Statistics*.

Sabatini F., (2008), Does social capital create trust? Empirical analysis of a community of entrepreneurs, *MPRA Paper*, 6781.

Sampson R. J., Morenoff J. D., Earls F., (1999), Beyond Social Capital: Spatial dynamics of collective efficacy for children, *American Sociological Review*, 64 [5], 633-660.

Schreiber J. B., Nora A., Stage F. K., Barlow E. A., King J., (2006), Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review, *Journal of Educational Research*, 99 (6), 323-338.

Scrivens K., Smith C., (2013), Four Interpretations of Social Capital: An Agenda for Measurement, *OECD Statistics Working Papers*, 2013/06.

Shane S., (1995), Uncertainty Avoidance and the Preference for Innovation Championing Roles, *Journal of International Business Studies*, 26 [1], 47-68.

Shane S. (1992), Why do some societies invent more than others?, *Journal of Business Venturing*, n. 7, 29-46.

Sloggett A., Joshi H., (1998), Deprivation indicators as predictors of life events 1981-1992 based on the UK ONS Longitudinal Study, *Journal of Epidemiology and Community Health*, 52, 228-233.

Snelgrove J. W., Pikhart H., Stafford M., (2009), A multilevel analysis of social capital and self-rated health: Evidence from the British Household Panel Survey, *Social Science and Medicine*, 68 [11], 1993-2001.

Snijders T. A. B., Berkhof J., (2007), Diagnostic Checks for Multilevel Models in *Handbook of Multilevel Model* edited by De Leeuw J. and Meijer E., *Springer Edition*.

Snijders T. A. B., Bosker R. J., (1999), Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, *London etc: Sage Publisher*.

Southerton D., (2003), 'Squeezing Time' Allocating Practices, Coordinating Networks and Scheduling Society, *Time Society*, 12 [1], 5-25.

Sturgis P., Patulny R., Allum N., Buscha F., (2012), Social Connectedness and Generalized Trust: A Longitudinal Perspective, *ISER Paper*, 2012-19.

Subramanian S. V., Kim D. J., Kawachi I., (2006), Bonding versus bridging social capital and their associations with self-rated health: a multilevel analysis of 40 US communities, *Journal of Epidemiology and Community Health*, 60, 116-122.

Subramanian S. V., Lochner K. A., Kawachi I., (2003), Neighborhood differences in social capital: a compositional artefact or a contextual construct?, *Health & Place*, 9, 33-44.

Svendsen T. G., Svendsen G. L. H., (2009), Handbook of social capital: the troika of sociology, political science, and economics, *Edward Elgar Edition*.

Swales K., (2015), Variations in community participation by neighbourhoods, *The State of Social Capital in Britain – Insights, opportunities and challenges*, Understanding Society, ESRC, National Council for Voluntary Organisations and Cooperative Councils Innovation Network, London, 11st November 2015.

Teachman J. D., Paasch K, Carver K, (1996), Social Capital and Dropping Out of School Early, *Journal of Marriage and Family*, 58 (3), 773-783.

The World Bank, (1998), Social Capital Initiative Working Paper Series, http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTSOCIALDEVELOPMENT/EXTTSOCIALCAPITAL/0,,contentMDK:20194767~menuPK:418848~pagePK:148956~piPK:216618~theSitePK:401015,00.html, Accessed from January 2012.

Tong S., Walther J. B., (2011), Relational maintenance and CMC. In K. B. Bright and L. M. Webb (Eds.), *Computer – mediated communication in personal relationships*, 98-118, New York: Peter Lang Publishing.

Tranmer M., Pickles A., Fieldhouse E., Elliot M., Dale A., Brown M., Martin D., Steel D., Gardiner C., (2005), The case for small area microdata, *Journal of Royal Statistical Society*, 168 [1], 29-49.

Tsai W., Ghoshal S., (1998), Social Capital and Value Creation: The Role of Intrafirm Networks, *Academy of Management Journal*, 41 (4), 464-476.

Twigg L., Moon G., (2002), Predicting small area health-related behaviour: a comparison of multilevel synthetic estimation and local survey data, *Social Science and Medicine*, 54, 931-937.

Twigg L., Moon G., Jones K., (2000), Predicting small-area health-related behaviour: a comparison of smoking and drinking indicators, *Social Science & Medicine*, 50, 7 [8], 1109-1120.

Twigg L., Moon G., Walker S., (2004), The smoking epidemic in England, *Health Development Agency Report*.

Tzavidis N., Salvati N., Pratesi M., Chamber R., (2008), M-quantile models with applications to poverty mapping, *Statistical Methods and Applications*, 17 [3], 393-411.

University of Essex. Institute for Social and Economic Research and NatCen Social Research, *Understanding Society: Waves 1-4, 2009-2013* [computer file]. *6th Edition.* Colchester, Essex: UK Data Archive [distributor], November 2014. SN: 6614, http://dx.doi.org/10.5255/UKDA-SN-6614-6.

University of Essex. Institute for Social and Economic Research, *British Household Panel Survey: Waves 1-18, 1991-2009* [computer file]. *7th Edition*. Colchester, Essex: UK Data Archive [distributor], July 2010. SN: 5151, http://dx.doi.org/10.5255/UKDA-SN-5151-1.

University of Essex. Institute for Social and Economic Research, NatCen Social Research. (2015). *Understanding Society: Waves 1-5, 2009-2014: Special Licence Access, Census 2011 Middle Layer Super Output Areas*. [data collection]. *4th Edition.* UK Data Service. SN: 7249.

University of Essex. Institute for Social and Economic Research, NatCen Social Research. (2015). *Understanding Society: Waves 1-5, 2009-2014.* [data collection]. *4th Edition.* UK Data Service. SN: 6614.

University of Essex. Institute for Social and Economic Research. (2010). *British Household Panel Survey: Waves 1-18, 1991-2009*. [data collection]. *7th Edition.* UK Data Service. SN: 5151.

University of Essex. Institute for Social and Economic Research. (2014). *British Household Panel Survey, Waves 1-18, 1991-2009: Special Licence Access, Census 2001 Middle Layer Super Output Area*. [data collection]. UK Data Service. SN: 7446.

Uslaner E. M., (2015), The roots of trust, in Li Y., (2015), Handbook of Research Methods and Applications in Social Capital, *Elgar Edition*, chapter 4.

Uslaner E. M., (2002), The moral foundations of trust, *Cambridge University Press*.

Van Deth J. W., (2000), Interesting but irrelevant: Social Capital and the saliency of politics in Western Europe, *European Journal of Political Research*, 37, 115-147.

Van Deth J.W. (1990), Interest in politics in Jennings M.K., van Deth J.W., Continuities in political action: A longitudinal study of political orientations in three western democracies, *Berlin/New York: De Gruyter and Aldine*.

Van Oorschot W, Arts W., Gelissen J., (2006), Social Capital in Europe: Measurement and Social and Regional Distribution of a Multifaceted Phenomenon, *Acta Sociologica*, 49 (2), 149-167.

Wang J., Staver J. R., Examining relationships between factors of science education and student career aspiration, *The Journal of Educational Research*, 94, 312-319.

Westermann O., Ashby J., Pretty J., (2005), Gender and social capital: The importance of gender differences for the maturity and the effectiveness of natural resource management groups, *World Development*, 33 [11], 1783-1799.

Whitworth A., edited by, (2013), Evaluations and improvements in small area estimation technologies, *National Centre for Research Methods – Methodological Review Paper*.

Whitworth A., (2012), Sustaining evidence-based policing in an era of cuts: Estimating fear of crime at small area level in England, *Crime Prevention and Community Safety*, 14, 48-68.

Wilding K., (2015), Concluding remarks, *The State of Social Capital in Britain – Insights, opportunities and challenges*, Understanding Society, ESRC, National Council for Voluntary Organisations and Cooperative Councils Innovation Network, London, 11st November 2015.

Wing S., Barnett E., Casper M., Tyroler H. A., Geographic and socioeconomic variation in the onset of decline of coronary heart disease mortality in white women, *American Journal of Public Health*, 82 [2], 204-209.

Wollerbæk D., Selle P., (2003), Participation and Social Capital Formation: Norway in a Comparative Perspective, *Scandinavian Political Studies*, 26 (1).

Woolcock M., (2001), The place of social capital in understanding social and economic outcomes, ISUMA *Canadian Journal of Policy Research*, 2, (1), 11-17.

Woolcock M., (1998), Social capital and economic development: Toward a theoretical synthesis and policy framework, *Theory and Society*, 27 (2), 151-208.

Wuthnow R., (2002), Religious Involvement and Status-Bridging Social Capital, *Journal for the Scientific Study of Religion*, 41 (4), 669-684.

Yli-Renko H., Autio E., Sapienza H. J., (2001), Social capital, knowledge acquisition, and knowledge exploitation in young technology-based firms, *Strategic Management Journal*, 22 (6-7), 587-613.

Zhang X., Holt J. B., Lu H., Wheaton A. G., Ford E. S., Greenlund K. J., Croft J. B., (2014) Multilevel Regression and Poststratification for Small-Area Estimation of Population Health Outcomes: A Case Study of Chronic Obstructive Pulmonary Disease Prevalence Using the Behavioural Risk Factor Surveillance System, *American Journal of Epidemiology*, 179 [8], 1025-1033.

Zhang X., Onufrak S., Holt J. B., Croft J. B., (2013), A Multilevel Approach to Estimating Small Area Childhood Obesity Prevalence at the Census Block-Group Level, *Preventing Chronic Disease*, 10 [E68].

Zheng W. (2010), A Social Capital Perspective of Innovation from Individuals to Nations: Where is Empirical Literature Directing Us?, *International Journal of Management Reviews*, 12 [2], 151-183.