Università degli Studi di Firenze
Dipartimento di Statistica "G. Parenti"
Scuola di Dottorato di Ricerca in Statistica Applicata
XXI ciclo – SECS-S/05

A Multilevel Approach to LC Models: Analysis of Social Exclusion among European Regions.

Elena Pirani

Tutor	Prof.ssa Silvana Schifini
Co-Tutor	Prof.ssa Carla Rampichini
Co-Tutor	Prof. Jeroen K. Vermunt
Director	Prof. Guido Ferrari

To all persons who believe in me.

Acknowledgements

This work would not have been possible without the contribution of several persons, that I wish to thank.

I sincerely thank Prof. Silvana Schifini who supervised me wisely and patiently, gave me support and advice, making this research possible. A special thanks to Prof. Carla Rampichini, for her help and advice, her helpfulness and the continuous motivation that she gave to me. I especially thank them for assisting me with perseverance and professionalism.

I wish to thank Prof. Jeroen K. Vermunt, who gave me the opportunity of spending some time at the Tilburg University, where I was able to take advantage of an intense and profitable period of work. I thank Prof. Vermunt for his precious contribution and his kindness, and all the Department of Methodologies and Statistics for their hearty welcome and hospitality.

I would also like to thank all the professors in my Department, particularly Prof. Guido Ferrari and Prof. Silvana Salvini, for their constant helpfulness and for the important things they taught me during these years of doctorate study.

A hearty thanks to Prof. Margherita Russo and Prof. Michele Lalla, of the University of Modena, who was the first to encourage me to take up my doctorate studies and who have always encouraged and motivated me. Moreover, I wish to thank Lara and Alessandro, who saw the beginning of this phase of my life and, even if away, have always encouraged and comforted me in case of need.

A though is for my doctorate workmates, with whom I have shared this experience. A special thank to Luca and Laura, with whom in this three years I have lived long days made of discussions and hard work, but also satisfaction and pleasant moments. I thank also Daniele, because conversing with him has represented a continuous spur for me.

This thesis is dedicated to my family, without whom I would have done nothing. I am very grateful to my parents, for having always supported me, endorsed my choices, and believed in me.

Finally, last but not least, a special dedication and a hearty thanks to my friends, who trust in me and, everyone in its own way, are close to me and facilitate my path, in particular: Tiziano, Alberto, Erika, Fabrizio, Michele, Giorgia, Letizia, Arianna, Christian, Veronica.

It is also because of all these persons that I am the person I have became.

Contents

Chapter 1

Introduction

In recent years, the term “social exclusion” has taken a prominent place in discussions concerning social policies and inequalities, in all European countries. While this notion has acquired important strategic connotations, by stressing structural and cultural/social processes, the precise meaning of the term remains somewhat elusive, irrespective of its frequent use in the academic and policy literature. Since the late 1980’s, the term “social exclusion” has became more and more important and used in political, institutional and, also, academic debate, till the Nineties, when there has been a shift in public and political discourse of several European countries from “poverty” to “social exclusion”. In European Commission documents, we remark nowadays a frequent reference to the decline in social cohesion and social solidarity, and the need to reintegrate the socially excluded into the mainstream society (European Commission 1993, 1997, 2000, 2007a). This terminology has emerged with reference to problems related to a “new poverty” that is not just monetary. “*Social exclusion refers to the multiple and changing factors resulting in people being excluded from the normal exchanges, practices and rights of modern society. Poverty is one of the most obvious factors, but social exclusion also refers to inadequate rights in housing, education, health and access to services.* (European Commission 1993, p. 1).

Social exclusion represents a multidimensional concept (Berghman 1995), which encompasses different forms of interdependent disadvantage and unease. These situations, whether prolonged in time, and in absence of intervention, may lead to exclusion of individuals from the mainstream society. Thus, individuals or social groups are mainly exposed to the risk of social exclusion when they experience, over long time, difficulties in relation to different dimensions of their life (Hills *et al.* 2002). The process of social exclusion refers to the persistence and the worsening of a disease condition *over time,* and it is connected to the individual’s past background and prospects for the future. Social exclusion is not only a negative condition per se, but also since it represents a disruptive element for social and economic development, both at individual and societal level. *[…] social exclusion emphasises the weaknesses in the social infrastructure and the risk of allowing a two-tier society to become established by default.*” (European Commission 1993, p. 1).

The fight against poverty and social exclusion is nowadays one of the central objectives of the European Union (EU) and of its member States, in a context where the links between economic and social spheres are assuming an increasing central importance (Atkinson *et al.* 2004). At the launch of the Lisbon strategy in

2000, the European Council invited member States and the Commission to take steps to make a decisive impact on the eradication of poverty and social exclusion by 2010. Within this strategy lies the adoption, at the December 2001 Laeken European Council, of a set of 18 statistical indicators of social exclusion – namely the Laeken indicators – with the objective to monitor social exclusion situations across European countries. This strategy arises out of the political agreement reached at the earlier European Councils in Amsterdam, Nice and Lisbon towards the promotion of social cohesion and inclusion as strategic goals. The recent proposal to designate 2010 as the European Year for combating poverty and social exclusion highlights the EU engagement to reaffirm and strengthen the initial political commitment.

The path towards 2010 has revitalized the current European debate about social exclusion, making it a key element of European socio-economic development. The socio-economic reality of the European Union is quite unbalanced today, with the growing presence of situations of social vulnerability. The quality of a society cannot ignore the extent of inequalities among its inhabitants.

Despite the growing interest around social exclusion issues, both at political and academic level, one still registers a lack of a comprehensive and common understanding about it. There are still many unresolved problems concerning the observation and the measurement of social exclusion. As opposed to analysis of poverty and economic disadvantages, few empirical studies investigating aspects of social exclusion can be found in the literature and, among them, there exists little agreement regarding its proper operationalisation. On the one hand, definitions, contents and interpretations of "social exclusion" are affected by the political, social and cultural model of the societies in which individuals live and relate themselves (Atkinson and Davoudi 2000, Mayes *et al.* 2001). On the other hand, one acknowledges the necessity of a shared theoretical framework, in order to develop a coherent approach of analysis.

The literature focusing on the definition of appropriate measures of social exclusion and on the identification of who is socially excluded, is increasing today (e.g. Burchardt *et al.* 1999, Tsakloglou and Papadopoulos 2002, Robila 2006). However, none of these measures seems to be completely satisfactory. Particularly, while the multidimensional nature of social exclusion is widely acknowledged, empirical studies have seemed to fail the multidimensional approach. The meaning of social exclusion is often mixed with related concepts, such as poverty, deprivation and marginalization, partially due to the lack of a conceptual clarity. As a result, there is a tendency to use poverty as a proxy for social exclusion, thereby undermining the multidimensional nature of exclusion. Moreover, most existing empirical applications regarding social exclusion actually deal mainly with economic aspects (e.g. Moisio 2004, Whelan and Maître 2005a), thus missing out the social dimension and outlining a limited picture of the problem. Moreover, while the idea that social exclusion lies in the perception of the individual of a weakness of his economic situation, of his rights and relations, has recently emerged (European Commission 1998, Eroglu 2007), there is a lack of applied researches which focus on the subjective elements with appropriate and rigorous statistical methods. For example, even Laeken indicators are objective measures which refer mainly to the economic domain. Finally, current analyses are limited to restricted context or segments of population, disregarding a comparative approach. The socio-economic context in which individuals are embedded is not accounted for.

This thesis moves from the acknowledgement of the existing lacks and from the need to develop appropriate methods to investigate social exclusion. We propose an operational approach to the study of social exclusion, in terms of *conceptual model*, *indicators* and *applied statistical method*, in such a way that an initial empirical analysis of social exclusion across European regions can be undertaken.

The absence of a shared definition of what is meant by the term "social exclusion" represents one of the causes of the deficiency in empirical research. Following recent literature (e.g. Bhalla and Lapeyre 2004, Berghman 1995, Hills *et al.* 2002), we first propose a conceptual model of social exclusion based on the identification of three principal dimensions we consider relevant on this issue: an economic dimension, a social dimension, and an institutional dimension. We believe that the insufficient participation in (one of) these three dimensions of human life might trigger social exclusion situations. The *economic* dimension refers principally to monetary and financial, in its different elements – namely income, wealth, saving capability, and so on. It is clear that this dimension relies directly to the concept of poverty. The *social* dimension concerns social relationships with family, friends, neighbours, local community, etc. Social relationships represent the principal focus of our approach to social exclusion, and the element that differentiates it from other similar concepts. In this dimension we place also elements like social recognition, social participation, identification in cultural and moral values. Finally, we identify an *institutional* dimension, which concerns relationships between people and the State, and includes the so-called active citizenship rights.

A list of the relevant elements referring to each of these dimensions is difficult to fulfil and it cannot be exhaustive. We will not address all these conceptual issues; rather, we will define an empirical framework that allows implementing a working approach that accords with our (admittedly imperfect) understanding of the concept. Other possible conceptions of social exclusion are conceivable, but our intention is to elaborate an approach and make it coherent and operational, attempting to overcome the weak points still present in the current social exclusion research. We believe that the proposed multidimensional conceptualization enables to highlight the fact that socially excluded people do not represent a homogeneous group but, rather, that social exclusion derives from the sum and the interaction of different kinds of risk factors and disadvantaged situations, and that this process is influenced by the social, cultural and economic context in which individuals live. In our approach, we consider relational and distributive aspects together. The role of social relations and cultural context may represent an irreplaceable support to compensate possible economic difficulties, especially when they are transitional. In less individualistic societies, for example, the family represents a mechanism to combat social exclusion in situations of exclusion from the labour market, supporting a process of social integration and damping down on the tie cause-and-effect between economic deprivation and social relations deprivation (Bhalla and Lapeyre 2004). In other cases, the protection and welfare system represents an important factor in attenuating the impact of unemployment and poverty.

Secondly, using the 56.1-2001 Eurobarometer survey, indicators about involvement in these dimensions are developed and analysed for all EU-15 countries. Reckoning with limited data availability, we select some indicators for each identified domain of exclusion, in order to take into account each different dimension and analyse the relations between them. Instead of build a composite indicator of cumulative disadvantage as typical in this context, we propose to use a set of indicators in a multivariate model. Indeed, in some cases, starting from basic indicators (namely Eurobarometer questions) synthetic indicators are built, in order to introduce in the analysis a large number of information, without increasing too much the numbers of indicators. All the selected elements, referring to different areas of human life, interact and influence themselves reciprocally, so exclusion in one dimension could determine, or makes worse, exclusion in the others.

The 56.1-2001 round of Eurobarometer Survey is especially useful for the study of social exclusion, due to its principal focus on monitoring poverty and social exclusion situations. Through a set of dedicated questions, it investigates the extent of these problems for respondents, and their involvement from different points of view: the main elements that people deem to be the main causes of social exclusion; economic and financial situation of respondents and characteristics of the place where they live; subjective feeling about social exclusion, sense of usefulness in the society, quality of life, etc. Moreover, this data source contains information for all EU member countries, allowing a comparative analysis. The disposable sample refers to the 15 countries of European Union before the recent enlargement, and the data structure allows performing the analysis below the national level, using the so-called NUTS regions at the first level of Eurostat classification (NUTS-1).

The selected dataset fits in with our conceptual framework almost naturally, allowing using both subjective and objective indicators for the most part of aspects under investigation.

The third step of our operational approach involves the choice of the most suitable statistical model. We propose to study social exclusion in a latent variable modelling perspective. Like many other concepts in social sciences, social exclusion represents a theoretical construct that can only be quantified or measured via indirect manifest indicators, which are assumed to be related in some way with it. It is a fuzzy and not well-determined concept, linked to several manifest indicators by *probability relations and not rigid laws* (Lazarsfeld and Henry 1968). The goal of Latent Variable models is to identify the pattern of the relationships among a plurality of relevant and observable variables by determining the characteristics and dynamics of the underlying latent concept. This modelling framework allows to take into account the multidimensionality of the data.

Particularly, we recur to Latent Class Analysis (LCA), which represents a powerful tool for investigation and analysis in situations in which one disposes of highly interrelated observed measures (categorical and/or metrical), and where association is due to some underlying unobserved factors assumed to be (or treated as) categorical. The Latent Class models do not require strong assumptions – namely the Normal one – concerning the data distribution. Since "*the social world seems to have been created with less multivariate normality than many researchers are willing to assume*" (McCutcheon 1987, p. 79) we believe that Latent Class Analysis could represent a useful tool on this issue. Through LCA, we deal with social exclusion as latent construct. In our empirical application, the latent classes represent the latent levels of social exclusion, which structure the cases with respect to a set of observed indicators.

The classical LC model is extended by introducing a hierarchical component, focusing not only on the differences between groups (European regions), but also on the latent distribution of each group. The first level of analysis is thus represented by individuals, for whom a set of responses variables (indicators) is provided; the regions in which individuals live represent the second level. The latent class approach is well known in the one-level framework (e.g. Hagenaars and McCutcheon 2002); in the multilevel framework it was first proposed by Vermunt (2003). Through a multilevel model, it is possible to highlight the differences between European regions and to verify whether and to what extent, the same risks and disadvantages determine the same perception of marginalization and exclusion in different political, economic, social and cultural contexts.

In our analysis, we consider also some covariates concerning both individuals and regions. While indicators serve to define and measure the latent classes, the covariates operate as explicative variables, useful to improve the description of the latent classes in terms of individual characteristics (e.g. age, economic situation, occupational status, living conditions, level of education and so on). Also elements operating at regional level are considered, in order to describe and accounting for the cultural context in which individuals are embedded and which may affect their responses. Covariates are gathered both from Eurobarometer survey and from Eurostat.

In order to take into account group differences in multilevel analysis, the random-effects approach is often used (Skrondal and Rabe-Hesketh 2004; Snijders and Bosker 1999). For the specification of the mixing distribution, we follow a nonparametric random-effects approach, introducing a *discrete latent variable* also at group-level. Concerning the analysis of social exclusion in a comparative approach, it is useful both from a substantial and technical point of view. A hierarchical LC models is implemented to take into account the existing multilevel structure of European population and to model regional differences in the distribution of the latent variable, allowing some parameters to differ across regions.

The proposed multilevel Latent Class model enables to attain simultaneously the identification of different profiles both of respondents and of regions, allowing for social exclusion to manifest itself in different ways for different subgroups across European regions.

The proposed approach of analysis represents a novelty in social exclusion literature. Our purpose is to analyse social exclusion in a multidimensional perspective, underlying the role of social relations on this issue and accounting also for subjective aspects. Moreover, the multidimensional framework, used to deal with the hierarchical structure of data, enables to highlight the presence of different structures of social exclusion and thus the relativity of this concept.

1.1 Thesis structure

In Chapter 2 we briefly review the origins of the concept of social exclusion used in institutional and academic European context (§ 2.1.1), defining its attributes and characteristics (§ 2.1.2) and describing the elements usually considered relevant to determine social exclusion situations (§ 2.1.3). In Section 2.2, we see how social exclusion is afforded by European Union, in terms of policies and indicators. The chapter concludes with a critical review of the existing data source about social exclusion (Section 2.3).

Chapter 3 describes the operational conceptualisation we propose. The characteristics and the elements referring to economic dimension of social exclusion are discussed in Section 3.2, with a particular stress on the differences between this concept and the poverty one; Section 3.3 deals with the social dimension in its different facets, like family models, social networks and unemployment. In Section 3.4, citizenship rights, social participation and welfare systems concerning the institutional dimension are discussed.

Chapter 4 deals with Latent Class Analysis. After describing the Latent Variable modelling framework (Section 4.1), the classical Latent Class Analysis is introduced. Generalizing for both continuous and categorical observed indicators, in Section 4.2 we describe the main features of the standard model, and in Section 4.3 its principal extensions. Then, in section 4.4 the multilevel Latent Class modelling used in the thesis is afforded. Since the aim of this research is to study data with one level of aggregation (individuals nested in regions), only two-level model is illustrated. Paragraphs 4.4.1 and 4.4.2 show different approaches to deal with hierarchical data structure in latent class framework, following, respectively, fixed and random effects approach. Finally, Section 4.5 presents technical aspects relating to the estimation, the fitting and the evaluation of these models. An effort to make different notations found in literature consistent, and to uniform them through the different models presented in this work, has been made.

Finally, in Chapter 5 we present the result of the multilevel Latent Class model performed in order to analyse social exclusion across European regions. After defining some issues concerning the study of social exclusion through Latent Class models (Section 5.1), Section 5.2 describes the dataset used (§ 5.2.1), the indicators (§ 5.2.2), the covariates (§ 5.2.3) and the territorial units (§ 5.2.4) introduced in the estimated model. Section 5.3 presents a brief description of the

EU context referring to these variables. Finally, after the specification of the model we implemented (Section 5.4), in Section 5.5 the main results of the multi-level Latent Class model are presented and discussed.

At the end of the dissertation, in Chapter 6 some concluding remarks summarize the strength points and the main results of the work. Furthermore, some limitations of the analysis are highlighted, together with some proposal for future research.

Chapter 2

Social exclusion: theoretical aspects

In the Eighties, the rising presence of negative indicators in different domains of life called attention of the EU institutions to the difficult and the inadequacy of Europe to provide a social quality to its citizens. At the same time, it indicated the necessity to introduce new concepts to evaluate the trend of European societies, and to direct a growing attention towards the social quality of the European citizens. In those years began a conceptual shift from poverty to social exclusion. In Nineties, issues relied to social exclusion and inclusion, as well the promotion of social cohesion as a strategic goal, attracted much attention in European debate as, i.e., demonstrated by the Treaty of Amsterdam (1997), the Nice Treaty (2001) and the Lisbon Treaty (2007). The designation, effected by the European Commission, of 2010 as the "European Year for Combating Poverty and Social Exclusion" reaffirms the global relevance of this topic and the EU's commitment (European Union 2007a, 2007b).

The terminology linked to social exclusion has emerged with reference to the problems related to a "new poverty" that is not just monetary, but includes deep changes both at individual and societal level. This change in terminology was a sign of the transformations in the economic systems of western European countries and, at the same time, it represented a forerunner of a change in European policy objectives. The term "social exclusion" has gradually replaced the term "poverty" in order to assess the appropriate intervention and to promote new coordination and cooperation policies at European level for the promotion of social inclusion and cohesion. Weakening of family ties, increasing of the job precariousness and unemployment rate, decline in social participation, and growing feeling of insecurity, are concrete current problems that cannot be adequately described by standard measures of poverty. As recent studies show (Earsterlin 2001), among the most industrialized European countries the link between wealth and social well-being is breaking off. Hence, one needs new tools to understand factors affecting social well-being of individuals and to evaluate and monitor the impact of social policies in different States.

In spite of the growing interest, social exclusion still represents a rather recent concept, for which there is not a full agreement on its definition. The meaning of social exclusion is often mixed with other related concepts such as poverty, deprivation and marginalization.

The aim of this chapter is to review and discuss the different conceptions of social exclusion used in European discussion, both at academic and institutional level. Clearly, they reflect different approaches towards its analysis and understanding, and each of them underlines a specific aspect. Particularly, we briefly review the different definitions of social exclusion (section 2.1.1), we describe the main characteristics of this concept (section 2.1.2), and the different ways in which social exclusion may affect life of individuals (section 2.1.3). Finally, we describe how the problem of social exclusion has been investigated till now by European institutions (section 2.2), and conclude highlighting some critical points still existing about data availability (section 2.3).

2.1 On defining exclusion

2.1.1 Origins of a new concept

The term “social exclusion” originated in France during the social policy of the Seventies (Percy-Smith 2000). It referred primarily to people living on the margins of society, and particularly to those who were excluded from the provision of the social insurance system. Thus, the socially excluded were those people administratively excluded by the State. In French original thinking, the core of the debate is the notions of “solidarity” and “integration”, in which the State plays the major role (Bhalla and Lapeyre 1997, Atkinson and Davoudi 2000). In this conceptualization, social exclusion is view as a failure of the Republican State in protecting the “cohesion of the society”, a process of social disaffiliation leading to a breakdown of the relationships between the society and the individual.

In Anglo-Saxon approach, the concept of social exclusion begins to be relevant in the Nineties, following the end of an economic growth period. Societal transformations, above all in the labour market structure, and the consequently reappearance of poverty and unemployment, revived attention about new forms of poverty and marginalization. In the UK, however, the term social exclusion did not appear in the official discourse prior to the Labour government in 1997[1]. Nowadays, the concept of social exclusion is part of the common British social policy debate. In Anglo-Saxon understanding, the concept of social exclusion is mainly relied to the lack of economic and financial resources. Unlike the French one, Anglo-Saxon tradition is rooted in the Liberal paradigm, where social integration is view in terms of relationships between individuals rather than a relationship between the individual and the society (Silver 1994). The society is view as a mass of atomized individuals in competition within the market place, therefore exclusion may result either from individual choices, or from system distortions, such as discrimination, market failures and unheard rights. In this context, social exclusion and poverty are frequently used in an interchangeable way.

From those years, the terminology linked to social exclusion has come to assume a more and more central place in discussions of social policies and inequali-

[1] In 1997 the term “social exclusion” became integrant part of the government’s policy with the setting up of the interdepartmental Social Exclusion Unit. For further details, see http://archive.cabinetoffice.gov.uk/seu/.

ties in European countries. However, although the original terminological shift from poverty to social exclusion has nowadays become a conceptually shift, the precise meaning of the term social exclusion remains somewhat elusive[2]. What is meant nowadays by "social exclusion"? Difficulties arise when one attempts to specify a precise or largely shared definition of social exclusion. Simply, there is no one univocal definition, but various interpretations of this phenomenon, each of them highlighting complementary aspects. In next paragraphs, we describe the main shared attributes of social exclusion concept, and we review principal definitions used in academic literature and institutional practice.

2.1.2 Main characteristics of social exclusion

Although a univocal definition of social exclusion seems far to be achieved, agreement has recently emerged in academic and institutional discussions regarding a number of attributes of social exclusion. These attributes substantially aim to conceptually differentiate between poverty and social exclusion. Indeed, even if most recent conceptualizations of poverty take into account a variety of elements (see e.g. Whelan and Whelan 1995a), the more comprehensive notion of social exclusion goes beyond and implies a wider analysis' perspective.

Above all, social scientists agree in considering social exclusion a multidimensional concept (e.g. Room 1995, Jordan 1996, Peace 2001). *Multidimensionality* implies that deprivation and lack of resources determining social exclusion have to refer to a broad set of quantitative and qualitative elements. People may be excluded from minimum consumption – due to a lack of economic resources – from employment, housing, education, welfare state, citizenships, or personal relationships (Silver 1994). In this perspective, the evaluation of the individuals' standard of life cannot be based merely on economic indicators – namely income measures – as it happens for the poverty concept. Multidimensionality implies that also opportunities and services available in the community of reference have to be accounted for when considering the level of resources available to individuals. Multidimensionality nature of social exclusion, moreover, involves the necessity to extend the analysis to the social relationships' field, namely social isolation, social support networks and social participation. Both relational and distributional factors are relevant in social exclusion issue (Bhalla and Lapeyre 1997, 2004). Weak social interactions (with family, friends and local community) and inadequate social participation represent a serious threat to social integration, both at individual and at collective level. People are often deprived of different things at the same time, thus focusing on multidimensionality of deprivation allows to account for this cumulative process.

Notion of social exclusion brings about another major theoretical contribution: it implies a focus on the relations and processes that cause multidimensional deprivation. The idea of social exclusion as *dynamic* concept or *process* is opposed to poverty seen as a static outcome. Dynamism refers to the persistence and

[2] According to Anthony Atkinson, the lack of a precise meaning might have contributed to determine the success and a widespread use of this concept (Atkinson 1998).

the worsening of a disease condition over time, which makes social exclusion a process rather than a status. From this perspective, social exclusion does not simply arise from an individual or group's current status, but it is connected to their past background and prospects for the future; social exclusion condition depends on how a situation and circumstances develop or are expected to develop. In this sense, also an across generation perspective (Atkinson 1998) could be relevant. Berghman offers an interesting synthesis of these two first characteristics of social exclusion. According to this author, the novelty of social exclusion, with respect other concepts, is given by its *comprehensiveness* and its *dynamic* character (Table 2.1).

The social exclusion could represent a most adequately tool, with respect the poverty one, to understand the dynamic nature of the processes of spatial accumulation of social disadvantages.

Table 2.1 – Conceptual differences between poverty and social exclusion

	Static outcome	***Dynamic process***
Income	Poverty	Impoverishment
Multidimensional	Deprivation	Social exclusion

Source: Berghman (1995).

The third characteristic of social exclusion is its *relativity*. Relativity means that an individual is socially excluded only with respect other members of his society, and it does not exist an "absolute" social exclusion condition. This connotation involves "exclusion" of people from a given society at a given time: to judge if a person is excluded or not, we have to observe the person relative to the context and the society he lives in. In this sense, in order to reach a meaningful understanding of factors determining social exclusion, one needs to adopt an appropriate spatial-temporal perspective. The nature of social exclusion and its causes are likely to vary a great deal from society to society[3]; across countries and, even, within national boundaries.

2.1.3 Excluded from what?

The multidimensionality nature of the social exclusion condition implies the necessity to identify the domains of life for which social exclusion may be relevant. We found many conceptualizations about social exclusion and, despite some points of contact, several conceptual and operational differences still remain.

[3] The context of the analysis, e.g. the socio-economic level of the area under investigation, is important also to determine the relative relevance of distributional and relational aspects (Bhalla e Lapeyre 2004).

"*Social exclusion refers to the multiple and changing factors resulting in people being excluded from the normal exchanges, practices and rights of modern society. Poverty is one of the most obvious factors, but social exclusion also refers to inadequate rights in housing, education, health and access to services.*" (European Commission 1993, p. 1). This definition clearly endorses the multidimensional factors and processes of social exclusion; the basic standard of living to which citizens have right is referred to a broad set of good and services, also public ones. Besides this general definition, European Union directly points out the attention on some critical risk factors in social exclusion situations. Thus, ethnic and religious minorities or migrants, and disabled person represent categories of population most exposed to the risk of economic and social weakness.

Jos Berghman and the researchers of the Observatory of National Policies to Combat Social Exclusion have been between the firsts to expand the concept of social exclusion beyond the traditional concept of poverty, stressing the importance of citizenship rights. They defined a comprehensive framework that refers to a "*breakdown or malfunctioning of the major societal systems that should guarantee full citizenship*". They proposed to conceive social exclusion in terms of "*the denial – or non-realisation – of citizenship rights*", that is civil, political and social rights (Berghman 1995, p. 19). This framework identifies four *systems*, each of them would promotes integration and inclusion in one of the principal dimensions of human life: democracy and legal system, which have to promote civic integration, labour market system, welfare system, and family and community system. In this definition, attention is directed not just to people's rights, but to society's institutions in which that rights can be fulfilled. Through this conceptualization, it becomes clear that poverty is merely a specific form of the most comprehensive social exclusion. Finally, another point stressed by Berghman is interesting. He states that the study of social exclusion has to be placed in a supranational analytical framework, in the specific case represented by the European one.

Other authors, emphasizing multidimensionality, deem that social exclusion derive from the sum and the interaction of various types of exclusion, each of them increasing or decreasing the vulnerability of individuals. For example, Kronauer (1998) identifies the exclusion from labour market, the economic exclusion, the institutional exclusion, the social isolation and the cultural exclusion. Moreover, accounting for the fact that people with limited economic and financial resources may concentrate in a geographic area, he proposes to consider the "territorial exclusion", which characterizes people living in scarce infrastructured areas[4].

Similar conclusions were drawn by Burchardt *et al.* (1999). According to them, social exclusion can be defined as "*a process which causes individuals or group, who are geographically resident in a society, not to participate in the normal activities of the citizens in that society*". (Burchardt *et al.* 1999, p. 230). Unlike Berghman, they conceptualize social exclusion in term of *activities* rather

[4] In a broader sense infrastructures may encompasses transport, medical services, shopping, events and cultural spaces, and so on. Substantially, it represents a transversal dimension, which cuts across and interacts with all the other ones.

than in term of *systems*, and in order to make operational their definition they identify five categories of "*normal activities*" for which it is important for citizens to participate: consumption activities, savings, production, political and social activities. The first three concern mainly economic and material aspects in their different connotations, while the last two introduce relational aspects, at different levels. These activities are connected each others, and the participation in one may influence the participation in another one. Talking about "*normal activities*" the authors want to stress the element of relativity since what is considered "normal" may evidently differ in time and place[5].

Also the conceptualisation of Bhalla and Lapeyre (1997, 2004) moves from the idea that social exclusion concerns both relational and distributional issues. They stress the importance of the income in determining social exclusion, but they acknowledge the possibility of social exclusion instances also in situation of material wellness. According to the multidimensionality, the authors identify three dimensions, which correspond to different fields in which human life develop itself: an economic dimension, a social dimension and a political dimension. They discuss in detail these dimensions, given clear indications about their meaning, contents and measurement. The economic dimension, relied to the concepts of poverty and deprivation, includes issues of income, production and access to goods and services market. The social dimension is though as the interaction of individuals (or of categories of individuals), and in this sense it refers to the access to the labour market[6], to the access to public goods and services market, and to the relational aspects among individuals and between individuals and State. Finally, through the definition of the political dimension – which includes aspects connected with personal security, freedom of expression, political participation and equal opportunities – also Bhalla and Lapeyre take advantage of the notion of citizenship rights.

The proposal raised in the end of Nineties from European Commission on "Non-monetary indicators of social exclusion" (European Commission 1998), also pointed attention on different dimensions representing the main spheres of human life in which potential causal factors may trigger social exclusion situations: a social, an economic, an institutional and a territorial dimension. The relevance of this definition stays mostly in a fifth added dimension, the "symbolic references", which refers to elements like identity, social visibility, self-esteem, interests, motivation, and future perspectives: all elements that combine to describe a society, their intangible elements like culture, motivation, and attitudes. Substantially, the Commission suggests including subjective and immaterial elements that depict a frame inside which to study a society. This one represented the first and unique attempt to introduce in the social exclusion framework some subjective aspects.

[5] Particularly, the authors test their conceptual framework through an empirical analysis of social exclusion in Great Britain between 1991 and 1995.

[6] In their conceptualization, the authors consider a broad notion of labour market and of its related problems, further than the dichotomisation employment-unemployment.

Finally, to complete this briefly review on social exclusion definitions, it is worthwhile mentioning some definitions developed by political national institutions that in last decade has occupied a central role in developing strategies to combat social exclusion at national level. In the front line to discuss and to implement actions and policies about these problems, we find the United Kingdom and Ireland. The UK Social Exclusion Unit[7] gives a definition focusing on social exclusion as outcome: "*Social exclusion is a shorthand label for what can happen when individuals or areas suffer from a combination of linked problems*" (Social Exclusion Unit 1999), identifying some risk factors such as unemployment, poor skills, low incomes, high crime environments, bad health and so on. The necessity of an integrating approach in order to combat social exclusion is afforded by The Scottish Office (1999) which adds: "*Action to promote social inclusion needs to be both comprehensive and co-ordinated: it must address the full range of issues facing an individual, a family or a community*". The Northern Ireland version explicitly considers the notion of citizenship rights: "*Social exclusion is a set of processes, including within the labour market and the welfare system, by which individuals, households, communities or even whole social groups are pushed towards or kept to the margins of society. It encompasses not only material deprivation but also more broadly the denial of opportunities to participate fully in social and civil life*" (Democratic Dialogue 1995).

From previous discussion, it emerges that social exclusion is a contested term, or at least an indefinite one. Each definition found in literature stresses, from time to time, various elements and a different theoretical approach. A lot of work has been made since the introduction of this concept. However, we think that the debate is still open and some crucial topics still remain. A first limit is represented by the fact that most definitions given in literature are scarcely operating. The presented frameworks refer mainly to conceptual approaches, without considering operational problems. Therefore, it could be the case that actual available information does not provide appropriate indicators of such dimensions, thus undermining the exploitability of these conceptual frameworks and nullifying their useful. We shall return on this topic in section 2.3. Finally, major deepening and clarity is needed concerning the "subjects" to which the analysis should be referred to. For example, in previous definitions, different kinds of "subjects" that could be affected by social exclusion – namely individuals, households, social groups or areas – are treated together indifferently. It is evident that the way to tackle the problem at these various levels will be strongly different, as different are the conceptualisations, the causes, the risk factors and, finally, the indicators and measures to use, included the possibility to exploit subjective ones.

[7] See foot note 1.

2.2 Social exclusion in EU debate

During the 1980s and the 1990s, consequently to the reappearance of poverty and unemployment instances, many of the assumptions underlying the social development of the EU began to be questioned. In this context first came to prominence the idea of a European *social policy*, in order to accompany common market policies toward a greater integration and social cohesion. Within the European Union, the term “social exclusion” first appeared during the Presidency of Jacques Delors (1985-1995). Initially, due to the influence of Delors’ own inclinations (Atkinson and Davoudi 1998), the meaning of social exclusion in EU discourses largely derived from French social policy and its associated welfare regime.

While *economic* cohesion has been a key goal since the early treaties establishing the European Economic Community, *social* cohesion has really become a focal point with the negotiations around the Maastricht Treaty. In the first and second European Programme against poverty (1975-80 and 1986-89), “disadvantage” and “deprivation” were the focus of European Council, but with the third Programme (1990-94) the terms *social exclusion* and *social cohesion* become the core of discourses and decisions. The change in terminology stresses the need to achieve, in a system of organisation based on market forces, the realization of an internal solidarity and mutual support, which ensures open access to benefit and protection for all members of society (European Commission 1997). The shift of the emphasis from poverty to social exclusion resulted from an increasing concern with the structural and multidimensional nature of processes and problems experienced by individuals, groups and areas. Social exclusion and social cohesion represented a term introduced to offer an alternative for innovative social policy that avoided the stigma of concepts such as poverty and deprivation. Exclusion was a more flexible concept[8]. Nowadays, the concept of social exclusion has become a core concept in the European Union and it is inserted in European programme as focus of European social policy.

2.2.1 Laeken indicators of social inclusion

Since Seventies, European Commission is engaged in actions and programs in order to investigate and to eradicate situations of poverty, but it is only in recent years that this phenomenon has become the object of more coordinated policies. The first step in this direction has been the Lisbon European Council in 2000. This summit culminated with the so-called “Lisbon Strategy”, a development plan for the European Union’s countries, that aimed to make the EU “*the most dynamic and competitive knowledge-based economy in the world capable of sustainable*

[8] Some authors critically argued that in European debate the term “social exclusion” was preferred to the “poverty” one, for political reasons, because of the opposition of some member States’ governments to apply the term poverty to their own countries. On the other hand, one could state that the concept of poverty was deemed inadequate considering that the welfare states in Europe guaranteed a minimum income and access to basic services (Cf. Percy-Smith 2000, Bhalla and Lapeyre 2004).

economic growth with more and better jobs and greater social cohesion, and respect for the environment by 2010" (European Union 2007b). The subsequent Laeken European Council in 2001 continued in this direction, in order to establish a higher coordination among national social policies, on the basis of shared programmes and indicators. Following the mandate from the European Council, the Social Protection Committee and its technical group was concerned with improving a set of common statistical indicators in the field of poverty and social exclusion. Such indicators, known as "Laeken indicators" (European Commission 2003, European Union 2001) were developed as a part of the Lisbon Strategy, with the purpose to observe, monitor and understand these phenomenon in a comparable, uniform and homogeneous way in all the European countries.

The 18 indicators has been conceived in order to give a balanced representation of Europe's social concerns. Given the large number of indicators needed to properly assess the multidimensional nature of social exclusion, the Social Protection Committee differentiated two levels of priority for the selected indicators[9]. The *primary indicators*[10] consist of a restricted number of lead indicators which cover the elements considered most important in determining social exclusion situations: income, unemployment, education, health, and so on. The *secondary indicators* refer to the same elements, but they would support the primary ones adding other dimensions and facets of the problem.

While it is fundamental this renewed attention of the European institutions about the economic and social living conditions of its citizens, and the enlargement of the focus on the social inclusion and social cohesion, it is straightforward to note that, once more, there is not a clear-cut distinction between poverty and social exclusion. In European Union discourses, we notice that the two different terms are used even if not as synonymous, at least as interchangeable. Moreover, also from an operational point of view we remark some weak points. From the list of Laeken indicators it emerges a statistical frame that considers the dimension connected to "poverty and inequality" as a key-factor in the process of social exclusion. The weight given to economic and poverty's indicators is relevant: among the primary ones, four out of ten refer to poverty's aspects, and between the secondary indicators five out of eight. These indicators represent a leap in quality compared to the past, especially in relation to their multidimensional perspective. Anyway, it seems that the actual relevance of the Laeken indicators as measures of social exclusion instances is not yet completely fulfilled.

---

[9] There may also be a third level, for indicators that are not strictly prevented at community level. In order to highlight specificities in their national areas and to help interpret the primary and secondary indicators, each member States may decide itself to adopt other third level indicators.

[10] The ten primary indicators are: low income rat after transfers, distribution of income, persistence of low income, median low income gap, regional cohesion, long term unemployment rate, people living in jobless household, early school leavers, life expectancy at birth, self perceived health status. The eight secondary indicators are: dispersion around the 60% median low income threshold, low income rate anchored at a point in time, low income rate before transfers, distribution of income (Gini coefficient), persistence of low income, long term unemployment share, very long term unemployment rate, persons with low educational attainment.

2.3 Some remarks about data

Studying social exclusion as individual attribute, one refers to numerous aspects of individuals' daily life. Information about all these aspects may be collected only by means of sampling surveys. On this concern, however, we remark some problems. On the one hand, there are some methodological problems concerning the collection and the measurement of the different elements under investigation; on the other hand, there is a scarcity of data, namely on relational aspects. Measurement of various aspects is lacking in, due to the absence of appropriate items and indicators that refer to a given theory. Since it is clear that one cannot develop a conceptual model on the basis of existing data, it would be to be hoped that *ad hoc* surveys were carried out.

Problems related to cross-national sampling surveys are well known in literature. The variation in sampling methods across countries and across years, and problems of vocabulary and translation, could compromise cross-national and cross-temporal comparison. A key issue arising in cross-national surveys is the questions wording. Despite the efforts to translate in the appropriate national languages, linguistic nuances could remain. In different countries, cultures, and systems of values, it could happen that the same word, even if correctly translated in the national language implies a slightly different concept, or at least different interpretation of the same concept. Social exclusion represents a typical example.

Concerning the study of social exclusion, besides these methodological problems one has to take into account also another critical issue, due to the complexity of this phenomenon. The data requirement for the operationalisation of the approaches previously described are discouraging, and the information required for their full implementation does not exist in any data set currently available. Different surveys have been developed to investigate the various elements addressing social exclusion issues, such as relations and social network, life satisfaction, involvement in social and political life, and economic conditions of people. All these surveys address social exclusion issues. Anyway, the current panorama of statistical surveys does not provide such a complete and comprehensive dataset in order to analyse in a complete manner the problem of social exclusion. In brief, it seems that the growing interest around social exclusion has not gone with an improvement in data collection. Notwithstanding the agreement about the multidimensional nature of social exclusion, researchers have to face toward a lack of data concerning relational aspects, social network, and social safety net. Moreover, we have underlined the necessity to include into the analysis belief, opinion and subjective evaluations, rather than simply objective measures, but also in this regard we remark a lack of data.

Among available socio-economic surveys, one of the most used is the European Community Household Panel (ECHP), which collects information on the living standard of the household of the EU member states. It contains detailed information on socio-economic conditions of individuals, however it focuses mainly on objective conditions and material deprivation (incomes housing amenities, consumer durables, employment conditions, and so on). Thus, ECHP does not allow examining in depth a number of aforementioned aspects of social exclusion, such as neighbourhood dimension, social relations, social safety nets, etc., as well as subjective evaluation of this condition. European Social Survey (ESS) poses a similar restriction. In this case, relational aspects are considered, but subjective evaluations are scarcely present. Moreover, this survey does not include all European countries. The Eurobarometer surveys represent another program of cross-national and cross-temporal comparative social research. The Eurobarometer series is designed to provide regular monitoring of the social and political attitudes among the European citizens[11]. This survey seems to represent, currently, the most suitable source of information in order to study social exclusion. In particular, Eurobarometer 56.1 (2001) refers explicitly to social exclusion. That survey investigates the respondents' point of view about social exclusion, thus allowing the adoption of a subjective perspective; for example, respondents were asked whether they felt left out of society, and which was in their opinion the main causes of social exclusion situations. Conversely, in ECHP and ESS, even investigating living standard, wellness, poverty and material deprivation, aspects that are all relied to social exclusion, one does not explicitly refer to this concept. Moreover, in Eurobarometer 56.1 survey we find rather satisfactory indicators for all relevant dimensions, the economic, the relational and the institutional one, even making some operational approximation. Although with some methodological and operational weak points, Eurobarometer surveys represent one of the most appropriate sources of information in order to study social exclusion following a comprehensive and a subjective approach.

An interesting characteristic of the Eurobarometer series is the continuity afforded by the repetition over time of some of the questions, included questions concerning poverty and social exclusion. The survey carried out in 2001 is the fourth in a series of studies of perception of poverty and social exclusion in European countries. The first was undertaken in 1976, the second in 1989, and the third in 1993. Additionally, in 2007 another survey focusing on poverty and social exclusion was undertaken. However, this continuity cannot be fully exploited, because of the change in terminology and the question wording across years. For example, several Eurobarometers investigate the principal elements that lead to social exclusion situations, but at each time one proposes diverse alternative items; more, we remark a change in the terminology adopted across time, which refers alternatively and indifferently to poverty, social exclusion, unease, situation of need, and so on. Particularly, the 2007 Eurobarometer survey focuses on material aspects of standard of living, neglecting relational issues. We acknowledge that changes in Eurobarometer questions wording during years may underline changes in social exclusion conceptual understanding, however, this temporal discontinuity prevents appropriate comparisons over time.

In summary, we think that data availability seriously hinder a comprehensive empirical approach to social exclusion. In social exclusion analysis, problems

[11] The issues investigated encompass a wide range of topics, e.g. quality of life, social security, consumer protection, environment, technology, social or ethnic exclusion health or family issues, and so on.

arise both from the intrinsic limitation of survey methodology, from the complexity and the extension of the issues addressed, and from the specific formulation of the survey's questions. We deem that, due to the relevance of the problem "social exclusion", a questioning of the adequacy of the current social survey urgently needs.

Chapter 3

An analytical and operational framework: dimensions of social exclusion

3.1 Introduction

In the previous discussion, we saw that it does not exist a clear and unanimously shared definition for the concept of social exclusion. Anyway, at the same time, we saw that different definitions agree in considering social exclusion a multidimensional concept that includes different forms of disadvantage and marginality. Social exclusion is a multidimensional concept that refers to different aspects of human life; that is, individuals may be excluded from different activities in their daily life. This implies that a wide range of living standard indicators may gather multidimensional deprivation. Indeed, each of them inevitably depicts only a single facet, a limited picture of the phenomena. These factors interact and it is the cumulating of these disadvantaged conditions that involve the higher risk of social exclusion. Therefore, an individual, or a social group, is most probably to be exposed to social exclusion risk when he experiences, for long time, difficulties in more dimensions of his life (Burchardt *et al*. 2002).

But which are these "multiple factors"? What are these forms of disadvantage and marginality? Which are the daily life activities relevant to determine social exclusion situations? As we discussed in Chapter 2, to date, the debate on social exclusion has been largely theoretical. Anyway, whether the objective is to achieve a deep understanding of the phenomena in order to give to policy makers a broad comprehension of the magnitude of the problem and of the different elements that interact, theoretical concept has to develop into an operational framework. We need an operational definition which, reckoning with data availability, identifies measurable domains of social exclusion. That is, after the definition of the relevant dimensions, it is necessary to pick out specific indicators that allow their quantification. Some attempts to operationalise the notion of social exclusion have been made (e.g. European Commission 1998, Burchardt *et al*. 2002, Ogg 2005, Böhnke 2008). Indeed, these works refer mainly to a theoretical approach, and they do not have to tackle operational problems. Therefore, it could be the case that some identified dimensions or categories of exclusion are impossible to measure by current available data. In other cases, the empirical application is limited to a particular context (e.g. Burchardt *et al*. 1999). Moreover, in social exclusion applied research there is still a tendency to focus primarily on the material dimension of social disadvantage (e.g. Moiso 2004, Whelan and Maitre 2005a) and to neglect the role of social relations. While in recent years the debate about social exclusion has turned attention to the multidimensionality of social disadvantage and its relational aspects, empirical applications have seemed to fail the multidimensional approach.

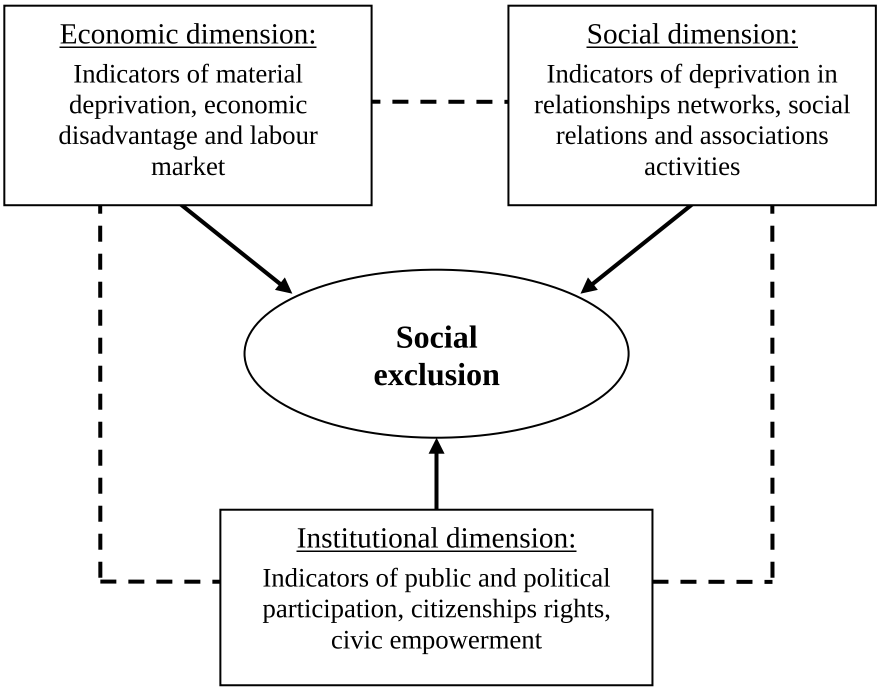
A comprehensive analysis of social exclusion requires a coherent framework. The attribute of "relativity" of social exclusion entails the requirement to define the context, in term of time and place, of the analysis. Different histories, cultures and demography, condition the relevant dimensions of exclusion; and the indicators or criteria to identify critical situations may have different weight depending on the reference frame[12]. In this chapter, we present the conceptual framework in which we perform our analysis. We consider both relational and distributive aspects, using an integrating approach: a modern society cannot disregard both an equally income distribution and the promotion of a high social cohesion. Referring to the fifteen countries belonging to European Union in 2001, there are three dimensions we consider appropriate to represent the spheres of human life in which it is most important for individuals to participate: the economic, the social and the institutional dimension. Social exclusion literature suggests similar conclusions (e.g. Commins 1993, Berghman 1995, Burchardt *et al.* 1999, Bhalla and Lapeyre 2004), but clearly this taxonomy is not unique and exhaustive, and not even definitive. Figure 3.1 depicts the proposed conceptual model. Social exclusion represents the multidimensional latent concept under investigation. Social exclusion has different components and it escapes to a precise and direct empirical measure[13]. It is a complex state that emerges when deprivation on material, cultural and social resources are so severe as to exclude people from the mainstream society. Thereby, we aim to describe and measure these complex and interrelated facets of social exclusion by means of a set of observed indicators. In order to yield a better understanding of how the processes of social exclusion trigger off, we decided to group these indicators referring to different dimensions. As the Figure shows, each one of the three identified dimensions is correlated to the other spheres of social exclusion.

---

[12] For example while racial problem is central in defining the significance and common understanding of integration in the United States, European people face the word "race" in a different way (Silver and Miller 2003). Or more, in discussions about social participation, citizenship rights may have a different value depending on the "part of the world" which one refers to (Bhalla and Lapeyre 1997). Moreover, in developing countries, the shortage of material and primary commodities has a different weight with respect to Western European countries. Whether both distributional and relational aspects of exclusion are relevant, distributional equity may be particularly important above all for low-income countries with very unequal income distributions and an inadequate presence of the social security system.

[13] Moreover, note that current social surveys neglect people who experience severe forms of social exclusion, such as the homeless or people in residential institutions, thus increasing the risk of bias.

Figure 3.1 – Proposed conceptual model



In next sections, we shall describe the main attributes of each dimension of social exclusion and the relations among them, describing the different aspects that indicators should take into account.

3.2 Economic dimension: multiple deprivation and disadvantages

In this section, we briefly review the existing principal concepts of poverty and deprivation, finding out the differentiations and the relations between them and the notion of social exclusion, in order to define our economic dimension.

The concepts of poverty and social exclusion are (inter)related and, to some extent, complementary. However, they do not imply the same elements and cannot be used as synonymous (Atkinson 1998). Both describe a situation of disadvantage and unease; both may be referred to different levels, namely individual and community; both imply elements of relatively and subjectivity. Anyway, they do not equal. Most literature recognize that even a minimum level of income remains a necessary requirement to ensure satisfaction of basic human need, it is not a sufficient one. Monetary and economic deprivation is merely a single, even if noteworthy, factor in social exclusion: while economic factors are undoubtedly a key aspect of social exclusion, social exclusion cannot be reduced to economic factors. There is evidence that throughout the EU well-off people are the most socially integrated, whereas poor or low-income ones run a higher risk of being socially disintegrated and excluded (Böhnke 2008). In general, the Mediterranean countries stand out as being characterized by a relatively minor degree of association between poverty and social disintegration, with respect to old EU countries of continental Europe and Scandinavia. However, heterogeneity among countries is large. What we want to underline is that people may be socially excluded without being poor (Atkinson 1998). And, conversely, they may be poor without being excluded from the mainstream society.

Our aim here is to define an economic dimension, in a way that could be useful to a better understanding of social exclusion. It is straightforward that the economic dimension relies directly to the concept of poverty, and it refers principally to monetary and financial aspects. Indeed, in a broader sense, it includes also people's capability to access to goods and services market and their employment condition. In this context, employment conditions are considered here in their elements linked to production and income.

When discussing about poverty, the key variable is customarily the "income". Both in absolute and relative terms, the income is used to define the individual (or the household) threshold of poverty. The *absolute* perspective of poverty builds on the assumption that there exist some primary goods needed for a physical survival of people, such as food, clothing and housing. So, individuals who cannot satisfy these minimal material necessities are classified as poor, or at least, at risk of poverty. Instead, following a *relative* approach, one has to compare the individual's socio-economic condition with that one of the others in the same society. In this sense the poverty risk arises when an individual fails in satisfying not just basically and primary needs but, more general, needs common in the society where he lives. Considering any Western European country, car, colour television, refrigerator, washing machine, going on holiday at least once a year, and so on, represent simple examples[14]. This set of needs is prone to continuous changes across countries but, above all, across time. The difference between the two concepts has been well remarked by Hagenaars and De Vos (1988, p. 212), who referred to the absolute and relative poverty respectively as:

a) Poverty is having less than an objectively defined, absolute minimum.

b) Poverty is having less than others in society.

With regard to the different methods of poverty measurement, we just briefly remind here that in order to measure poverty in absolute terms, one refers to a minimum threshold of income objectively defined as necessary to access a set of basic goods. Obviously, debates can be made about what the expression "basic and primary needs" refers to. Some limits of this definition are overcome using the relative definition, which, on the one hand, presents the advantage to consider individual situation within a particular socio-economic context, making comparative analysis more meaningful. On the other hand, the relative approach considers the income distribution in a given society (namely a Nation), and conventionally fixes the poverty line, trough which one discriminates between poor people and

[14] Note that we refer to needs in a broader sense, in terms of services and commodities.

not poor people, at 60% of the national median equivalised disposable income[15] (European Commission 2003, 2005). In this way, using in income–based measure for poverty and deprivation, the relative approach still accounts just for monetary aspect.

The definition of poverty given by Townsend seems to be more helpful for our purpose. "Individuals, families and groups in the population can be said to be in poverty when they lack resources to obtain the types of diet, participate in the activities and have the living conditions and amenities that are customary, or at least widely encouraged or approved, in the societies to which they belong. They are, in effect, excluded from ordinary living patterns, customs and activities." (Townsend 1979, p. 31). Clearly, this one is a relative definition. More, it stresses the fact that poverty affects in a general way the life of individuals. People's condition has to be qualified not only with respect to their financial resources, but also to their actual living conditions. Focusing on the multidimensional nature of poverty and disadvantage, Townsend emphasizes the fundamentally social character of poverty, since living conditions are not limited to material factors but include a large set of "necessities". In this definition, poverty is associated with lack of various kinds of resources, such as education, knowledge, employment, etc., which in turn can represent a way to participate in normal activities of life. Poverty and deprivation are not merely a negative condition per se, but also since they prevent people from participating in social life and exploiting all its possibilities (to participate in cultural and sporty association, to be engaged in recreational and leisure activities, to travel, to have an assurance coverage, and so on). Secondly, the use of deprivation indicators offers the possibility to measure poverty directly, opposite to the indirect measure assessed on the disposable income of individuals or households.

According to this conceptualisation, it seems to be more meaningful to broad the economic dimension in order to include also the access to goods and services market. In this sense, a low level of income may imply difficulties in the access to good and services market, thus originating exclusion from a suitable level of consumption. The most extreme situation is represented by homelessness, a situation that refers strongly to the absolute concept of poverty. But in a modern society, like that one of the industrialized European countries, the context is more complicated. In modern societies, some needs are not "essential" in a narrow sense, but they could be important to make people included in peer groups and neighbourhood activities. Once more, it emerges the *relativity* of social exclusion: it stands to reason that the relevance of exclusion in this domain is strongly related to context – time and place – where people lives.

Although the concept of poverty in last decades has been revised in a more comprehensive sense, another definition of poverty has gained ground in recent years. Hagenaars and De Vos (1988) add a third category that refers to people's feeling that they do not have enough to get along, which underlines the *subjective*

[15] Issues relating the rigid partition given by the criteria of the 60% of the median income are not treated here.

aspect of poverty, regardless of the absolute level of income. Two individuals or households with the same level of absolute resources may feel the situation differently and assess differently the difficulties they have to face with (Strobel 1996). Asking people about perceptions of their economic situation makes it possible to define subjective poverty indicators[16]. In this conceptualization, poverty is relative since in giving their responses people think to the general situation of the society where they live; but, at the same time, it is absolute in the sense that people have to think, even if implicitly, at an income threshold they consider sufficient to not be poor. Subjective poverty may be considered a measure of "economic pressure" (Robila 2006), which for some purpose is more useful than objective economic evaluation in assessing the overall quality of life. Research showed that the proportion of people who feels poor, is much greater than that of people considered as poor according to objective measures of poverty (Bhalla and Lapeyre 2004).

Referring to economic dimension at aggregate level, in addition to the overall level of income (individual or global), some authors suggest to consider also the inequalities in income distribution. Extreme imbalances in the income distribution within a State represent an alarming symptom that entail higher social expenditure and could trigger social conflicts and social exclusion.

To conclude, we showed how poverty and social exclusion are strictly connected and how the economic status may influence the degree of exclusion from certain social and institutional activities. Although, recently, more widespread meanings of poverty have been developed, the economic dimension still represents a partial aspect compared to social exclusion. Hence, in our framework the economic dimension will be accounted for using several indicators side by side, both from an objective and from a subjective perspective of the individual's standard of living.

3.2.1 Poverty, unemployment and social exclusion

Numerous studies dealt with the mechanisms that yields from unemployment towards poverty and finally towards social exclusion. Occupational status is relied to the individual capability of income production, thus an unemployed person, without a sure and fixed income, has surely a higher risk to fail in affording daily expenses. It is straightforward that one of the principal aims of the employment is to get money that allows participating in normal activities of human life, and to access to good and services market. According to the European Council, employment is the best safeguard against social exclusion (European Commission 2000). However, the findings in this regard are controversial: the empirical evidence does not seem to show a consistent relationship between social exclusion and unemployment across Europe. Atkinson (1998) argues that the extent to which a fall in unemployment generates social exclusion depends on the reasons

[16] In order to measure subjective poverty one usually uses appropriate survey questions asking how well do people get by with his household's income. Another solution often used is to ask for the amounts the households considered to be "sufficient" for their situation, in order to get a subjective minimum income level.

for the unemployment; moreover, he suggests that the creation of new jobs does not automatically contribute to fight social exclusion, but this relation depends on the kind of these new jobs, their remuneration, their features and their perspectives.

Indeed, we believe that exclusion from labour market is something more complex than unemployment; it might take into account other changes affecting the labour market, such as job security, precariousness and low-paid.

In recent years, unemployment rates have been reduced in most of European Union countries, but at the same time, the labour market has become more and more complicated and problematic. The proportion of working poor – persons that, even employed, do not have an adequate standard of living and face the problem of poverty – is augmented. These problems are connected to the low-paid and precarious works (Bhalla and Lapeyre 2004). The lack of a stable work and the difficulties to find a new job in case of temporary unemployment involve difficulties in planning one's own life for the future. The prolonging of these situations may determine a rise of physical and psychological pressure for individuals and, thus, worsening their sense of vulnerability. Particularly relevant it is the long-term unemployment, which involves higher social costs in a modern society, both at individual and at global level. The absence from the labour market for long time entails not just an output reduction (that is income for individual, and product for the entire system), but also a depreciation of human capital, skills and know-how, loss of dignity, and distrust toward future perspectives of reintegration into the labour market. These are all elements that may compromise the social fabric and the social stability.

Unfortunately, to date we do not dispose of appropriate data for this labour market segment – qualitatively and quantitatively – especially at European level. This aspect is important and noteworthy, and it should be object of further analysis, also in social exclusion issues.

3.3 Social relations

In this section, we intend to furnish a comprehensive view concerning social relations, combining theoretical questions and practical issues, the latter being relative mainly to the availability of appropriate indicators. We shall highlight the differences and the points of contact with the other dimensions, and explain in what sense "social dimension" is important for the complete development of human beings. Social relationships represent the core of the notion of social exclusion. They are the element that differentiate social exclusion in comparison to other similar and connected concepts (and often used, mistakenly, as synonymous), like poverty, deprivation, unease. Several authors have attempted to give a definition of social relations.

The term "social" is often used in different contexts and with different meanings, probably due to its ambiguous and broad-based connotations. Bhalla and Lapeyre (2004), for example, include in the social sphere of exclusion elements like crime, juvenile delinquency, homelessness. Others add also drug and alcohol addiction, and other deviant behaviours. All these elements are surely indicators of a break down of the social fabric, and they refer mainly to the society in its complex. High levels of these indicators should represent an alarm bell of a weakening of the social inclusion of some persons in the mainstream society. Anyway, it is not sure that deviant people, like drug addicted, feel excluded. Clearly, the term "social" deals also with other aspects of daily life, such as education and health. To some extent, these elements refer to social exclusion, in the sense that they could represent risk factors for individuals; in a recent research Ogg (2005) for example shows that the likelihood of experiencing social exclusion decreases with the number of years of education. Moreover, the presence of a bad health, disability or some kind of handicap, whether not adequately supported, could represent a limitation in the possibility of social participation, as well as a source of economic difficulties.

Anyway, what we mean here with "social dimension" refers primarily with the domain of relations among individuals. In the social dimension we include relational aspects with relatives, neighbours, friends, workmates or schoolmates, local community, and so on. As seen in Chapter 2, different authors (Berghman 1995, Kroenauer 1998, Burchardt *et al.* 1999) underline the relevance of social relations. Though using a differentiate terminology – family and community system, social activities, social networks – they all consider social interactions with family, relatives and friends, but more general, participation in social groups (cultural, sportive or other) and in community life.

We have to note that the social isolation, that is the weakening of the net of social ties, could depend on other dimensions of social exclusion. We showed in the previous paragraph how economic exclusion may entail difficulties in social participation. But territorial exclusion (Kroenauer 1998) also matter in this issue. Room (1995) states that social relations are the element that differentiates poverty from social exclusion, the former refers most on *distributional* issues and lack of material resources, while the latter deals with *relational* ones. We agree with Room in stating that the distinguishing features of social exclusion are the difficulty to participate in social activities of a community and the lack of social ties to the different level of a community. Conversely, we disagree with him when he considers the two concepts separately. We believe that the relation between the two concepts is more complex.

At individual level, social ties developed within family, friends, work environment or local community provide emotional and moral support. But we think that all these relational issues affect lives of individuals also in another manner, and in this sense the connection with distributional aspects discussed by Room comes into play. At individual level, relationships networks may be viewed as forms of social capital that can be activated when necessary, thus providing not only emotional support but material assistance too. Social relations may represent an irreplaceable support in compensate for possible economic difficulties, mainly in case of emergencies or transitional troubles. In this sense, it emerges that deprivation and material difficulties are an aspect of social exclusion, which in turn represents a wider concept. The presence of a reliable social network around individuals, constitutes a "life net" that may allow triggering mechanisms of solidarity. The family system represents an important factor in mitigating the impact of precarious material situations (Böhnke 2008). Social relations also operate as facilitators of access to information and contacts, playing an important role in overcoming unemployment (Granovetter 1985). Particularly, it is not merely the actual existence of social ties that matters, but also the potentiality, for individuals, to have confidence in one's own personal networks and to can rely upon them whether the need arises. The subjective expectation of remaining isolated in situations of need and the personal dissatisfaction with one's family life and participation in society, are warning symptoms of social exclusion. One signal of the existence of effective social relationships and networks is represented by the amount of practical and emotional support potentially available to individuals in times of need.

At macro level, all these elements combine to determine the sense of solidarity of a society and its social cohesion. The social participation of individuals in all its different forms represents an important indicator of integration, raising the sense of belonging to a social community. The strength of social networks and the possibility of feeling part of a wide community are conceived as a basic human need. Through social relations, people develop their own personality and realize themselves. Family cohesion, besides other elements like sense of solidarity, social responsibility, role of the social networks, religiosity and so on, contributes to delineate the "cultural profile" of a country.

Current research shows differences across European countries in cultural attitude toward familial and social relations, and in types and degrees of social support. In Mediterranean countries, family cohesion and solidarity prevail, while in Scandinavian ones, social contacts are less family-centred and concern mainly friendship relations and organized activities (Böhnke 2008). In Southern Europe, the family represents the main provider of care, and a mechanism to combat social exclusion in situation of labour market exclusion. Family is the first unit of social relations, and it remains the main provider of support in the absence of any state-protected welfare provision. Here the role of the (extended) family is to support a process of social integration even in case of economic deprivation. In these less individualistic societies, the tie cause-and-effect between economic deprivation and social relations deprivation is damped down on, and poor people is anyhow connected to a network of powerful and solid social relations. Thereby, one may find people that suffer for economic exclusion but not for relational one (Bhalla and Lapeyre 2004). However, nowadays, the traditional family structure of southern Europe is changing, and this fragmentation, together with the rising difficulties in the labour market, risks failing in protecting individuals from falling into a cumulative process of deprivation. This form of support, which should mitigate against some elements of social exclusion, risks to break down. Instead, in Nordic countries, people can find the major protection against poverty, unemployment and social exclusion typically in social and welfare system. Moreover in cases of need it seems more important the support from friends, workmates or neighbours, the role of the family being more limited.

Our hypothesis is that differences in national cultural models affect social exclusion processes. Moreover, it is conceivable to hypothesize that some differences remain also within each country, particularly in bigger ones, or in States which particular historical and cultural paths. One of our research hypothesis is, indeed, to verify if and to what extent differences in social relations patterns exist also at regional level, and whether they bear on social exclusion situations of their inhabitants (cf. Chapter 5).

As in the case of poverty, also social dimension might be view from a subjective perspective. The qualitative aspects are as important as the quantitative ones to explain social exclusion situations. Thus, social dimension should account for social contacts and for the subjective perception of these relationships, two aspects that may go differently according to contextual conditions. Moreover, individual appraisal is the unique manner to take into account elements like social recognition, sense of usefulness, self-esteem, and the identification of individuals in prevailing community's cultural and moral values.

3.3.1 Unemployment and social relations

As we saw before, unemployment is often view as a relevant risk factor in social exclusion situations, especially because of its association with poverty. In European discourse, fight against poverty and social exclusion aims to the eradication of unemployment, mainly the long-term one. Unemployment is analysed in terms of scarce economic resources, and economic and social dependence. Anyway, it may be associated also to a lack of recognition and usefulness.

The lack of employment determines a rising of social costs: it entails a loss of income (Atkinson 1998) both for individuals and for society, and more, it involves a break in the productive role of individuals, determining a loss of social legitimacy and social status as regards other people. As Amartya Sen observed employment is not only an income source, but it "*gives a person the recognition of being engaged in something worth his while*" (Sen 1975, p. 5). In this context, also the type of job and its characteristics become relevant. Various aspects relied to occupation, such as work satisfaction, the opportunity to take part in decision-making, the ability to learn, the quality and variety of tasks, etc., contribute strongly in the perception of marginalization and social exclusion situations. Secondly, unemployment may involve a weakening and a change in social network of individuals (Negri e Saraceno 2000): more long a person is unemployed, more probably his social net is formed by other unemployed.

Through employment, people have the possibility to become well-adjusted in the social structure and to have a clear role in the society, that increase their sense of utility. Participation in the labour market is a way to develop social contacts and social interaction; having a job that guarantee an income may be considered the most effective mechanism for integrating individuals into society and giving them social legitimacy.

3.4 Individuals are embedded in a society: the institutional dimension

While social dimension refers to relationships between individuals, the institutional one concerns relations between individuals and the State. In a sense, the former accounts for the private sphere of people, whereas the latter focuses on individuals as citizens.

Citizenship rights have been introduced in social exclusion conceptualization during the first half of the 1990s (Atkinson and Davoudi 2000), in order to bring together the French approach with the Anglo-Saxon one (cf. Chapter 2). The notion of citizenship rights refers to the denial of human, civil and political rights for certain population groups (Marshall 1964). From this point of view, the promotion of an active citizenship involves a wider opportunity of civil dialogue, thus assuring social cohesion through national and European societies. Conversely, non-participation contributes to disempowerment (Percy-Smith 2000): lack of interest and disengagement towards different forms of political, public and civic involvement may trigger mechanisms of social marginalization and social disorder. The citizenship rights issue recurs also in discussions and documents of European Union. In "Non-monetary indicators of social exclusion" (European Commission 1998) one includes "justice, education, health, political rights, bureaucracy" as risk factors for social exclusion situations. Literature about social exclusion usually adopts the Marshall's theory of citizenship (see Bhalla and Lapeyre 2004), according to which State should offer to its citizens:

- civil rights, like freedom of expression or right to justice;
- political rights, such as voting right and more general participation in political power;
- socio-economic rights, that is equality of possibilities, personal security, right to a minimum health care, and so on.

The denial or the violation of such rights and liberties for an individual, or for a group of individuals, may entail a situation of social exclusion. Clearly, the State has an irreplaceable role in the active defence and promotion of these rights; the ability of the States to promote and to encourage these rights may influence strongly social inclusion/exclusion processes. In this sense, the institutional processes – defined as the way and the degree to which individuals and groups are embedded within institutional systems and the effects these institutional systems have on individuals/groups (Atkinson e Davoudi 1998) – are crucial to the understanding of social exclusion.

In our conceptual framework, we identify an institutional dimension that includes relations between citizens and public institutions. These relations may be measured from an objective perspective in term of offer and enjoyment of civil, political and socio-economic rights. Thus, indicators concerning the access to right to justice or the limitation of personal freedom, the participation in the exercise of political power, or the right to personal security, to a minimum health care and so on, seem to well describe this dimension. Indicators that describe the environment in terms of rights, possibilities and public participation, fulfil the objective to approximate the socio-economic context in which individuals are embedded. In practice, however, there is a certain difficulty to account for some of these objective indicators at individual level.

Beside, one could consider also the subjective perception of how these rights work. For example, the extent to which people are satisfied by the medical services, the social security policies, and other kinds of civil, political and socio-economic rights, provides a measure of the social empowerment. This perspective allows introducing in the analysis an element concerning the detachment of individuals from public institutions and policies, and how it affects perception of social exclusion and social exclusion situations.

Finally, two points is worthwhile to stress here. Firstly, the institutional dimension is a clear example of how the relevance assumed by these rights, their accessibility and their significance depend on the context of reference. In democracy context, the freedom of speech assumes surely a strongly different meaning respect to countries where these human basic rights are denied. Secondly, exclusion from citizenship rights prescinds from the fact that an individual wants actually take advantage of those rights; he should have the possibility to enjoy rights and liberties independently from the fact that he actually exercises them. Thereby, the recognition (implicit or explicit) of citizenship rights does not automatically imply their actual fulfilment for everybody. This is particularly true for certain social groups (immigrants, racial minorities, women, elderly people, and so on).

3.4.1 The role of Welfare systems

We discussed about the differences in sociability and relational models across Europe, and about the way they affect social exclusion situations. Together with familial models, also social policy models play an important role in the relations between unemployment and poverty, and between poverty and social exclusion. Economic difficulties, limitation in the access to goods and services market, situations of disease or disabilities, could be mitigated by the effect of social protection policies (Atkinson 1998, Mayes *et al*. 2001), e.g. through social transfers and measures of social assistance. Tsaklogou and Papadopoulos (2002) found that the relationship between a country's welfare regime and the risk of social exclusion its population faces varies across countries. The effectiveness of social transfers and social benefits in reducing the proportion of people at risk of poverty varies greatly in different countries (Bhalla and Lapeyre 2004). The total expense per capita on social protection, the way the tax and benefit systems are structured, the age structure of the population, the overall income distribution and the general economic situation, are the main factors that determine these differences in efficiency.

These findings cannot be ignored. We need tools and methodologies in order to measure and to account for these elements, too. Regional and welfare system differences could be analysed through different statistical methods. Current comparative studies usually introduce dummies in the model; for example, a strategy is to consider the complete sample of the European population together, and to include a set of country-specific dummy variables and/or a set of dummy variables corresponding to the welfare regimes that can be found in EU. Another possibility, not much exploited in this field, is to perform a multilevel analysis. In presence of a hierarchical structure of the data (e.g. individuals nested in regions), which implies dependence among observations within the same group, multilevel analysis allows examining whether differences in social exclusion situation/perception are due to the regional or welfare regime context rather than to individual characteristics. The latter one represents the methodology we shall apply to account for the different contexts in our empirical analysis (see Chapter 4 and Chapter 5).

Chapter 4

Studying unobservable concepts: Latent Class Analysis

4.1 Latent Variable Models

Many empirical researches in the field of social sciences involve the study of not directly observables concepts: intelligence or skills, attitudes, medical conditions, personality traits, preferences and perceptions, and so on. We believe that social exclusion is one of them. In these cases, researcher may only observe direct indicators he assumes to be related in some way with the underlying latent concept of interest.

The properties of the latent variable(s) must be inferred indirectly, using a statistical model connecting unobserved variables to observed ones. The latent variables can be thought as representing "true" variables or constructs, and the observed variables as indirect or fallible measures (Skrondal and Rabe-Hesketh 2007).

The traditional classification of the different kinds of Latent Variable models (Bartholomew *et al.* 2002) considers the assumptions concerning the level of measurement of the manifest and the latent variables (Table 4.1). Factor Analysis (FA) and Latent Trait Analysis (LTA) deal both with continuous normally distributed latent variables. The difference is that FA is concerned with manifest and latent variables, both representing continuous indicators, while in LTA the manifest indicators are discrete. In Latent Class Analysis (LCA) all the variables used in the analysis are (or are treated as) discrete, whereas Latent Profile Analysis (LPA) represents a combination of discrete latent variables and continuous manifest ones. This scheme represents only the traditional and simplistic classification.

It is worthwhile noting that sometimes this distinction is not so fundamental at all, depending on choices made by the researchers. Usually, the specification of the conditional distributions of the indicators follows naturally from their scale type, whereas the latent variables can be assumed to be both continuous and categorical, or a combination of the two.

There are some connections among these four kinds of statistical methods, but it is straightforward that their differences do not pertain only to the scale type of the variables. For example, although LCA is often considered the categorical-data analogue to FA, the latter is concerned with the structure of variables (i.e.

their correlation), whereas LCA is more concerned with the structures of cases, providing classification of individuals. Moreover, LCA and LTA are commonly considered as two variants of the latent structure analysis: the difference is that the latent variable that determines data structure is nominal with the LCA, and continuous (i.e. a latent continuous trait) with LTA. The possibility to accurately approximate the distribution of a continuous latent variable in a LT model by a discrete one without a lack of fit is for example discussed in Heinen (1996) and Bartholomew *et al.* (2002). Moreover, in between LCA and LTA, one may place the "*discrete latent class models*" (Heinen 1996), for which the latent variable is discrete and unidimensional as with LCA, but the classes are viewed as ordered along a latent continuum, as with LTA.

Forasmuch as the difficulties to distinguish empirically among the Latent Variable models[17], the most appropriate modelling framework should be selected by the researcher depending on the specific application and its objectives, considering at the same time methodological issues of the Latent Variable models and substantive and practical meaning of the problem under investigation.

Table 4.1 – Classification of Latent Variable modelling

	Manifest variables	
Latent variable(s)	*Continuous*	*Categorical*
Continuous	Factor Analysis	Latent Trait Analysis (or IRT)
Categorical	Latent Profile Analysis	Latent Class Analysis

Source: Bartholomew et al. (2002).

Indeed, it is a common practice to introduce a mixture of metrical and categorical variables as manifest indicators, thus expanding the conceivable types of models. Thus, the above classification could be broadened differentiating for continuous variables or restricted continuous variables, categorical-dichotomous, categorical-nominal, categorical-ordinal or categorical-quantitative variables, or considering other possibilities that mix all these types. As remarked by Clogg *et al.* (1995) there are as many types of Latent Variable models as the types of variables and of their combinations.

Actually, latent variables are widely used to capture a wide variety of statistical concepts, although under different names and different forms (Muthén 2002): common factors, latent classes, random effects, underlying variables, frailties, components of variation, missing data, finite mixtures. In recent literature, various

[17] See also Bartholomew *et al.* (2002) and Bartholomew and Knott (1999) for further details.

authors, recognizing the mathematical similarity of a wide range of Latent Variable models, propose very general frameworks for Latent Variable modelling, integrating specific methodologies in a global theoretical context. The Generalized Linear Latent and Mixed Models (GLLAMM) framework proposed by Skrondal and Rabe-Hesketh (2004) includes models for continuous and discrete latent variables. GLLAMM consist of two building blocks (Skrondal and Rabe-Hesketh 2007), formally described by two equations:

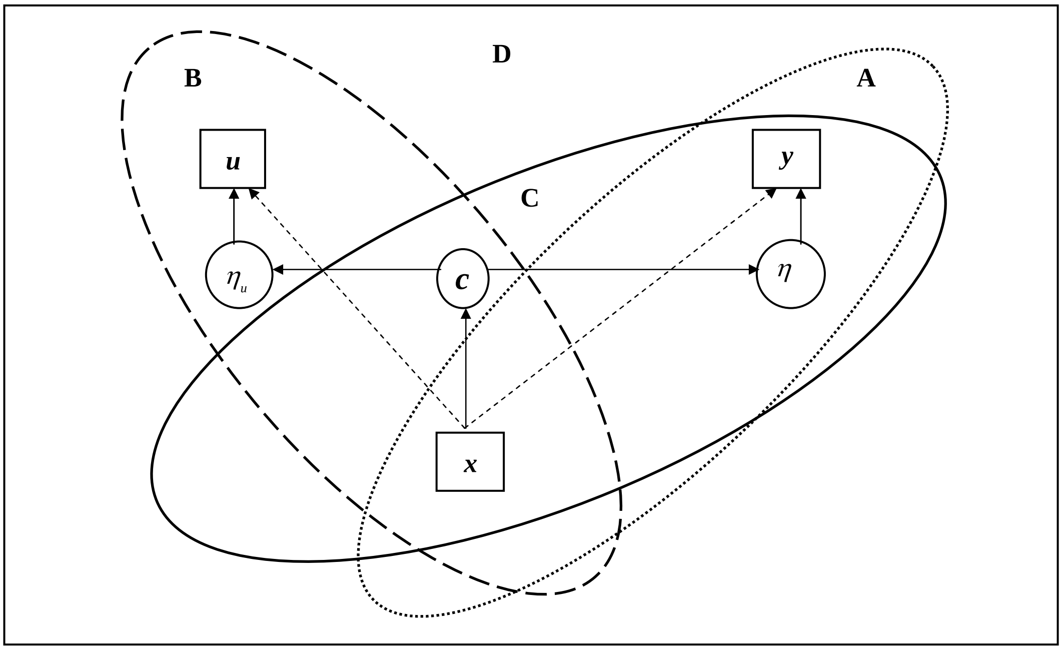
1. the conditional distribution for the response variable(s), given the latent variables and the observed covariates. The conditional expectation of the responses y, given the (vector of) latent variable(s) X and the covariate(s) Z, is linked to the linear predictor via a link function $g[\cdot]$;
2. the structural model for the latent variable(s), which specification depends on the nature of the latent variables, either continuous or categorical, or both.

The choice of the appropriate components specification with respect the nature of the phenomena under study, yields a specific Latent Variable modelling, including hierarchical models with unobserved variables at different levels. The models are implemented in the GLLAMM package of the Stata software (Rabe-Hesketh *et al.* 2004).

Another attempt to unify and extend a wide variety of common types of Latent Variables analysis into a general single modelling framework refers to Muthén (2002). The generality of the model is achieved by considering both continuous and categorical latent variables. This global framework has been developed together with the software Mplus (Muthén and Muthén 1998-2007), through which it is possible to implement models with different specifications concerning the variables types.

In this general modelling and software framework represented in Figure 4.1, one identifies the well-known case of continuous latent variables (sub-framework described by ellipse A in Figure 4.1), and the case of categorical latent variables (ellipse B, that includes, e.g., latent class analysis and latent class growth analysis). Ellipse C is characterized by adding categorical latent variables to A, thus yielding models that combine continuous and categorical latent variables, such as finite mixture modelling, latent profile models, growth mixture modelling and mixture SEM. Finally, by adding categorical indicators to framework C, Muthén defines the framework D, which shows the modelling generality, achieved by a combination of continuous and categorical latent variables. This framework contains many special cases and suggests many new types of models, deriving from model combinations. Note that Figure 4.1 is merely a simplification, the general framework allowing also other direct effects between variables. This framework has been developed to include multilevel models too (see Muthén and Muthén 1998-2007 for further details).

Figure 4.1 – The general Latent Variable modelling framework proposed by Muthén



Note: the squares represent observed variables, where y denotes a vector of continuous response variables, u a vector of categorical response variables, and x a vector of covariates. The circles represent latent variables, both continuous (η) or categorical (c). The arrows represent direct effects between variables.

Source: Muthén (2002).

Finally, Jeroen Vermunt (2003, 2007, 2008) proposes a framework that allows to define models with any combination of categorical and continuous latent variables. This framework is implemented the 4.5 version of Latent GOLD software (Vermunt and Magidson 2008). Latent GOLD is a software for conducting Latent Class Analysis. However, the option to include one or more continuous latent variables in a LC model – that traditionally deals with categorical latent variables – extends Latent GOLD to a more general Latent Variable modelling program. More generally, in the Latent GOLD framework, LC model refers to any statistical model in which some of the parameters differ across unobserved subgroups. Particularly, the syntax version of the program includes a list of options, variable specifications and equations that provide flexibility for developing alternatives. Models can be handled specifying the appropriate distribution with respect to the required data organization and the nature of all types of variables concerned (latent, dependent, independent, covariates and grouping variables). In this way, Latent GOLD can be used to implement the most important social science application types of LC and finite mixture models, and a wide repertoire of Latent Variable models: factor analytic models, IRT models, latent growth models, discrete factor and continuous factor models, path models, random effects regression models (Vermunt and Magidson 2005b). Another important extension encompasses mixture and multilevel variants of these models, which allow defining models with any combination of discrete and continuous latent variables at each level of the hierarchy.

The Latent GOLD framework is strongly related to GLLAMM (Vermunt 2007). The models proposed by Vermunt fit very naturally into the general modelling framework developed by Skrondal and Rabe-Hesketh (Skrondal and Rabe-Hesketh 2004, 2007). The most important extension of Latent GOLD compared to the GLLAMM approach is that it allows defining models with any combination of discrete and continuous latent variables at each level of the hierarchy (Vermunt 2007). Particularly, multilevel extensions of non-traditional LC model are implemented, allowing the inclusion of group-level latent classes and/or group-level continuous factors.

This dissertation concerns with Latent Class Analysis, thus we find in Latent GOLD a natural framework and the possibility of a flexible implementation of our application (see Chapter 5). After a brief history of Latent Class models (§ 4.1.2), in this chapter we review the theoretical literature of the standard Latent Class models underlying their principal features (section 4.2), we analyse their principal extensions (section 4.3), then we describe the possibility for LC models to account for hierarchical structures of the data (section 4.4). Finally, issues concerning model estimation and model evaluation are treated (section 4.5). An attempt to unify different notations and formulations found in literature has been made.

4.1.2 A brief history of Latent Class Analysis

Latent Class analysis was originally introduced in the fifties by Lazarsfeld (1950). His aim was to build typologies based on dichotomous observed variables, in order to explain respondents' heterogeneity in survey response patterns. More than 20 years later, Goodman (1979) developed an algorithm for obtaining maximum likelihood estimates for the model parameters, making a fundamental contribution for the applicability in practice of these models. He also formalized and extended models to polytomous manifest variables and multiple latent variables, and did important work on the issue of model identification (Goodman 1974a). Over the same period, Haberman (1979) showed the connection between LC models and log-linear models for frequency tables with missing (unknown) cell counts, and the work of Day (1969) and Wolfe (1970) contributed to the emergence of the related field of finite mixture models for multivariate normal distributions.

In last decades, many important extensions of the classical LC model have been proposed: introduction of covariates (categorical and continuous), relaxing of local independencies, introduction of several latent variables or of ordered latent variables, and repeated measures. These developments have broadened the fields of applications of Latent Class models (Haagenars and McCutcheon 2002), enhancing the potentiality of LCA as technique in social and behavioural research. A general framework for categorical data analysis with discrete latent variables was proposed by Hagenaars (1990) and extended by Vermunt (1997). Furthermore, as underlined in Goodman (2002), in presence of heterogeneous populations LC analysis can be viewed as a probabilistic cluster analysis tool, where each latent class represents a hidden cluster: detecting the presence of latent categories or latent types of individuals, LCA allows the study of multiple unidentified groups that behave differently regarding the problem at hand. Other interesting application areas deal with measurement error in nominal and ordinal indicators. Finally, LC and finite mixture models can be useful in several other areas as well. One of these is density estimation, in which one makes use of the fact that a complicated density can be approximated by a finite mixture of simpler densities (Vermunt and Magidson 2004).

4.2 The classical Latent Class model

In social and behavioural sciences it is frequent to dispose of categorically scored observed measures that are highly interrelated, and where association is due to some underlying unobserved factors (or dimensions). Latent Class Analysis offers a powerful tool for investigation and analysis of such situations, in both explorative and confirmative ways. The latent classes are the levels of a categorical latent variable, which structure the cases with respect to a set of observed indicators allocating each unit to one of the classes. The latent classes are mutually exclusive, and they are supposed to account for the dependencies among the observed indicators.

The primary object of LCA (McCutcheon 1987) is to identify a latent categorical variable accounting for the dependencies among the observed indicators. That is, to identify a number of mutually exclusive classes within which the manifest indicators are locally independent. Secondly, the latent class membership, which is not known but inferred from observed data, steps up the understanding about the interrelationships and the associations between the manifest variables, thus representing a method of data reduction for complex datasets (Bartholomew *et al.* 2002). Enlarging attention also to non-observable conditioning variables, the methods of LCA basically seek for possible structures of conditional independence in multivariate data (Clogg 1995). Goodman (2002) illustrates some different uses of Latent Class Analysis: as measurement model, as a tool to investigate the unobserved heterogeneity among the respondents, and as a probabilistic classification method of each respondent on the basis of their responses.

From a statistical point of view, a LC model consists of three parts (Vermunt and Magidson 2005b):

1. the assumed probability structure, which defines the relevant set of conditional independence assumptions among the variables in the model;
2. the assumed distributional forms for the response variables, which depend on the scale types of the variables selected;
3. the regression-type constraints used to gain parsimony in the description of the relationships between the observed variables in the model.

Originally, LCA applied on datasets where both latent and manifest variables were categorical (Lazarsfeld and Henry 1968, Goodman 1974b, Haberman 1979). Anyway, due to recent developments concerning the nature of manifest indicators, we prefer to use a notation that applies in a more general framework, allowing to work with observed variables of any scale type. For the cases concerning only discrete variables, one can substitute the symbol $P(\cdot)$ instead of $f(\cdot)$ to indicate one is dealing with a probability instead of a density function. Moreover, here we propose a notation that attempts to make different formulations found in literature consistent. Various authors use a different notation to explain the characteristics of LC models. However, since we review this modelling, beginning from the basic unrestricted Latent Class model, exploring their characteristics and arriving to more complex models, we think we take advantage of using a consistent and congruent notation.

Suppose data are available for units (i.e. individuals) denoted $i = 1,...,I$. $\mathbf{Y}_i$ refers to the full vector of responses of the same unit i on a set of K indicators (items), with $k = 1,...,K$. X_i represents the underlying latent discrete variable, for which a particular latent class is denoted by t and the number of latent classes by T. We formulate the probability structure of the responses as follows:

$$f\left(\mathbf{Y}_i\right) = \sum_{t=1}^{T} P\left(X_i = t\right) f\left(\mathbf{Y}_i \middle| X_i = t\right) \qquad [4.1]$$

The idea is that respondents belong to one (and only one) of the T latent classes, and that the multiple responses are generated by class-specific densities (or probabilities). A Latent Class model aggregates probabilities of classification in the latent classes $P\left(X_i = t\right)$, as well as conditional densities $f\left(\mathbf{Y}_i \middle| X_i = t\right)$ of outcomes for each $\mathbf{Y}_i$ within latent class. These two sets of probabilities and densities represent the model parameters.

Equation [4.1] shows the idea underlying the LC models: the density $f\left(\mathbf{Y}_i\right)$ corresponding to a certain response pattern of the i-th unit, is a weighted average of the T class-specific densities $f\left(\mathbf{Y}_i \middle| X_i = t\right)$. The weight is the proportion of persons belonging to the latent class t, $P\left(X_i = t\right)$. In other words, the density $f\left(\mathbf{Y}_i\right)$ results from a mixture of class-specific densities $f\left(\mathbf{Y}_i \middle| X_i = t\right)$ (the *mixture distribution*) and *mixing weights* $P\left(X_i = t\right)$.

Actually, LC models represent a case of the more general class of mixture models. Particularly, Latent Class models have the same structure of the Finite Mixture (FM) models, which seek to separate out data that are assumed representing a mixture of a finite number of different subpopulations (McLachlan and Peel 2000). The subpopulations membership is not known but is inferred from the data. In Latent Class Analysis, these different subpopulations can be seen as different dimensions of the same (unknown) latent construct. While generalized linear mixed models create a mixture of linear predictor values using a latent variable having a continuous distribution (Agresti 2002), Latent Class models use a mixing distribution that is qualitative rather than quantitative. In recent years, the two fields of LC models and FM models have come together[18], and LC models is the name usually used in social sciences, while in more general statistical context the term finite mixture models is preferred. In this thesis the two terms will be used as synonymous.

4.2.2 Local independence assumption

At the basis of LC models is the local independence assumption. Local independence assumption implies that the response variables are mutually independent given the latent variable. This means that within the latent classes only random relationships among variables remain: LCA structures the units into latent classes so that the indicators are uncorrelated within each class. Actually, this assumption is common to all Latent Variable models, such as factor or item response models, where the responses are conditionally independent given the *continuous* latent variables (factors or traits).

Using Y_{ik} to indicate the response of unit i on a generic indicator k, the local independence assumption can be expressed by

$$f\left(\mathbf{Y}_i \middle| X_i = t\right) = \prod_{k=1}^{K} f\left(Y_{ik} \middle| X_i = t\right) \qquad [4.2]$$

i.e. the observed items Y_{ik} are mutually independent given the individual's score on the latent variable. Using [4.2] to define the local independence structure, the LC model [4.1] becomes:

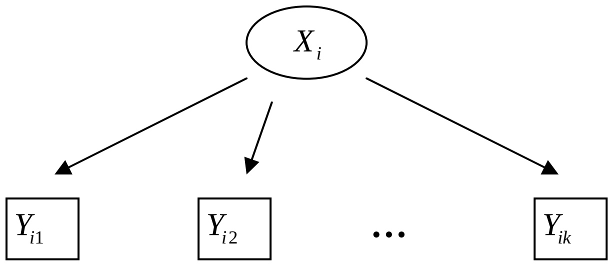
$$f\left(\mathbf{Y}_i\right) = \sum_{t=1}^{T} P\left(X_i = t\right) \prod_{k=1}^{K} f\left(Y_{ik} \middle| X_i = t\right) \qquad [4.3]$$

Figure 4.2 shows the path diagram for the standard urestricted Latent Class model. Following the convention of path diagrams, the circle represents the latent variable, the rectangles the manifest indicators, and the arrows the relations. In the figure, the absence of arrows connecting individual responses Y_{ik} stands for the local independence assumption: they are not connected directly, but only indirectly through the common latent variable X_i. The latent variable is assumed to explain all the associations among the manifest variables. Determining the small-

[18] Anyway, while the terms latent class and mixture model are sometimes used synonymously, they have distinct early literatures, Latent Class models focusing on categorical outcomes (Goodman 1974, Lazarsfeld 1950, Lazarsfeld and Henry 1968) and finite mixture models focusing on continuous outcomes (Day 1969, Wolfe 1970).

est number of latent classes sufficient to account for the associations among the manifest variables is the principal aim of traditional LC analysis.

Figure 4.2 – Path diagram of the Latent Class model [4.3]



Anyway, it is worthwhile noting that, for some applications, conditional independence may represent an inappropriate or unrealistic assumption. Consider for instance a survey that presents two very similar items, such that responses on them are probably associated. Moreover, sometimes the violation of this assumption determines a lack of model fit (Clogg 1995). The presence of local dependence between two indicators implies some overlapping information between them, so (partially) relaxing the independence assumption could yield to a better classification, when LC analysis is used as a clustering tool (Hagenaars 2002). We shall account for this specification in the following sections.

4.2.3 Parameters of LC models

In Latent Class models the parameters consist of *unconditional* and *conditional* densities (probabilities). The first ones, also called *latent class probabilities*, describe the distribution of the latent variable; the *conditional densities* (*probabilities*) characterize the distribution among the indicators (observed variables) conditional on the latent classes.

The *latent class probabilities* $P(X_i = t)$ are the probabilities associated to the levels of the latent variable within which the observed measures are (locally) independent of one another. The *number of classes* represents the number of latent levels defined by the LC model for the observed crosstabulation. The *relative size of the classes* provides information about the distribution of the population among the T classes: the population can be evenly distributed among classes, or some latent classes can represent relatively large segments of the population while other class relatively small ones (McCutcheon 1987). These parameters are particularly helpful when the objective of the analysis is the comparison of populations: two populations could have similar latent structures but differ in the class size distribution.

Since the marginal distribution of X implies a mutually exclusive and exhaustive classification, the sum of the latent class probabilities over the T latent classes must equal one:

$$\sum_{t=1}^{T} P\left(X_i = t\right) = 1 \qquad [4.4]$$

with $P\left(X_i = t\right) \geq 0$. Given [4.4] there will be at most $T-1$ nonredundant parameters for X. This restriction implies that there is a latent class for each of the possible response patterns observed in the data.

The second type of parameters in LC models are the *conditional densities* (*conditional probabilities*) $f\left(Y_{ik} \middle| X_i = t\right)$, $k = 1,...,K$, representing the probability densities for an individual at level t of the latent variable X, to have a certain response pattern. E.g., considering categorical indicators, for each of the T classes of X there is a set of conditional probabilities equal to the number of levels measured for each of the observed indicators.

As for latent class probabilities, within each of the T latent classes the conditional densities for each of the observed variable integrates to one

$$\int_{S_k} f\left(Y_{ik} \middle| X_i = t\right) = 1 \qquad [4.5]$$

for continuous indicators, and sum to one

$$\sum_{S_k} P\left(Y_{ik} \middle| X_i = t\right) = 1 \qquad [4.6]$$

for nominal indicators, where S_k is the support for the k-th indicator.

While latent class probabilities serve to describe the distribution of the population among the latent classes, the conditional densities allow characterizing the nature of the latent variable, defining the profiles identified by each latent class. As McCutcheon remarks (1987), the conditional densities of LC models are comparable to the factor loadings in factor analysis, which represent the correlation between the observed variable and the factors[19]. For practical purposes, conditional densities/probabilities enable the researcher to characterize the structure of the latent typology, and they are used to name the latent classes. The inspection of these parameters allows discovering to which of the T classes an individual with a certain response pattern is most likely to belong. Units within the same latent class are homogeneous on certain criteria, and conversely, units in different latent classes are dissimilar on others.

[19] However, in contrast to factor analysis, LCA enables also to classify respondents.

4.2.4 Conditional distributions

To completely specify the joint response distribution under the conditional independence assumption of the standard LC model [4.3], one has to define the distributional form for the latent and the observed variables.

The categorical values of the latent variable X_i are assumed to follow a multinomial distribution, parameterized as follows:

$$P\left(X_i = t\right) = \frac{\exp\left(\eta_t\right)}{\sum_{t'=1}^{T} \exp\left(\eta_{t'}\right)} \qquad [4.7]$$

In case of a single nominal latent variable, this yields a standard multinomial logit model, which linear predictor is:

$$\eta_t = \gamma_{0t} \qquad [4.8]$$

The intercept parameters γ_{0t} are subjected on identifying constraint, that is $\sum_{t=1}^{T} \gamma_{0t} = 0$ in case of effect coding, or $\gamma_{01} = 0$ or $\gamma_{0T} = 0$ in case of dummy coding[20].

Recent versions of the standard model extend to the case of multiple and/or ordinal latent variables. In these cases it would be easy to modify the multinomial probability defining a standard and/or multivariate version of the adjacent-category ordinal logit model (e.g. Agresti 2002, Vermunt and Magidson 2005b).

A particular distributional form and a linear predictor has to be assumed also for the response manifest variables Y_{ik} , and it stands to reason that the scale types of the indicators are relevant for the distributional form.

Usually, a multinomial distribution is preferred for nominal variables so, for a generic response Y_{ik} – assuming that response variable k is independent of the other response variables – the distribution is of the form:

$$P\left(Y_{ik} = s_k \left| X_i = t\right.\right) = \frac{\exp\left(\eta_{s_k t}\right)}{\sum_{s'=1}^{S_k} \exp\left(\eta_{s't}\right)} \qquad [4.9]$$

with linear predictor:

[20] Effect coding means that the parameters will sum to zero over the categories of the nominal variable concerned, so the category-specific effects should be interpreted in terms of deviation from the average. Conversely, in dummy coding one selects a reference category for whom the corresponding parameters are equated to zero. In dummy coding the category-specific effect interpretation is in terms of difference from the reference category.

$$\eta_{s_k t} = \beta_{0 s_k} + \beta_{1 s_k t} \quad [4.10]$$

where β_{0s_k} represents the indicator intercept, and β_{1s_kt} is the class-specific main effect of the latent variable X_i on the indicator, for each $t = 1, ..., T$.

For identifiability reasons, a set of constraints has to be imposed, i.e. $\sum_{s'=1}^{S_k} \beta_{0s'} = 0$ and $\sum_{s'=1}^{S_k} \beta_{0s't} = 0$ in case of effect coding; $\beta_{01} = 0$ or $\beta_{0S_k} = 0$ and $\beta_{01t} = 0$ or $\beta_{0S_kt} = 0$ in case of dummy coding.

Note that in unrestricted standard LC models, only the single variable parameters along with the two-variable parameters between the latent variable and each of the indicator are included, whereas all higher-order terms involving combinations among the indicator variables of the loglinear LC model are set to zero. In this way, one expresses the assumption of local independence among the indicators in the logit models.

For an ordinal response variable, the nominal logit model can be replaced by an adjacent-category ordinal logit model, with linear predictor:

$$\eta_{s_k t} = \beta_{0 s_k} + \beta_{\cdot 0t} \cdot y^*_{s_k} \quad [4.11]$$

where $y^*_{s_k}$ is the score assigned to category s_k of the k -th indicator.

If Y_{ik} is a continuous response, the univariate normal distribution $N\left(\mu_{k,x}, \sigma^2_{k,x}\right)$ (or a multivariate normal distribution in the multivariate case) is the usual choice:

$$f\left(Y_{ik} \,|X_i = t\right) = \frac{1}{\sqrt{2\pi\sigma^2_{kt}}} \exp\left\{-\frac{1}{2}\frac{\left(Y_{ik} - \mu_{kt}\right)}{\sigma^2_{kt}}\right\} \quad [4.12]$$

The nominal, ordinal and continuous responses are the most common in empirical applications, but other scale types of indicators can be easily accommodate. Vermunt and Magidson (2005b) illustrate the case of counts or number of events, censored or truncated data.

4.3 Extensions of the standard model

4.3.1 Introduction of covariates

A natural extension of the standard model implies the introduction of individual covariates (e.g. Clogg 1981, Dayton and McReady 1988, Hagenaars 1993), that may affect both latent classes and conditional probabilities, thus helping to predict class membership. While indicators are used to define and measure the latent classes, the covariates operate as explicative variables, useful to improve the description of the latent classes in terms of individual characteristics (e.g. demographic). Their introduction may improve the prediction and reduce classification error. Covariates are considered exogenous variables, in the sense that the associations among them are not explained by the latent variable (Magidson and Vermunt 2004).

Extending the standard LC model [4.3], the general mixture model that defines the relationships between the exogenous (Z_i), latent (X_i) and response ($\mathbf{Y}_i$) variables at individual level, can be formulated as follow:

$$\begin{aligned} f\left(\mathbf{Y}_i \middle| Z_i\right) &= \sum_{t=1}^{T} P\left(X_i = t \middle| Z_i\right) f\left(\mathbf{Y}_i \middle| X_i = t\right) \\ &= \sum_{t=1}^{T}\left(X_i = t \middle| Z_i\right) \prod_{k=1}^{K} f\left(Y_{ik} \middle| X_i = t\right) \end{aligned} \qquad [4.13]$$

The multinomial logistic regression model expressing the probability that individual i falls in class t of the latent variable X_i as a function of the covariate is:

$$P\left(X_i = t \middle| Z_i\right) = \frac{\exp\left(\eta_{tZ_i}\right)}{\sum_{t'=1}^{T} \exp\left(\eta_{t'Z_i}\right)} \qquad [4.14]$$

where the linear predictor[21] is

$$\eta_{tZ_i} = \gamma_{0t} + \sum_{r=1}^{R} \gamma_{rt} \cdot z_{ri} \qquad [4.15]$$

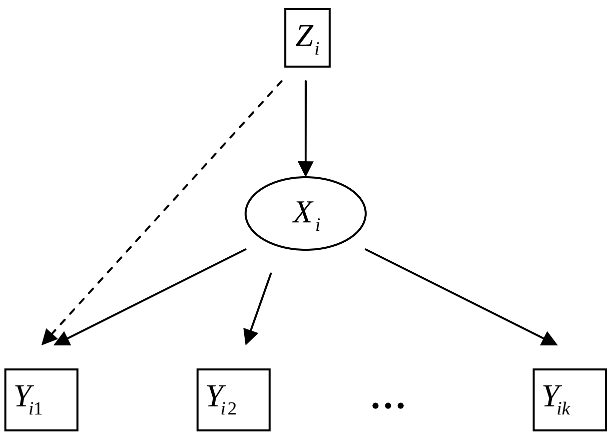
under the usual identifying constraint on the intercepts in the effect coding $\sum_{t=1}^{T} \gamma_{rt} = 0$, or in the dummy coding $\gamma_{r1} = 0$ or $\gamma_{rT} = 0$.

[21] The logit model may provide for interactions between covariates, even if, actually, it is not common.

As in the standard model [4.3], the most right-hand side of equation [4.13] expresses the local independence assumption $f\left(\mathbf{Y}_i \middle| X_i = t\right) = \prod_{k=1}^{K} f\left(Y_{ik} \middle| X_i = t\right)$. Note that model [4.13] implies an additional set of conditional independence assumptions: i.e. the indicators are assumed to be independent of the covariates given the latent variable X_i (Vermunt and Magidson 2005b).

A usual (and useful) further implementation concerning covariates involves the inclusion of *direct effects* of covariates on indicators. In Figure 4.3 the dotted arrow between the covariate Z_i and the indicator Y_{ik} represents this effect. Following this approach, it is assumed that the covariate Z_i and the observed indicators Y_{ik} are not conditionally independent given the latent variable. In this way, as respect what is illustrated by the [4.13]-[4.15], one relaxes the assumption that the influence of the covariates on the indicators goes completely by the latent variable. We will account formally for this situation in next paragraph, which deal with the treatment of the unexplained association in the estimated model.

Figure 4.3 – Path diagram for a Latent Class model including covariates



4.3.2 Relaxing local independence assumption

As stated before, lack of fit of a LC model may result from the violation of the local independence assumption. Whether a certain solution in terms of number of latent classes does not completely account for the relations between the manifest indicators, the traditional strategy is to increase the number of classes until an acceptable fit is obtained. Anyhow, different reasons could be mentioned (Hagenaars and McCutcheon 2002) in favour of less traditional alternatives, leading to more parsimonious, as well as congruent, models.

Sometimes local dependency is due to many redundant variables that yield overlapping information among different indicators. In other situations, it happens that an external factor, unrelated to the latent variable, is responsible for the local dependence. In social science research this latter case is frequent in presence of a similar question wording used in two (or in battery of) survey items. In certain situations it is possible that the meaning of an indicator is somewhat different for different subgroups (e.g. by sex, age, and so on) distinguishable in the population in terms of covariates. In these cases the assumption of local independence as well as involving a scarce fit, would be also an unrealistic hypotheses.

The proposed non-traditional solution to deal with this local dependence (Hagenaars 1988, Magidson and Vermunt 2004), is to include one or more direct effects in the model. Thus, both associations between two indicators and direct effects of selected covariates on selected indicators can be added, according to different situations.

In order to analyse formally the introduction of direct effects, let us consider again model [4.13], where we included K indicators and one covariate Z_i. Suppose we want to relax two local independence assumptions by assuming that the two indicators Y_{i1} and Y_{i2} are directly related, and that a third indicator Y_{i3} is affected by the covariate. These assumptions modify the conditional densities formulation as follows:

$$f\left(\mathbf{Y}_i \middle| Z_i\right) = \\ = \sum_{t=1}^{T}\left(X_i = t \middle| Z_i\right) f\left(Y_{i1}, Y_{i2} \middle| X_i = t\right) f\left(Y_{i3}, \middle| X_i = t, Z_i\right) \prod_{k=4}^{K} f\left(Y_{ik} \middle| X_i = t\right) \qquad [4.16]$$

Two changes as regard the standard formulation of the LC models given in equation [4.13] have to be noted in [4.16]. Firstly, the dependent variables Y_{i1} and Y_{i2} now appear as joint dependent variable, in order to allow the existence of a residual association between them. Secondly, in [4.16] the covariate Z_i appears also in the conditional density of Y_{i3}, implying that the manifest indicators Y_{ik} are mutually independent conditional on both latent class membership and the exogenous variable. Now the conditional densities are related directly to the covariate. Clearly, one may suppose different combinations of all these direct effects.

The multinomial logit model containing direct effect between two indicators h and k will be:

$$P\left(Y_{ik} = s_k, Y_{ih} = s_h \middle| X_i = t\right) = \frac{\exp\left(\eta^{kh}_{s_k s_h t}\right)}{\sum_{s''=1}^{S_h} \sum_{s'=1}^{S_k} \exp\left(\eta^{kh}_{s' s'' t}\right)} \qquad [4.17]$$

with linear term:

$$\eta^{kh}_{s_k s_h t} = \beta_{0s_k} + \beta_{1s_k t} + \beta_{0s_h} + \beta_{1s_h t} + \beta_{2s_k s_h} \qquad [4.18]$$

The relaxing of local independence assumption is expresses here by the inclusion of a higher-order term involving combination between the two indicators for which one assumes the existence of a residual association. In equation [4.18] β_{0s_k} and β_{0s_h} still represent the indicators intercepts, and $\beta_{1s_k t}$ and $\beta_{1s_h t}$ the class specific main effects of the latent variable on the indicator (cf. eq. [4.10]); $\beta_{2s_k s_h}$ now represents the additional parameters accounting for the effect between the two indicators h and k.

When the indicator is categorical, the introduction of a direct effect of the r-th covariate on the indicator involves the multinomal logit model

$$P\left(Y_{ik} = s_k \left| X_i = t, Z_{ri}\right.\right) = \frac{\exp\left(\eta_{s_k t Z_{ir}}\right)}{\sum_{s'=1}^{S_k} \exp\left(\eta_{s' t, Z_{ir}}\right)} \qquad [4.19]$$

where

$$\eta_{s_k t Z_i} = \beta^k_{0s_k} + \beta_{1s_k t} + \sum_{r=1}^{R} \beta_{rs_k} \cdot z_{ri} \qquad [4.20]$$

is the linear predictor.

From a technical point of view, the relaxing of local independence assumption is implemented introducing into the [4.10] extra effect parameters β_{rs_k} and $\beta_{s_k s_h}$ that represent the desired direct effects. From an interpretative point of view, these extra parameters account for the residual correlation between pairs of variables that is not explainable by the latent classes. Moreover, these direct effects may imply that the indicators have a different meaning for different subgroups.

4.3.3 Imposing parameter restrictions

As any statistical method, LCA may be used both as an exploratory and as a confirmatory tool. In the former case, the Latent Class models are said to be unrestricted, the principal aim being to identify a set of levels for the latent variable on the basis of a set of observed measures. Since this approach is typically used as a first step in a modelling procedure in absence of an explicit theory (or to provide information concerning the adequacy of an existing theory), it does not impose a priori constraints on model parameters. Conversely, restrictions are commonly imposed in confirmatory Latent Class Analysis, that is, when the objective is to test certain hypothesis regarding the nature of the latent variable. In order to verify whether hypothesized characteristics of the latent variable actually correspond to those found empirically, a priori restrictions on either the conditional probabilities, the latent class probabilities, or both, can be imposed. Several variants of the basic Latent Class model may be obtained in order to test different types of hypotheses and achieve more parsimonious models. Our objective here is merely to give an overview of these possibilities.

In literature one usually distinguishes three types of constraints: equality, inequality and specific value ones (McCutcheon 1987). Equality restrictions refer to the fact that one imposes the same value to two or more latent class parameters. Interpretatively, equality constraints applied on latent class probabilities test the hypothesis that the size of the latent classes is equal, i.e. $P\left(X_i = 1\right) = P\left(X_i = 2\right)$. For example, this kind of constraint appears particularly useful when comparing the latent classes of two or more populations. Instead, equality in conditional densities hypothesizes that observations in two or more classes are equally likely to be found at a given level of an observed variable. For example, setting $P\left(Y_{i1} = 1 \middle| X_i = 1\right) = P\left(Y_{i2} = 1 \middle| X_i = 1\right)$ and $P\left(Y_{i1} = 2 \middle| X_i = 2\right) = P\left(Y_{i2} = 2 \middle| X_i = 2\right)$, one equates a priori conditional probabilities for two marginal dichotomous indicators. Such manifest indicators are called *parallel* indicators of the latent variable (Hagenaars 2003). In this way, for two (or more) classes, the conditional probabilities are identical for two (or more) indicators.

Inequality restrictions are relevant in problem concerning classes ordering. Sometimes, a reasonable hypothesis is that the classes of the latent variable represent a scale on which the respondents can be ordered. This problem may be faced by imposing for each dichotomous item[22] inequality restriction on the response probabilities that yields to the so-called ordinal LC analysis: $P\left(Y_{i1} = 1 \middle| X_i = 1\right) \geq \ldots \geq P\left(Y_{i1} = 1 \middle| X_i = t\right) \geq \ldots \geq P\left(Y_{i1} = 1 \middle| X_i = T\right)$.

The third type of restrictions, also called deterministic model restrictions, applies when one or more of the parameters are fixed to equal an a priori specified value. For example, they apply when one imposes latent class probabilities being equal to a specific value required by the theory, or test the hypothesis that a given class accounts for a specific proportion of the population. Nevertheless, it is worthwhile noting that, in practice, deterministic restrictions of latent classes probabilities are rarely used in social research. Conversely, deterministic constraints may be imposed to conditional response probabilities, in order to define an indicator as a *perfect* indicator (Hagenaars 1993) of the latent variable. In this instance, setting the appropriate conditional response probabilities equal to one or zero – for example $P\left(Y_{i1} = 1 \middle| X_i = 1\right) = 1$, and consequently $P\left(Y_{i1} = 2 \middle| X_i = 1\right) = 0$ – implies that class 1 respondents will have probability one to give a certain response to the first item (and a zero probability to give the other response).

[22] For polytomous items with more than two response categories the system of inequalities can be generalized. We refer to Croon (2002) for further details.

Through the imposition of restrictions, one reduces the number of parameters of a given model. Thus, it may represent an alternative strategy when the augmentation of the number of classes poses identifiability problems.

4.4 LC analysis in a multilevel framework

The fundamental assumption of standard LC models is that observations are independent. However, this assumption is often violated. Applied social science research often involves the analysis of populations that are hierarchically structured: individuals nested in countries, pupils in schools, workers in organizations, and so on. Sharing the same group-specific influence, observations within a group tend to be more alike than observations coming from different groups[23], that entails dependence among observations within the same group. Ignoring this *intra-group* correlation and treating within-group observations as the same as between-group ones, may produce invalid standard errors (Agresti 2002), particularly when the clustering of units is considered a phenomenon of interest rather than a mere disturbance.

The treatment of latent structures among groups can be done in different ways. The first development in this sense is due to Clogg and Goodman (1984, 1985) who implemented the so-called *simultaneous latent structure analysis* for several groups, or *multiple-group analysis*. Simultaneous latent structure analysis enables researchers to compare latent structures in different populations, in terms of the number of the resulting latent classes and of their conditional densities (or probabilities) across groups. For example, two populations could have the same number of classes but different structure across groups, or, again, the same number of classes and similar latent structures but different distribution of the classes.

Random coefficients models (Agresti *et al*. 2000), also called multilevel or hierarchical models (Skrondal and Rabe-Hesketh 2004, Snijders and Bosker 1999), are a more recent approach dealing with various types of dependent observations. The aim of multilevel analysis is to deal with dependence among observations, and to disentangle group-level effects from individual-level ones.

Recent latent variables literature has developed different multilevel statistical models allowing researchers to evaluate the effects of the shared group on the individual outcome. Table 4.2 depicts a four-fold classification of multilevel Latent Variable models. Models A and B are extensions to the multilevel framework of continuous Latent Variable modelling (e.g. factor and IRT models). For example, the multilevel mixture factor analysis is a hybrid Latent Variable modelling that combines elements of multilevel factor models and latent class ones, including both continuous and categorical latent variables in order to model heterogene-

[23] Note that the "groups" can be conceived as different populations (so, for the examples in the text: countries, schools, organizations) for which identical measures are obtained from their members, but also as independently drawn samples of the same population at different times (McCutcheon 1986). This latter case applies, for example, for repeated measures or item responses that are nested within individuals.

ity in the observed items. In general (Nylund *et al.* 2007), the latent class variable is used to identify distinct groups in the population, and the continuous latent variable (i.e. the factor) is used at the lower-level to describe a continuum existing within the classes (case B). Instead, cases C and D identify multilevel extensions of Latent Class models. They refer to finite mixture models to which one adds either continuous or discrete random effect at the higher level of nesting. These random effects pick up variation in LC model parameters across higher-level units. Note that finite mixture models B and D allow the researchers to handle the population heterogeneity classifying individuals into smaller and homogeneous latent subpopulation (Asparouhov and Muthén 2008).

Table 4.2 – Classification of multilevel Latent Variable models (Cf. Vermunt 2008)

	Higher level latent variable(s)	
Lower level latent variable(s)	*Continuous*	*Discrete*
Continuous	A. Multilevel FA and IRT models with continuous random effects	B. Multilevel mixture FA and IRT models
Discrete	C. Multilevel LC models with continuous random effects	D. Multilevel mixture LC models with discrete random effects

The full framework described in Table 4.2 is implemented in the Latent GOLD computer program (Vermunt and Magidson 2005b, 2008). In addition to these four cases, hybrid variants can be defined by means of this software, combining discrete and continuous latent variables at both levels of the hierarchy (Vermunt 2008).

The multilevel Latent Class model described in this section and implemented in the empirical application concerning social exclusion among European regions in Chapter 5, is an example of case D in Table 4.2. The model has been estimated using the syntax module of Latent GOLD 4.5 (Vermunt and Magidson 2005b)[24].

The basic idea in multilevel Latent Class models is to take into account the multilevel structure in order to model group differences in the distribution of the latent variable, allowing some parameters to differ across groups. Whereas the multiple-group analysis models these differences introducing fixed-effects (that is

[24] As regards multilevel Latent Class models, GLLAMM software allows to estimate LC models where latent variables at both levels are discrete, and where only the responses depend on the higher-level class membership. Also Mplus deals with LC models with continuous and discrete group-level random effects (Muthén and Muthén 1998-2007).

group dummies), a random-effects approach assumes that group-specific coefficients come from a particular distribution, whose parameters should be estimated. Latent Class models in the multilevel framework were first proposed by Vermunt (2003).

In the following paragraph 4.4.1 we describe basic features of the traditional multiple-group approach in LC analysis, then we switch to the random-effects approach (§ 4.4.2), underlining the differences and their respective advantages and disadvantages.

4.4.1 The fixed-effects approach to account for the hierarchical data structure

In order to simplify the exposition, in this Section we shall refer only to nominal observed indicators. However, notice that also in the multilevel framework continuous indicators are allowed.

In the hierarchical framework, consider Y_{ijk} indicating the response of the individual i (level-1 unit) within group j (level-2 unit) on the indicator or item k, whereas $\mathbf{Y}_{ij}$ refers to the full vector of K responses of individual i in group j. A particular level of item k is denoted by s_k and its number of categories by S_k, while $\mathbf{s}$ is the K-vector of the response pattern. X_{ij} represents the underlying latent discrete variable for the i-th subject of the j-th group, for which a particular latent class is denoted by t and the number of latent classes by T. Expanding notation given in previous Sections for the standard case, the probability of observing a particular response pattern $P\left(\mathbf{Y}_{ij}=\mathbf{s}\right)$ in presence of second-level units, can be expressed by:

$$P\left(\mathbf{Y}_{ij}=\mathbf{s}\right)=\sum_{t=1}^{T} P\left(X_{ij}=t\right)\prod_{k=1}^{K} P\left(Y_{ijk}=s_k \,\middle|\, X_{ij}=t\right) \qquad [4.21]$$

The local independence assumption is included in the formulation of class-specific probabilities [4.21]: $P\left(\mathbf{Y}_{ij}=\mathbf{s}\,\middle|\,X_{ij}=t\right)=\prod_{k=1}^{K} P\left(Y_{ijk}=s_k \,\middle|\, X_{ij}=t\right)$. Moreover, we see that the class-specific probabilities $\prod_{k=1}^{K} P\left(Y_{ijk}=s_k \,\middle|\, X_{ij}=t\right)$ are weighted with the probability $P\left(X_{ij}=t\right)$ that unit i in group j belongs to latent class t.

Marginalizing for the j-th group, we obtain:

$$P\left(\mathbf{Y}_{j}=\mathbf{s}\right)=\prod_{i=1}^{n_j}\sum_{t=1}^{T} P\left(X_{ij}=t\right)\prod_{k=1}^{K} P\left(Y_{ijk}=s_k \,\middle|\, X_{ij}=t\right) \qquad [4.22]$$

The effect due to the inclusion of a higher level of hierarchy becomes more evident writing logit equations. To take into account the multilevel structure we allow parameters in equations [4.7] and [4.9] to be dependent of the j-th group to which an individual belongs, that is:

$$P\left(X_{ij}=t\right)=\frac{\exp\left(\eta_{tj}\right)}{\sum_{s'=1}^{S_k}\exp\left(\eta_{s'j}\right)} \qquad [4.23]$$

$$P\left(Y_{ijk}=s_k \left| X_{ij}=t\right.\right)=\frac{\exp\left(\eta_{s_k t_j}\right)}{\sum_{s'=1}^{S_k}\exp\left(\eta_{s' t_j}\right)} \qquad [4.24]$$

Without other specifications, the model defined by equations [4.21]-[4.24] correspond to a multilevel LC model in which all parameters are group specific (Vermunt 2003), and it is equivalent to an unrestricted multiple-group LC model (Clogg and Goodman 1984). Essentially, multiple-group LC analysis model differences among groups by including group dummies in the model, which is equivalent to use the so-called fixed-effects approach.

As all parameters are group specific, one will obtain as many set of probabilities as the number of second-level units. So, this method leads to a number of parameters that increases rapidly with the number of level-two units. A second disadvantage is represented by the fact that this kind of models does not allow to distinguish group-level from individual-level effects: all group differences are "explained" by the group dummies. This could represent a strong limitation when the aim of the research is also the study of the effects of level-two covariates on the probability of belonging to a certain latent class.

4.4.2 A non-parametric random-effects approach in the LC framework

The mentioned disadvantages of the fixed-effects approach can be avoided adopting the random-effects approach. The novelty of random-effects approach compared to multiple-group one, lies in assuming that group-specific coefficients come from a certain distribution, whose parameters should be estimated. In this way, one overcomes the problem associated with the estimation of a set of different parameters for each group.

The basic idea of a multilevel Latent Class model (Vermunt 2003) is to allow certain model parameters to differ randomly across higher-level units, in order to deal with the dependence among lower-level observations nested within them. In the multilevel latent class framework, the possibility of choice for the mixing distribution at group level is double: differentiation of LC model parameters across groups is achieved by introducing group-level *continuous random variable*(*s*) (namely continuous random effects) or group-level *discrete random variable*(*s*). Depending on whether the form of the mixing distribution is specified or not, either a parametric or a nonparametric random-effects approach is obtained. The adoption of a random-effects approach involves the assumption that the group-specific effects (the random coefficients) come from a certain distribution, typically the Normal one in case of a continuous u_j. This specification (e.g. Vermunt 2003) yields a multilevel LC model in which the linear predictor in equation [4.23] is $\eta_{tj} = \gamma_t + \tau_t \cdot u_j$ with $u_j \sim N(0,1)$ and identifying constraint on the γ_t and the τ_t. Alternatively, one can assume discrete random effects to account for the nested structure. That is, following a non-parametric approach, one assumes a multinomial distribution for the latent variables at both levels. This model specification yields a multilevel Latent Class model in which, besides a latent variable at the individual level, there is a latent variable also at the group level. Unlike the parametric approach, in this case the second level latent variable serves to structure the second level units (i.e. groups) into a small number of latent classes.

Considering the empirical application treated in this thesis, we will concentrate the subsequent discussion on the second option[25], analysing models that assume latent classes both for individual level units and for group level ones. This approach equals to work with a discrete unspecified mixing distribution.

A non-parametric approach is more flexible. It does not rely on strong distributional assumptions about the random effects, avoiding in this way possibly misleading inferences from model assumptions. Moreover, parametric assumptions (typically the Normal one) are often "*inappropriate and unverifiable*" (Aitkin 1999). The assumption of normality for random effects is popular and it has attractive features, but the impossibility of a closely check could imply misspecification and, then, possibly harmful effects.

Secondly, a non-parametric approach is useful when the random effects distribution is not itself of direct interest (Agresti 2002). Furthermore, from a substantive point of view a discrete unspecified mixing distribution at the higher level can be more appropriate and/or practical when it is more meaningful to classify groups into a small number of classes instead of placing them on a continuous scale (Vermunt 2003). Even more so, this approach turns out to be extremely helpful when the researcher has reason to believe that the latent concept which structure the second level units is not strictly ordinal, and different patterns could stand out assuming a discrete latent variable also at group level. Strong assumptions yield to more parsimonious solutions and sometimes could be preferred, anyway it would bear in mind that conclusions are sensitive to the chosen distributional form. Moreover, Vermunt and Van Dijk (2001) demonstrate as a discrete unspecified mixing distribution, in addition to fit much better compared with the assumption of multivariate normality of the random coefficients, involves much shorter computation time.

The extension of a standard LC model in the multilevel framework can be implemented as follows (Vermunt and Magidson 2005b): beginning from a stan-

[25] For an extended discussion about the parametric approach, we refer to Vermunt (2003). Note that it is even possible to combine the two approaches (Vermunt 2008).

dard finite mixture model at individual level, one adds another finite mixture model, at the higher level. So, the multilevel data structure and the homogeneity within group is dealt with the random-effects introduced by means of a finite mixture model which represents itself a nonparametric random-effects model (Aitkin, 1999). This yields a model with a separate finite mixture distribution at each level of nesting.

Assuming observed responses nested within individuals, who are in turn nested within groups, it follows that we have to define two finite mixture models. The probability structure at the individual level is similar to the standard LC model of equation [4.3], except that now we have to take into account that individuals belong to different groups for which we assume another latent structure. Thus, we denote W_j a discrete latent variable at group level with $m = 1,...,M$ classes or mixture components unknown a priori. Notice that one knows to what group individuals belong to, but the membership of the groups to the second level latent classes is unknown a priori, as well as it is unknown the membership of individuals to the first level latent classes. Now, instead of $P(\mathbf{Y}_i)$ the first part of the model refers to the conditional probabilities $P\left(\mathbf{Y}_{ij} = \mathbf{s} \middle| W_j = m\right)$:

$$\begin{aligned} P\left(\mathbf{Y}_{ij} = \mathbf{s} \middle| W_j = m\right) &= \sum_{t=1}^{T} P\left(X_{ij} = t \middle| W_j = m\right) P\left(\mathbf{Y}_{ij} = \mathbf{s} \middle| X_{ij} = t\right) \\ &= \sum_{t=1}^{T} P\left(X_{ij} = t \middle| W_j = m\right) \prod_{k=1}^{K} P\left(Y_{ijk} = s_k \middle| X_{ij} = t\right) \end{aligned} \qquad [4.25]$$

that defines the conditional probabilities of observing a certain response pattern for each individual given the class membership of the higher-level unit.

Secondly, one adds an element that connects the first-level units belonging to the same group, so accounting for their dependence. The probability for the full vector of responses of all individuals in group j – $P(\mathbf{Y}_j)$ – is obtained with the additional assumption that its n_j members' responses are independent of one another, conditional on group class membership, subsequently summing over the latent classes at group level:

$$P\left(\mathbf{Y}_j = \mathbf{s}\right) = \sum_{m=1}^{M} P\left(W_j = m\right) \prod_{i=1}^{n_j} P\left(\mathbf{Y}_{ij} = \mathbf{s} \middle| W_j = m\right) \qquad [4.26]$$

The first part of equation [4.26] contains the assumption that each group belongs to one and only one of the M classes of the higher-level discrete latent variable with probabilities equal to $P\left(W_j = m\right)$. The second part assumes that the n_j observations within a group are mutually independent given the class membership of the group.

Hence, substituting equation [4.25] in equation [4.26], we obtain the hierarchical or multilevel Latent Class model:

$$P\left(\mathbf{Y}_j = \mathbf{s}\right) = \\ = \sum_{m=1}^{M}\left[P\left(W_j = m\right)\left[\prod_{i=1}^{n_j}\sum_{t=1}^{T}P\left(X_{ij} = t \middle| W_j = m\right)\prod_{k=1}^{K}P\left(Y_{ijk} = s_k \middle| X_{ij} = t\right)\right]\right] \qquad [4.27]$$

Comparing equations [4.22] and [4.27] makes evident that, instead of estimating a separate latent class distribution for each group (like in multiple-group analysis), following this non-parametric random-effects approach, one assumes that each j-th group belongs to one of the latent classes. That is, we are defining a discrete latent variable both for lower-level (X_{ij}) and higher-level units (W_j). This situation can be useful depicted in a path diagram (Figure 4.4).

Figure 4.4 shows clearly that we are dealing with a three-levels structure. In most cases, indeed, LC analysis is basically a technique for analyzing two-level data structures, for example repeated measures or item responses nested within individuals (e.g. De Boeck and Wilson 2004). Thus, the introduction of a higher-level in which individuals are nested, represents a three-level extension of the LC model (Vermunt 2008). At the top we find the discrete latent variable for the groups, W_j; this one influences the membership to the latent classes of the latent variables for the level-1 units $X_{n_j j}$. On the bottom there are the observed responses Y_{ijk}. Without any other specification, the multilevel LC models assume that the discrete latent variables $X_{n_j j}$ are mutually independent given W_j, and the individual responses Y_{ijk} are mutually independent given X_{ij}.

Figure 4.4 – Path diagram of a multilevel Latent Class model

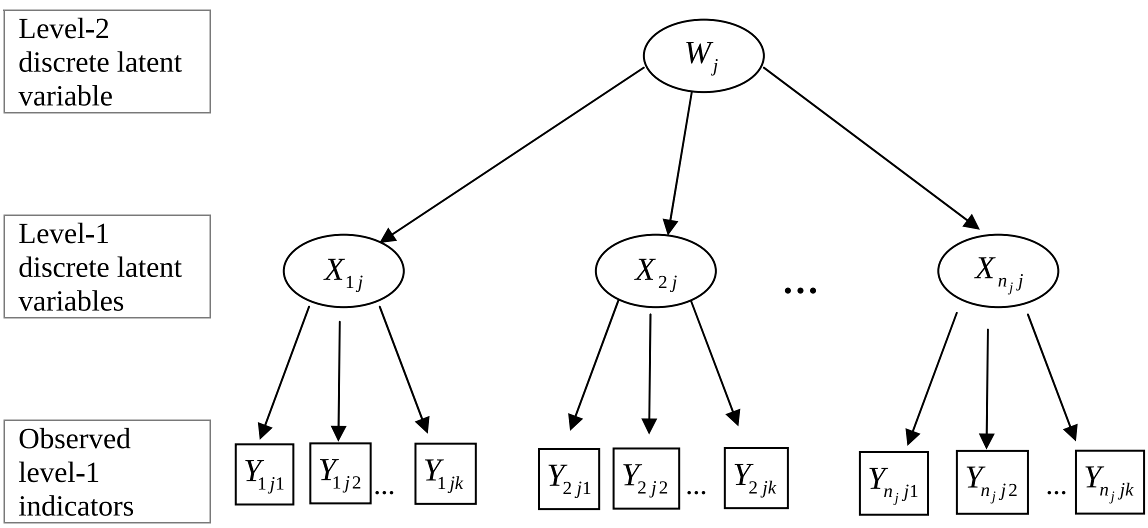


Figure 4.4 depicts a multilevel LC model in which the latent classes at the lower-level capture all the association between the responses within lower-level units Y_{ijk} , whereas the higher-level classes capture the association between subjects within groups. The objective of this model specification is to achieve a meaningful identification of the lower-level classes taking into account the multilevel data structure.

By means of loglinear models, the membership probability of the j -th group to the m -th latent class at group level is:

$$P\left(W_j = m\right) = \frac{\exp\left(\eta_m\right)}{\sum_{m'=1}^{M} \exp\left(\eta_{m'}\right)} \qquad [4.28]$$

where the linear predictor

$$\eta_m = \alpha_{0m} \qquad [4.29]$$

is subjected to the identifying constraint $\sum_{m=1}^{M} \alpha_{0m} = 0$ in case of effect coding, or $\alpha_{01} = 0$ or $\alpha_{0M} = 0$ in case of dummy coding.

We can parameterize the multinomial probability that individual i belongs to a particular latent class t at lower level, given the higher level latent class membership, as follows:

$$P\left(X_{ij} = t \middle| W_j = m\right) = \frac{\exp\left(\eta_{tm}\right)}{\sum_{t'=1}^{T} \exp\left(\eta_{t'm}\right)} \qquad [4.30]$$

with linear predictor

$$\eta_{tm} = \gamma_{0tm} \qquad [4.31]$$

This component captures the key differences between the classes of the discrete latent variable at the group level, W_j .

The conceivable extensions of the multilevel model are numerous, like in standard LC models. Thereby, they can involve the relaxing of local independence assumption for some pairs of indicators, or the introduction of covariates, now both at individual and group level. In this case, the group level covariates will be useful to profile the group level latent classes. In addition, in this framework researcher can assume that covariates affect Y_{ijk} , X_{ij} or W_j , or that level-1 covariates depend on the mixture variable W_j , for example structuring a model for the latent classes in which the intercept and the covariate effects differ across classes at the higher-level.

Equation [4.27] describes a model in which the model part linking the lower level class membership to the responses is the same for all groups. However, in hierarchical Latent Class models, one can assume that conditional response probabilities depend on the group-level latent variable. Assuming group-level classes to have direct effects on the indicators equals to account for the fact that individuals belonging to different groups respond to certain items in a different manner. As suggested by Vermunt (2003), this is a way to deal with a phenomenon sometimes referred to as item bias. Mathematically, this corresponds to change the structure of the conditional probabilities, allowing the regression coefficients to differ across clusters of higher-level units:

$$P\left(\mathbf{Y}_{ij} = \mathbf{s} \middle| W_j = m\right) = \sum_{t=1}^{T} P\left(X_{ij} = t \middle| W_j = m\right) \prod_{k=1}^{K} P\left(Y_{ijk} = s_k \middle| X_{ij} = t, W_j = m\right) \quad [4.32]$$

In such a case,

$$P\left(Y_{ijk} = s_k \middle| X_{ij} = t, W_j = m\right) = \frac{\exp\left(\eta_{s_k tm}\right)}{\sum_{s'=1}^{S_k} \exp\left(\eta_{s'tm}\right)} \quad [4.33]$$

where

$$\eta_{s_k tm} = \beta_{0 s_k} + \beta_{1 s_k t} + \beta_{2 s_k m} \quad [4.34]$$

β_{0s_k} represents the intercept, and β_{1s_kt} and β_{2s_km} are the main effects of the latent variables at both levels, X_{ij} and W_j respectively. In equation [4.32] both the lower-level class proportion and the class-specific probabilities depend on W_j.

Another special case assumes instead that response probabilities depend on the group-level class membership, but lower-level class membership does not: $P\left(X_{ij} = t \middle| W_j = m\right) = P\left(X_{ij} = t\right)$. This model is similar to the multilevel factor analysis model proposed by Muthén (1994) in which the variation in a multivariate response vector is attributed to common latent factors at two levels of a hierarchical structure (Vermunt in press). It is very similar also to a standard three-level regression model that involves the variance decomposition into independent parts: the higher-level classes capture the common variation of all responses within a group, while the lower-level classes the common residual variation within subjects (Vermunt 2008).

Some of the extensions listed above, which involve hypothesis typical in social sciences research, will be deepened in more detail in the sections describing the empirical application (Chapter 5).

4.5 Model fitting

Due to its importance, in next paragraph the main characteristics of the traditional EM algorithm will be reviewed briefly (§ 4.5.1); next, we present the modified EM procedure implemented by Vermunt (2004) for the parameter estimation in multilevel LC models (§ 4.5.2). Finally, aspects relied to model identifiability, model evaluation and selection are tackled (§ 4.5.3 and 4.5.4).

4.5.1 ML estimation and the EM algorithm

The most important contribution to estimation procedure of LC models has been implemented in the Seventies by Goodman (1974a, 1974b, 1979). He developed an algorithm for obtaining maximum likelihood estimates (MLEs) of the conditional and latent class probabilities. Although others (see Lazarsfeld and Henry 1968, and the references therein) had previously suggested efficient methods for estimating LC parameters, Goodman's estimators are considered a "*breakthrough beyond earlier approaches*" (McCutcheon 1987) for their simplicity and generality. Estimation of LC models is somewhat difficult due to the presence of the latent variable, which involves an incomplete data matrix. This results in an expression for the log-likelihood with a complex structure. The mostly frequently numerical methods used in literature for obtaining maximum likelihood estimates are the Newton-Raphson algorithm (Haberman 1988), the scoring method (Haberman 1979), and the EM algorithm. Since the first procedure is not very stable numerically, and the second one presents the drawback to difficultly find initial values for the parameters that will lead to convergence of the algorithm (Heinen 1996), as a result, they are not widely used for parameter estimation in LC models. The preferred alternative is nowadays represented by the EM method. Applying the iterative procedure of Expectation-Maximization (EM) algorithm (Dempster *et al.* 1977) to the initial estimates proposed by Goodman leads to ML estimates for LC models.

The EM algorithm is an efficient iterative procedure to compute maximum-likelihood estimate in the presence of incomplete or hidden data, or in problems that can be posed in a similar form, such as mixture model estimation. Sometimes hidden variables may be introduced purely as an artifice for making the maximum likelihood estimation of unknown parameters tractable. Since the latent variable is not directly observed, it can be seen as a missing one. By treating the latent variable X_i as missing (or, equally, unobserved), parameter maximum likelihood estimation of LC models can be made by means of the EM algorithm.

For a standard LC model with covariates, the log-likelihood to be maximized is:

$$\log L = \sum_{i=1}^{I} \log f\left(\boldsymbol{Y}_i \middle| \boldsymbol{Z}_i, \boldsymbol{\vartheta}\right) = \sum_{i=1}^{I} \sum_{t=1}^{T} \log P\left(X_i = t \middle| \boldsymbol{Z}_i, \boldsymbol{\vartheta}\right) f\left(\boldsymbol{Y}_i \middle| X_i, \boldsymbol{Z}_i, \boldsymbol{\vartheta}\right) \qquad [4.35]$$

where for a generic individual i (and a total number of cases equals to I), $\boldsymbol{Y}_i$ denotes the k vector of the response variables, $\boldsymbol{Z}_i$ the p vector of the covariates, and $\boldsymbol{\vartheta}$ the vector containing the unknown γ and β parameters.

The iterative procedure of Expectation-Maximization algorithm enables to make ML estimation in presence of missing data, bringing back to a problem of complete data estimation. Briefly, starting from observed data and current estimates of the model parameters, each iteration of the EM algorithm consists of two steps, the E-step and the M-step:

1. the expectation or E-step provides estimation of missing data, using the conditional expectation. In the context of Latent Class models, this means that the probabilities for the complete data matrix are estimated using the observed proportions (Heinen 1996);
2. in the maximization or M-step the estimates of the missing data obtained in E-step are used in place of the actual missing data, in order to maximize the likelihood function under the assumption that the missing data are known. Then, these new model estimates are used to perform another E-step, in order to obtain new estimates for the complete data matrix, and so on.

The two steps are performed alternately until the procedure converges, convergence being assured since the algorithm increase the likelihood at each iteration.

4.5.2 A modified EM algorithm for the hierarchical framework

The standard EM algorithm briefly described above, cannot be used for a nonparametric hierarchical model with more than one level of nesting. Alternative solutions to provide ML estimates with EM algorithm also for mixed-model having more than two hierarchical levels, have been implemented recently. In this paragraph we describe a variant of the EM algorithm for hierarchical latent structures proposed by Vermunt (2004), which requires a special implementation of the E step. A hierarchical LC model involves a large number of unobserved variables, i.e. $1+n_j$, so the computation of the joint posterior distribution $P\left(W_j, X_{1j}, X_{2j}, ..., X_{n_j j} \middle| \boldsymbol{Y}_j\right)$ of the discrete variable with missing values, with a total of $L \cdot T^{n_j}$ entries, prevents from the application of the E step of the standard EM algorithm. This problem can be overcome exploiting conditional independence assumption implied by the hierarchical LC models. In the following we describe the procedure *upward-downward* proposed by Vermunt and implemented in Latent GOLD (Vermunt and Magidson 2005b, 2005c). This modified EM algorithm will be applied for the parameters estimation of the empirical application described in Chapter 5.

Starting from current estimates for the conditional and latent class probabilities of the model, the E-step computes the expected value of the complete data log-likelihood, providing the marginal posterior probabilities $P\left(W_j = m, X_{ij} = t \middle| \mathbf{Y}_j\right)$. In the multilevel Latent Class model (see equation [4.27][26]) the expected value of the complete data log-likelihood has the form:

$$\begin{aligned}\log L_j = P\left(\mathbf{Y}_j = \mathbf{s}\right) & \\ = \sum_{m=1}^{M} P\left(W_j = m \middle| \mathbf{Y}_j\right) \log P\left(W_j = m\right) & \\ + \sum_{m=1}^{M} \sum_{1=1}^{n_j} \sum_{x=1}^{T} P\left(W_j = m, X_{ij} = t \middle| \mathbf{Y}_j\right) \log P\left(X_{ij} = t \middle| W_j = m\right) & \\ + \sum_{m=1}^{M} \sum_{1=1}^{n_j} \sum_{x=1}^{T} P\left(X_{ij} = t \middle| \mathbf{Y}_j\right) \log P\left(\mathbf{Y}_{ij} \middle| X_{ij} = t\right) &\end{aligned} \qquad [4.36]$$

As class membership of the i-th individual (X_{ij}) is independent of the other observations in the same group given the higher-level class membership of the group (W_j), the marginal posterior probabilities needed to perform the E step can be written using the following decomposition:

$$\begin{aligned}P\left(W_j = m, X_{ij} = t \middle| \mathbf{Y}_j\right) & \\ = P\left(W_j = m \middle| \mathbf{Y}_j\right) P\left(X_{ij} = t \middle| W_j = m, \mathbf{Y}_j\right) & \\ = P\left(W_j = m \middle| \mathbf{Y}_j\right) P\left(X_{ij} = t \middle| W_j = m, \mathbf{Y}_{ij}\right) &\end{aligned} \qquad [4.37]$$

where

$$P\left(W_j = m \middle| \mathbf{Y}_j\right) = \frac{P\left(W_j = m\right) \prod_{i=1}^{n_j} P\left(\mathbf{Y}_{ij} \middle| W_j = m\right)}{P\left(\mathbf{Y}_j\right)} \qquad [4.38]$$

and

[26] In order to simplify the notation and the explanation, we omit in this explanation the vector of covariates $\mathbf{Z}_j$ from the model.

$$
\begin{aligned}
P\left(X_{ij}=t\middle|W_j=m,\mathbf{Y}_{ij}\right)&=\frac{P\left(X_{ij}=t,\mathbf{Y}_{ij}\middle|W_j=m\right)}{P\left(\mathbf{Y}_{ij}\middle|W_j=m\right)}\\
&=\frac{P\left(X_{ij}=t\middle|W_j=m\right)\prod_{k=1}^{K}P\left(\mathbf{Y}_{ij}\middle|X_{ij}=t\right)}{P\left(\mathbf{Y}_{ij}\middle|W_j=m\right)}
\end{aligned}
\qquad [4.39]
$$

The independence assumption can also be seen from the absence of direct relations between the k indicators and between the n_j level-one latent variables in Figure 4.4: level-one observations are mutually independent given the value of W_j, and subsequently X_{ij} is independent of the information of the other level-two group members.

The modified procedure uses, in the E-step, the so-called *upward-downward* algorithm (Vermunt 2003) to compute the posterior probabilities given current estimates for the unknown model parameters[27]. In the *upward* part of the algorithm, the n_j sets of probabilities of equation [4.39] $P\left(X_{ij}=t,\mathbf{Y}_{ij}\middle|W_j=m\right)=P\left(X_{ij}=t\middle|W_j=m\right)P\left(\mathbf{Y}_{ij}\middle|X_{ij}=t\right)$ for each individual are computed, then they are collapsed over X_{ij} to obtain $P\left(\mathbf{Y}_{ij}\middle|W_j=m\right)$; secondly, this term is used in turn to obtain $P\left(W_j=m\middle|\mathbf{Y}_j\right)$ for each j-th group (equation [4.38]). To conclude, the *downward* step uses $P\left(W_j=m\middle|\mathbf{Y}_j\right)$ and $P\left(X_{ij}=t\middle|W_j=m,\mathbf{Y}_{ij}\right)$ to compute – through equation [4.37] – the bivariate joint posterior probability $P\left(W_j=m,X_{ij}=t\middle|\mathbf{Y}_{ij}\right)$ for each individual.

Substantially, this procedure moves *upward* and *downward* through the hierarchical structure: latent variables are summed out going from the lower to the higher level, and subsequently, the marginal posterior probabilities are computed going from the higher to the lower level (Vermunt and Magidson 2005b). This procedure makes the estimation feasible, because with this algorithm the computer storage and time to obtain estimation increases only linearly with the number of cases per group instead of exponentially (as would happen using standard EM algorithm).

Then, as usual, in M step of EM algorithm, standard complete data methods can be used to maximize the expected complete data log-likelihood with respect the parameters update.

[27] The name is in analogy with the procedure forward-backward of the Baum-Welch EM algorithm, which moves forward and backward through the hidden Markov chain (Baum *et al.* 1970).

Note that, as for other statistical models, ML estimates can be computed with either the EM algorithm or the Newton-Raphson one. For a Latent Class model a well-known problem (e.g. Goodman 1974a, Hagenaars and McCutcheon 2002) is the possibility to have a local maximum for the log likelihood equations, rather than the global one, that is, there may exist more than one set of conditional and latent class probabilities for any specified number of T latent classes[28]. Therefore it is advisable to perform the fitting process several times starting from different guesses for the parameter values. As the EM algorithm tends to be less sensitive to the choice of starting values and it is more stable even when it is far away from the optimum, a common practice is to begin with the EM algorithm and then switch to the Newton-Raphson algorithm to exploit his speed. In the software Latent GOLD local optima are avoided by using multiple set of random starting values. Moreover, Latent GOLD uses a combination of EM and Newton-Raphson algorithm to find the Maximum Likelihood estimates for the model parameters. Precisely, the estimation process starts with EM iterations and, when either the maximum number of EM iterations or the relative change in the parameters is small, the program switches to the Newton-Raphson algorithm. Finally, it stops when the maximum number of NR iterations or the overall convergence criterion is reached[29]. The algorithm used in Latent GOLD to maximize the log-likelihood, as well as the nature and number of the indicators, affect the computational time necessary to estimate a multilevel latent variable model, which increase strongly when covariates are used.

4.5.3 Model identifiability

Another problem that can arise in fitting Latent Class models is the under-identification, that is the number of parameters to be estimated is too large as regards the available data. That is, parameters cannot be determined uniquely. A necessary (but not sufficient) condition for a model to be identifiable is that the number of degrees of freedom is not negative, being the number of estimable parameters limited by the availability of degrees of freedom from the observed variables. This equals to say that the number of independent unknown parameters to estimate has not to exceed the number of independent known ones (the cell frequencies in the observed table).

Goodman (1974a) has formulated a sufficient condition for local identifiability (Heinen 1996). Models that are not identified can be made identifiable by putting restrictions on the parameters.

[28] Particularly, the risk of multiple local maxima represents a serious problem as the number of T classes increases, or the sample size decreases (Bartholomew and Knott 1999).

[29] Usually, 250 EM iteration and 50 NR ones are enough to obtain satisfactory results. In the empirical application presented in Chapter 5, we set 350 EM iteration and 100 NR, the convergence reached after a small number of iterations.

4.5.4 Model evaluation and model selection

Once the model parameter estimates are obtained, the question is whether the model fits the data in an acceptable way. This means that one has to verify if the theoretical model succeeds in explaining the observed data. Model evaluation encompasses both the assessment about the global fit of a certain model and the comparison of that model with others. Above all, one could assess overall goodness-of-fit by carrying out an unconditional statistical test. Moreover, another issue is the comparison of nested models, that is, in LCA, models with a different number of classes, for example t and $t+1$. In order to choose between different models, different techniques are available. Here some tests to evaluate the global goodness-of-fit, some information criteria, and other statistical-substantive criteria are briefly introduced highlighting some critical aspects.

The standard goodness-of-fit tests – that compare the observed frequencies f_{s} and estimates of the expected frequencies under the model being tested $\hat{F}_{\mathrm{s}}$, across the response patterns – apply also in LCA. Two types of goodness-of-fit tests have been commonly employed: the likelihood ratio chi-squared statistic L^2 and the Pearson chi-squared goodness-of-fit test statistic X^2. For a given data pattern $\mathbf{s}$[30], they are defined as:

$$L^2 = 2\sum_{\mathbf{s}} f_{\mathrm{s}} \ln\left(\frac{f_{\mathbf{s}}}{\hat{F}_{\mathbf{s}}}\right) \qquad [4.40]$$

$$X^2 = \sum_{\mathbf{s}} \frac{\left(f_{\mathrm{s}} - \hat{F}_{\mathrm{s}}\right)^2}{\hat{F}_{\mathbf{s}}} \qquad [4.41]$$

Both the statistics are a measure of the association between the variables that remains unexplained by the model. The Chi-squared statistic evaluates whether the ML estimate for the expected frequencies $\hat{F}_{\mathbf{s}}$ based on a certain model, differ from the corresponding observed frequencies $f_{\mathbf{s}}$: the larger the value, the poorer the model fits the data. When the null hypothesis holds, both statistics are distributed approximately as a χ^2, with a number of degrees of freedom equal to the number of cells in the full multi-way table minus 1 minus the number of distinct parameters. Thus, a model fits well the data if the value of L^2 or X^2 is sufficiently low to be attributable to chance. X^2 is demonstrated to be distributed less like a chi-squared than is L^2 (see Collins *et. al.* 1993 for a more detailed discussion).

[30] In models containing covariates, one also has to group cases with identical covariates. Moreover, it is worthwhile noting that the program Latent GOLD uses a somewhat unconventional formula for X^2, summing over the nonzero observed cells only (Vermunt and Magidson 2005b).

L^2 and X^2 present two main disadvantages that could reduce their usefulness in most empirical applications. First, the χ^2 statistic is especially sensitive to sample size, and its performance worsens as sample size increases. Particularly, when sample size is large the chi-square statistic tends to be conservative, leading to not refuse the null hypothesis also for modest parameters. Secondly, problems may arise in case of sparse data, that is when a large number of observed variables or a large number of categories for these variables, leads to a lot of empty cells[31]. Under these conditions, the chi-squared distribution should not be used to compute the *p*-value because the chosen goodness-of-fit statistics may deviate considerably from known distributions (Read and Cressie 1988). The problem of sparseness of data is particularly present in Latent Class Analysis.

Finally, the statistics described above, even if used to assess global fit, cannot be used in model comparison since the difference in the likelihoods for two nested models, i.e. models with different number of classes, is not chi-square distributed (McLachlan and Peel 2000). In the case of LC models with differing numbers of classes, regularity conditions are not met (Nylund *et al.* 2007), and standard difference testing is not applicable.

Different alternatives have been proposed in literature. McLachlan and Peel (2000) initially proposed the bootstrap technique in the contest of mixture of Normals. In brief, bootstrap approach empirically estimates the *p*-value associated with the L^2 statistic by means of a parametric bootstrap, instead of assuming it to follow a known distribution. The model of interest is estimated not only for the sample under investigation, but also for some *n* replication samples generated from the probability distribution defined by the maximum likelihood estimates. The procedure applies also to the *p*-value corresponding to the log-likelihood difference distribution used to test two nested models.

Another approach that allows tackling both problems at once, is represented by the information criteria. Forasmuch as any model is a simplification of reality (Agresti 2002), a simple model may be preferred to a more complex one because it tends to provide better estimates of certain characteristics of the true model and it has the advantage of model parsimony. Following this rationale, information criteria are based on the computation of some indexes representing a penalized form of the likelihood. As the likelihood increases with the addition of some parameters, it is penalized by the subtraction of a term accounting for the complexity of the model: in this way information criteria penalize models with many parameters. The most used model selection criteria are BIC (Bayesian Information Criterion), also known as the Schwarz's criterion, AIC (Akaike Information Criterion), and CAIC (Consistent Akaike Information Criterion):

$$BIC_{L^2} = L^2 - \log(N)\,df \qquad [4.42]$$

[31] As suggested by Vermunt and Magidson (2005b), the best indication of sparseness of data is when the number of degree of freedom is (much) larger than the total sample size N.

$$AIC_{L^2} = L^2 - 2df \qquad [4.43]$$

$$CAIC_{L^2} = L^2 - \left[\log(N) + 1\right] df \qquad [4.44]$$

where N is the number of observations, and df are the degrees of freedom. Taking into account both model fit and parsimony, the information criteria are particularly useful for models comparisons: the lower they are, the better the model.

Otherwise, using the log-likelihood $LL = \sum_{i=1}^{I} \hat{f}\left(\mathbf{Y}_i\right)$, more general formulations of the information criteria are:

$$BIC_{LL} = -2LL - \log(N)\, np \qquad [4.45]$$

$$AIC_{LL} = -2LL + 2np \qquad [4.46]$$

$$CAIC_{LL} = -2LL - \left[\log(N) + 1\right] np \qquad [4.47]$$

that use np (number of free parameters to be estimated) instead of df. Following Vermunt and Magidson (2005b) the information criteria based on L^2 or LL should yield the same result, however the latter should be preferred for extremely large df.

A debate has come up in recent years referring to the term N in the computation of the different information criteria presented above in the case of multilevel analysis. While in standard LCA one refers to N as the number of observations, or equivalently, the sample size, getting on to the multilevel framework it is not clear which sample size should be used in the procedure. For example, in a two level Latent Class model, the choice is between the number of groups J (number of higher-level observations), and the number of individuals N (number of lower-level observations). There are no clear-cut hints in favour of one of the options: either the number of groups or the number of individuals could be used, sometimes depending on whether the principal aim is to determine the number of classes at the higher or at the lower level. Nevertheless, it should be noted that in latent class modelling, the use of J represents the predominant approach (Skrondal and Rabe-Hesketh 2004). A simulation study of Lukociene and Vermunt (in press) has recently provided further evidence in this sense. In the Latent GOLD software, the information criteria values are computed using the number of cases as N.

Sometimes, a lack of fit may be due to a certain amount of association not explained by the current model. The identification of any pair of variables responsible for some local dependencies can be traced by means of bivariate residuals. The bivariate residuals (BVR) represent a diagnostic statistic corresponding to a Pearson chi-square statistic divided by the degrees of freedom (Vermunt and Magidson 2005b); they are computed on the observed frequencies in a two-way table of a pair of variables using the expected frequencies estimated under the corresponding LC model. A value of the bivariate residual substantially larger than 3.84 suggests that the association in the corresponding two-way table is not well explained by the t-class model. The ability to identify specific 2-way tables where lack of fit may be concentrated is of use in suggesting the most suitable alternative model. The presence of residual associations even after controlling for the latent variable, may be tackled either increasing the number of latent classes, or deleting one or some of the redundant indicators, or adding a direct effect between the involved indicators.

An important issue in evaluating the model fit in LCA concerns the decision about the number of classes assumed to describe the unobserved heterogeneity in the population under investigation. Note that in some applications there may be no "right" solutions about the number of latent classes. The usual approach is to begin by fitting a 1-class model, that is the baseline model that assumes independence of the data. Then, one increases the number of classes $T = 2, 3, \ldots$ up to the maximum plausible number of latent classes until an adequate fit is reached[32]. In general, as the number of classes is fewer, models fit the data worse, and a point is reached after which models are rejected by the L^2 criterion. Selection among different models may be achieved also comparing BIC (or other information indices) of each of them. An "heuristic" approach that may complement these approaches, consists in considering the percent reduction in the (log-)likelihood ratio statistic. Thus, one can compare the statistic associated with a certain LC model for which $T > 1$, with the baseline statistic associated with the 1-class model, i.e. $T = 1$, in order to determine a percent reduction measure which represents the total association explained by the model. Finally, in decision about the number of classes, one may take advantage also of the examination of the parameter estimates and of the characteristics of the resulting latent classes.

Finally, it is worthwhile to mention other measures provided by the Latent GOLD software (see Vermunt and Magidson 2005b for further details). These measures, although less formal and statistically meaningful, could help research from a substantive point of view whether used jointly with other statistical tests. The Dissimilarity Index (DI) (Clogg 1995) compares observed and estimated cell frequencies:

$$DI = \sum_{\mathbf{s}} \frac{\left| f_{\mathbf{s}} - \hat{F}_{\mathbf{s}} \right|}{2N} \qquad [4.48]$$

[32] In multilevel latent class analysis, one may estimate models for alternative values of the number of latent classes at individual and at group level, then compare them using information indices.

It is a descriptive measure that allows an indirect quantification of the proportion of the sample that should be moved to another cell to get a perfect fit.

The degree to which the latent classes are distinguishable by the data and the model can be assessed by using the estimated posterior class membership probabilities for each response pattern $\mathbf{s}$. Using this latent classification, a table can be constructed containing information on how well one can predict to which latent class cases belong, given observed values. Moreover, the proportion of classification errors can be computed. This proportion is not a fit measure, but it is an important measure to evaluate the distinctiveness of different classes[33].

Finally, in LC framework another information criterion may be added to the others previous described. The Average Weight of Evidence (AWE) criterion also considers classification performance, besides fit and parsimony:

$$AWE = -2LL^c + 2\left(\frac{3}{2} + \log N\right)np \qquad [4.49]$$

where LL^c is the so-called classification log-likelihood, which is equivalent to the complete data log-likelihood[34].

In spite of the richness of contributes pertaining to the different criteria existing to evaluate the goodness-of-fit of LC models, and to choose between different models, there is not common acceptance of which is the best one (Nylund *et al*. 2007). One of the reasons is connected to the variety of mixture models. We found a variety of simulation studies that explore the issue of goodness-of-fit in Latent Class models, and for the number of class determination (Collins *et al*. 1993, Nylund *et al*. 2007, Lukociene and Vermunt in press). They all conclude that some measures may perform better in certain situations and worse in others, depending on different factors, e.g. the sample size, the level of separation of the classes, the scale type of the response variables, or the assumptions made about the models adopted.

In our empirical application, we turn back on this issue, showing that our decision arises from a combination of criteria. All things considered, in empirical research evaluation of a model should look upon both statistical criteria and substantial and practical issues, always keeping in mind the problem under investigation.

[33] We shall discuss these issues in a more detail in the empirical application reported in Chapter 5, referring on classification on both levels.

[34] The classification log-likelihood (Vermunt and Magidson 2005b) is the log-likelihood value under the assumption that the true class membership is known, i.e.

$$LL^c = \sum_{i=1}^{I}\sum_{t=1}^{T} \log \hat{P}\left(X = t \,|\, \mathbf{Z}_i\right) \hat{f}\left(\mathbf{Y}_i \,|\, X = t, \mathbf{Z}_i\right).$$

Chapter 5
Social exclusion in European regions: an application of Multilevel LC models

5.1 The research hypothesis

5.1.1 Studying social exclusion through multilevel LC models

In social and behavioural sciences, it is common to dispose of some observed variables that are associated because of some underlying unobserved factors. Like many other concepts in social sciences, *social exclusion* represents a theoretical construct that can only be quantified or measured via indirect manifest indicators, which are assumed to be related in some way with it. In presence of a set of categorically observed measures that are highly interrelated, Latent Class Analysis offers a powerful tool for investigation and analysis, both in explorative and in confirmative way. Social exclusion is a fuzzy and not well-determined concept, linked to several manifest indicators by *probability relations and not rigid laws*, using of Lazarsfeld and Henry's words (1968). This is the reason why we consider Latent Class Analysis a fundamental approach in this area.

We stated that social exclusion is a multidimensional concept. Anyway, often it is not analysed following a multidimensional approach. We identified several indicators expressing social exclusion – the lack of social relations, the perception of the respondents, their economic situation are only few examples – anyway, they still remain imperfect measures. Even for poverty, an element that theory and research found to be strictly related to individual social exclusion, this relation is not deterministic and always true. Notwithstanding the growing interesting demonstrated in recent years towards the issues linked to social exclusion, we found a worrying lack of empirical applications on this topic. There are few applications and empirical studies that attempt to understand how the process of social exclusion trigger, and which are the most important risk factors; most of them focus on the economic aspects or on deprivation, even if in a broad significance.

In theoretical and empirical research, we found some attempts to face the problem of multidimensionality of social exclusion through the building of a composite index. In a composite index several factors and indicators are combined together, for instance in a weighted manner (Bhalla and Lapeyre 1997). Empirical studies (e.g. Barnes *et al.* 2002, Hills *et al.* 2002, Schifini and Petrucci 2007) are generally based on the construction of composite objective indicators, which are used to compare geographical areas or to develop regional rankings. However, the construction of a composite indicator that encompasses such different aspects, involves problems of weighting and interpretation (Bhalla and Lapeyre 2004) that may hide the relative importance of the different components, preventing to understand which are the domains and the priorities that need an intervention. Whether the construction of a single composite indictor has the advantage of synthesize a complex phenomenon, it fails in capturing the detailed information that determine such a result, providing a net combined effect that could hide priorities and critical aspects. Our aim is not to range individuals or regions based on their final result on "social exclusion" but to identify the main risk factor according to individuals or regions' characteristics. Our objective is to study the social exclusion phenomenon structure, and how this structure changes across the European regions.

Some authors perform regression analysis for a single country or for specific segments of the population (children, senior citizens, immigrants, and so on), which seek to describe the changes in certain indicators of social exclusion without giving a comprehensive framework of the phenomena (Burchardt *et al.* 1999, Gordon *et al.* 2000, Kieselbach 2003, Scharf *et al.* 2005). In the most recent literature, we found some attempts to propose analysis based on Latent Variable models (Dewilde 2004, Moisio 2004, Whelan and Maitre 2005b, Robila 2006). These researches represent a first and useful attempt in understanding the dynamics of the social exclusion process; however, they are based primarily on objective and economic data. We recognized a lack of applied researches that focus on the individual elements (perceptions, values, social relations and so on) with appropriate and rigorous statistical methods.

Through LCA, we deal with social exclusion as latent construct. We dispose of highly interrelated observed measures (categorical and metrical, objective and subjective), which association is due to some underlying unobserved factors assumed to be (or treated as) categorical. In our empirical application, the latent classes represent the latent levels of social exclusion, which structure the cases with respect to a set of observed indicators. Even if it would be possible to think the level of social exclusion lying on a continuum, we think that the categorization of the concept yields to a more meaningful and operational distinction. All the more so, given the multidimensionality of the concept under investigation, there are reasons to believe that the latent concept is not strictly ordinal, thus assuming a discrete latent variable could allow different patterns to stand out. Working with a discrete variable enables to define different patterns of social exclusion according to the different dimensions, and analyse their features.

Starting from the conceptual framework depicted in Chapter 3, social exclusion can result from breakdowns in any of the identified dimensions. But it seems likely that we can only truly talk of social exclusion when, for individuals or groups, several of these systems break down. In point of fact, the major risk is that a single breakdown triggers a mechanism of instability also in the other dimensions of human life, as a chain reaction. Since these elements refer to different area of human life that interact and influence reciprocally, exclusion in one dimension could determine or make worse exclusion in the others. For instance, whether individuals are employed but poorly integrated in terms of family or community system, an unexpected (long-term) unemployment may lead to social isolation, which in turn will accentuate tendencies of poverty and civic marginalisation, culminating in social exclusion. On the other hand, the situations in which people are excluded from all the dimensions, contemporaneously and for long time, are very rare. This conceptualisation led some authors to conclude that it would be preferable to analyse separately each dimension of exclusion, rather than to think at socially excluded as a homogeneous group (Burchardt *et al.* 2002). Anyway, we consider a multidimensional analysis more useful to address social exclusion issues.

The identification of latent classes, structuring the cases with respect to the observed indicators, enables to identify different profiles of respondents. Taking into account, at the same time, several indicators describing different domains and sub-domains, leads to a better understanding of the weakest points according to different situations.

The proposed model enables us to attain simultaneously the identification of different profiles of the respondents and of the regions in which respondents are nested. That is, we allow social exclusion to manifest itself in different ways for different subgroups in the European population. In fact, the classical LC model is extended by introducing a hierarchical component, focusing not only on the differences between groups (namely European regions), but also on the latent distribution of each group. Individuals, for whom a set of responses variables (indicators) is provided, represent the first level of the analysis; the regions in which individuals live represent the second level. For the specification of the mixing distribution, we follow a nonparametric random-effects approach, introducing a *discrete latent variable* also at group-level. Random-effects approach is often used to take into account group differences in multilevel analysis (Skrondal and Rabe-Hesketh 2004; Snijders and Bosker 1999). Regarding the analysis of social exclusion in a comparative approach, it is useful both from a substantial and technical point of view. The basic idea of hierarchical LC models is exploited to take into account the existing multilevel structure of European population to model regional differences in the distribution of the latent variable, allowing some parameters to differ across regions. Hierarchical Latent Class models enable to account for the fact that individuals belonging to different groups respond to certain items in a different manner. An objective of our analysis is to verify if European citizens have different perception of poverty and social exclusion depending on the context in which they are embedded. We choose a categorical variable also at group level because our aim is to map European regions according to the similarities in their level and features of social exclusion, instead of ranking them.

Preliminary analysis (Section 5.3) highlight that there exist differences across European regions in the individual's perception of social exclusion. Particularly, areas characterized by a high perception of social exclusion include not only poor regions, but also some areas that would not be classified as disadvantaged based on objective indicators: the role of economic conditions seems to be reduced introducing also elements of subjective perception (Pirani and Schifini 2008). This represents an interesting result, since at political and institutional level it is usual to consider only objective socio-economic indicators to evaluate and measure social exclusion. Secondly, different elements connected to social exclusion, namely income, unemployment, social ties and so on, have different importance in different European regions. Introducing the second level of nesting shows how geographical differences explain a certain degree of heterogeneity in the process of social exclusion and in the characterization of weakest population groups. This supports the hypothesis of presence of different latent structures of the theoretical construct of social exclusion, endorsing a multilevel approach. Our preliminary findings are consistent also with previous researches. Using a different methodology, Tsaklogou and Papadopoulos (2002) and Ogg (2005) for example showed a significant relationship between a country's welfare regime and the risk of social exclusion its population faces, even after controlling for individual characteristics.

5.1.2 Regions vs. nations

Studies concerning poverty, social exclusion, inequalities in well-being and life conditions, undertaken in a comparative approach, impose the choice about the most appropriate geographical level of analysis. To date, analysis involving Europe, pay attention predominantly to the national level. This is a consequence of the policy targets of European Union, traditionally addressed to the nation as a whole, an attitude confirmed by adopted indicators, measured at national level. The most recent example refers to Laeken indicators of poverty and social exclusion (European Commission 2003). Although recognising the matter of geographical differences not only across but also within countries (Atkinson *et al.* 2002), Laeken Indicators have been formulated as measures – of poverty and inequality, health and education as well as unemployment – at national level, the only exception being one measure about regional disparities.

Different studies have emphasized the fact that in comparative approaches the unit of "region" (i.e. a subnational administrative unit) is assuming a central role and a broad significance as unit of investigation. On the contrary, the use of "nation" is losing its relevance and can no longer be considered an exhaustive unit of measurement in most social science fields. Looking at the internal situation within each nation, for example Petrucci and Schifini (2004) found that there are zones joined together by similar economic, social, and cultural behaviour, which do not correspond to national borders; in addition, there are areas similar to other areas outside their national boundaries. Again, a regional breakdown allows to examine how a country's best-performing region matches up to the best or worst-performing region in another State. Both considering economic profile and social attitude, it is notorious that deep differences exist within nations. These wide imbalances would be hidden under a too aggregate analysis. The existence of significant intra-country disparities for different socio-economic indicators, such as the poverty or the unemployment rates, is well documented (Stewart 2003, Heidenreich 2003). It is also true that the degree of regional disparities depends on the indicator examined, since is not the same region within each country that performs best or worst at all times.

In the specific case of analysis of social exclusion, another element turns to be relevant. We stated that social exclusion is a relative concept. This implies exclusion from someone (a social group that tends to exclude another one) and from somewhere. That is, exclusion could manifest at different levels. Burchardt *et al.* (2002, p. 7), in representing the social structure like an "onion" diagram, argues that individual level is influenced by many other levels, namely family, community, local, national and, finally, the global one. This is true, hence, it is clear that individuals are first excluded from their local community, and it does not make sense in a practical purpose to consider a global dimension without considering a level closer to individuals. In Chapter 3 we argued that the feeling of exclusion of marginality depends also on the comparison with the others, the comparison group being the place where people live, namely their town or region. To participate in a "minimum acceptable way of life" people need to be able to dress, eat and travel in similar ways than their friends and colleagues (Atkinson 1998). The use of a reference group brings the definition of a context closer to the social reality where people under investigation live. At the same time, researchers and policy makers may need to evaluate whether and to what extent the lack of services and opportunities in a local community prevents individuals to participate in normal activities and changes. Whether considering multidimensional aspects of exclusion, the differentiation in these elements and in surrounding conditions arises to be relevant in differentiating situations and processes.

5.2 Sample and data

5.2.1 The Eurobarometer sample

For our empirical application, we attempt to provide an operational approximation to the approach defined in Chapter 3 using the information available in the 2001 "Eurobarometer 56.1" [35]. In the Eurobarometer 56.1 there are several questions that can be theoretically assigned to the dimensions of social exclusion. We are aware that there are significant omissions in the chosen dataset (e.g. household composition, active citizenship, access to goods and services market), but at the same time it represents the unique source that covers, even if in an incomplete manner, all the dimensions we identified as relevant for the study of social exclusion. This source also enables to adopt a subjective perspective. We acknowledge too that these data are not really update. This prevents to draw conclusions about the current situation in European Union, and we can only represent a

[35] Cf. Gallie and Paugman 2002. The data were collected between September and October 2001 by a consortium of market research agencies under the overall co-ordination of INRA (Europe), at the request of *Directorate-General Employment and Social Affairs, Unit E2 Social Protection and Inclusion Policies – European Opinion Research Group*. The complete Eurobarometer 56.1 questionnaire is available on the website http://www.gesis.org/en/data_service/eurobarometer/.

picture referred to the past. Anyhow, the object of this thesis is to check the relations between the different dimensions of social exclusion and to develop a comprehensive methodology that could be applied using future data. For our purpose, Eurobarometer represents the unique source of information currently available for European regions, which looks on social exclusion in a broader sense. Although implicitly, this survey considers the different dimensions of social exclusion, and through a set of dedicated questions, it investigates the respondents' point of view about poverty and social exclusion.

Eurobarometer data are furnished for all European countries[36] and the survey covers the population aged 15 and over. In each EU country, a number of sampling units are drawn with a probability proportional to population size and – for a total coverage of the country – to population density[37]. The sample size is about 1,000 for each country[38] except small countries like Luxembourg (600) and Northern Ireland (300). The samples of the Eurobarometer surveys are designed for comparative analysis among national populations, and the sample size allows equally precise estimates for small and large countries, as well as comparisons between sub-groups with respect to basic demographic variables.

The complete available sample of Eurobarometer 56.1 includes 15,943 individuals nested in 77[39] regions belonging to 15 countries. As typical in social sciences research, some individuals in the sample did not answer to all questions, giving rise to missing values (see Table A.1 of the Appendix A for the missing data in our dataset). The usually adopted strategies are the substitution of the missing values, or the exclusion of those cases without complete data. In Latent Class Analysis, methods to deal with missing values are different depending on the type of variable concerned. Considering missing values on observed indicators, a solution is to include all the cases in the analysis, the likelihood contribution

---

[36] Data concerning new members of European Union and candidate countries are gathered since 2001 through proper surveys. However, since not all information is available for them, we decided to concentrate our analysis only on the 15 countries that belonged to European Union in 2001.

[37] Detailed information about sampling can be found in Eurobarometer websites: http://www.gesis.org/en/data_service/eurobarometer/ and http://ec.europa.eu/public_opinion/. In brief, since 1989 in all member states it is used a multi-stage random probability sampling design. The sampling is based on a random selection of sampling points after stratification by the distribution of the national resident population in terms of metropolitan, urban and rural areas. These primary sampling units (PSU) are selected from each of the administrative regions (NUTS level 2) in every country. In the second stage a cluster of addresses – chosen systematically using standard random route procedures – is selected from each sampled PSU. Then, in each household, a respondent is selected by a random procedure. New independent samples are drawn for each Eurobarometer survey.

[38] Although the Eurobarometer 56.1 refers to the fifteen countries belonging to European Union in 2001, the survey counted 17 sampling areas: Germany was divided into East and West, and United Kingdom into Great Britain and Northern Ireland.

[39] More precisely, the NUTS level-1 would lead to 72 regions for the fifteen EU member states. Anyway, for some countries constituted by a unique NUTS at first level, we used NUTS level-2, whereas possible. It was the case of Ireland, Finland and Portugal. Conversely, for Sweden, Denmark and Luxembourg we did not have more detailed data, neither Eurobarometer nor Eurostat one, thus we decided to consider these countries at the national level. Finally, we note that concerning Greece, the Attiki region is not a separate unit in Eurobarometer data.

will be based only on the observed information. For complete data, the likelihood contribution is based on the K observed indicators, while for respondents presenting H missing values, one computes likelihood based only on the $K - H$ indicators (Vermunt and Magidson 2005b). That is, parameters are estimated using all available information for each case. Applying this approach, the results obtained from the respondents are valid under the missing at random (MAR) assumption (Little and Rubin 1987, Vermunt 1997, Skrondal and Rabe-Hesketh 2004). An analogous specification is possible when assuming the presence of local dependencies within latent classes. Also in this case the multivariate probability containing correlated indicators is based on the observed variables for the person concerned, so excluding missing values. Concerning missing values on the covariates, one can chose to exclude those incomplete records from the analysis or, alternatively, to imputing missing values. Following the second approach, missing data on covariate are usually replaced by the mean over all cases without a missing value.

In our application, we decided to include in the analysis cases with missing values on some indicators, while no missing values on covariates were present in the dataset.

5.2.2 The multidimensional approach: indicators of social exclusion

When trying to operationalise the concept of social exclusion in a multidimensional way, researchers have to deal with many theoretical, methodological and practical problems that could influence the results. Major research problems involve choices related to the available data, the life domains to be included, the selection of indicators for each domain, and the choice of a method to evaluate the relative importance of the indicators. The main problem that researchers have to face in analysing social exclusion is the difficulty in measuring all the relevant aspects. We do not dispose of all necessary information and many of these elements are not exactly measurable. The lack of a shared definition of what is meant by social exclusion represents one of the causes of this data incompleteness; on the other hand, the multidimensional nature of social exclusion implies that there is no one "true" indicator for it. A way to approach the problem is thus to manage a set of indicators that refer to different aspects of social exclusion, and then attempt to understand the manner in which those indicators are related to it. That is, we treat social exclusion as a latent concept that can be only measured via indirect indicators, which represent their different facets.

In Chapter 3 we identified three principal dimensions of social exclusion, i.e. economic, social and institutional dimension. Referring to the previous discussion about the content of these dimensions, in this section we present the available indicators from Eurobarometer survey that we assume to offer a useful characterization of each dimension. The three dimensions of social exclusion are strictly connected each other, and they overlap for some aspects. Anyway, it is possible, at least at a theoretical level, to consider them separately. For the most part of aspects under investigation, we dispose of information both from a subjective and from an objective perspective. The available variables potentially referring to the three dimensions are described in the following, and explorative analysis showed their relevance. In Table 5.1 we present the complete list of the indicators. Anyway, only the not redundant variables have been included as indicators in the final model discussed in this thesis. Even if we dispose of a large dataset, LC models imply the estimation of a large number of parameters that limit the number of indicators that one can add. In some cases, we preferred to aggregate some indicators or categories of them, in order to obtain a more parsimonious model. Subsequently (5.2.3), we briefly describe also covariates, both at individual and regional level. Finally, a discussion concerning the appropriate level of aggregation for macro data is afforded (5.2.4).

Economic dimension

Nevertheless its limits, the actual income of individuals remains an important indicator of social exclusion. The objective measure of the economic and financial situation may be introduced by means of the income quartile of individuals. As expected, the income quartile variable contains an high rate of missing values (about 30% in our dataset. Cf. Table A.1 of the Appendix A). For comparative purposes the self-rated measure of income is an useful indicator. Eurobarometer 56.1 asks individuals how well they get by with their income, with four categories of response: with great difficulty, with difficulty, easily, very easily. Generally, there is a significant correlation between the self-rated measure of economic difficulty and the objective one (Bhalla and Lapeyre 2004). Anyway, the simultaneously introduction of both indicators enables to highlight discrepancies between the actual economic situation and the economic situation perception.

In order to balance this perception we can add an indicator of the economic difficulties that people actually coped with in last twelve months, using the responses given to questions concerned the occurrence of problems like paying rent or mortgage, paying bills, paying food and repaying loans. Using together these four variables, we build a composite indicator measuring the overall magnitude of economic difficulties[40]. Another available indicator concerning the material life condition is the satisfaction with one's own house or flat.

Finally, to better characterize the economic dimension from a subjective perspective we may refer to the degree of agreement expressed by the respondents to the question: "Some people look down on me because of my income or job situation". We deem that this indicator may account for the subjective perception of a sense of inferiority perceived by people, that evidently could present an high degree of variation, due to the cultural environment and the level of disparities among regions.

[40] The original responses to the four questions (no problems, some problems, a lot of problems, enormous problems) have been recoded in a synthetic indicator taking the following modalities: a lot of difficulties, some difficulties, no difficulties.

Social dimension

To be engaged in significant interaction nets with family or friends, and to be identified with a cultural group or a community, represent ways to feel part of a society, thus enhancing a more social cohesive society. Using this idea as a starting point, our purpose is to identify a set of indicators that capture non-material aspects of exclusion, introducing the deprivation also in the domain of social relations.

First of all, the Eurobarometer dataset provides information about the frequency of the relationships with the "immediate" sphere of relations of people. Particularly, it is asked if people meet their friends, relatives and neighbours several times a week (yes or not). Using together these three variables, we build a composite indicator measuring the overall magnitude of personal relationships[41]. Moreover, to capture the existence of effective social networks, respondents were asked how much practical and emotional support they would expect to get from members outside their household in three situations of need: whether they feel depressed, they help need to find a job, they urgently need to borrow money. Besides these kinds of personal relationships, Eurobarometer offers the possibility to investigate also the participation in social activities like leisure or sport clubs, voluntary or charitable organisations.

Among the others, an interesting Eurobarometer question is attention-getting for the purpose of our analysis: "Do you feel left out of society?". Respondents had to say whether they agree (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree) with this statement. Using this subjective perception of social exclusion we can investigate to what extent risk factors traditionally relied to social exclusion are really decisive in individual perception. The Eurobarometer dataset allows studying also another element from a subjective perspective. People were asked the degree of agreement with the statements: "I don't feel that the value of what I do is recognised by the people I meet" and "I don't feel that I have the chance to play a useful part in society". To be engaged in activities which are positively valued by others is important for the psychological wellbeing of people, and may contribute to enhance social relations and social participation. On the contrary, the perception to not have a useful role in society may increase social instability, marginalization and thus disrupt the capability of citizens to engage in social activities in their community.

In Chapter 3 we remarked the importance to extend the analysis also to indicators referring the labour market, particularly the quality of job (in terms of relations, environment, competences and career), the job security, the precariousness, an so on. However, even if such indicators would be somewhat available in our dataset, we decided to not to use them. In fact, the introduction of these indicators would exclude from the investigation people who do not participate in labour

[41] This indicator takes the value "high" for individuals who respond to have frequent social contacts with friends, relatives and neighbours; "medium" when the social contacts are frequent for two categories of subjects; "low" when one has frequent contacts only for one category, and "very low" whether all kinds of social contacts are scarce.

market, limiting our analysis which, on the contrary, aims to study social exclusion for the entire population.

Institutional dimension

The institutional dimension represents the most difficult one to measure at individual level. Nevertheless, using some questions provided by Eurobarometer, we attempt to account, to some extent, for the attachment between citizens and public institutions, and their satisfaction about them. Data are provided about the subjective evaluation of the respondents about the medical services in their local area, their social entitlements in case of sickness, invalidity and unemployment, but also about the travel and the shopping facilities in the area, the level of noise and the job opportunities. For all these elements, respondents were asked to express their satisfaction through the scale: very good, fairly good, very bad. These categories of response have been recoded into good *vs*. bad.

A set of questions investigated how people perceive the presence of violence, vandalism, theft and drug addition, the state of buildings and the reputation of the area, in the place where they live. These indicators may be useful to evaluate the sense of security of individuals and to depict their opinion about the quality of local environment.

Concerning the voting participation, Eurobarometer survey 56.1 did not gathered any information. To evaluate the political engagement, the only available information concerned the membership to a political party. But, evidently, such a question investigate a specific involvement, more than the voting participation.

An overall evaluation about the political and public environment may be derived using the responses to the question "Would you say you are very satisfied/fairly satisfied/not very satisfied/not at all satisfied with the way democracy works in your country?".

Table 5.1 – Potential indicators of social exclusion provided by Eurobarometer survey 56.1, 2001

ECONOMIC SITUATION		
Indicator	***Description***	***EB question***
Income perception	How well do you get by with your income?	Q5
Income quartile		D29
Economic difficulties	In last year have you had problems in paying rent/mortgage, bills, food, loans?"	Q9
Inferiority perception	"Some people look down on me because of my income or job situation"	Q21.7
Evaluation of own house or flat	What do you think about your house or flat?	Q20.1

SOCIAL RELATIONSHIPS		
Indicator	***Description***	***EB question***
Frequency of social contacts	I meet my friends several times a week I meet up my relatives several times a week I talk to my neighbours almost every day	Q5 Q6 Q7
Potential support	Is there anyone you could rely on to help you if you were feeling depressed Is there anyone you could rely on to help you if you needed help finding job Is there anyone you could rely on to help you if you needed to borrow money to pay urgent bill	Q19
Participation in other activities or associations	I am a member of a leisure or sports club	Q18.10
Perception of social exclusion	I feel left out of society	Q21.04
Perception of family exclusion	I feel left out of my family	Q21.05
Perception of useful in society	I don't feel that I have the chance to play a useful part in society	Q21.06
Perception of own value in society	I don't feel that the value of what I do is recognised by the people I meet	Q21.02
INSTITUTIONAL DIMENSION		
Indicator	***Description***	***EB question***
Satisfaction about medical services	What do you think about medical services in the area? (very good, fairly good, fairly bad, very bad)	Q20.08
Satisfaction about social entitlements	What do you think about social entitlements in case of sickness, invalidity, unemployment, old age? (very good, fairly good, fairly bad, very bad)	Q20.09
Perception of democracy	Which is your degree of satisfaction with the way democracy works in you country?	Q46.05
Quality of local environment	What do you think about: - shopping facilities in the local area - job opportunities in the local area - travel facilities	 Q20.11 Q20.12 Q20.05
Quality of local environment	With reference the area in which you live, do you agree with the following statements? - it has buildings in a bad state of repair - there are problems of drug abuse - it has a lot of vandalism and theft - there is a lot of violence - it has not got a good reputation	 Q21.09 Q21.11 Q21.12 Q21.13 Q21.14

5.2.3 Individual and contextual covariates

In this study we analyse social exclusion situations considering the population in its complex. The entire population forms a very heterogeneous group, in terms of age, economic situation, occupational status, living conditions, and so on. Moreover, there is a certain degree of heterogeneity, due to the fact that respondents belong to different geographical areas. Here, we provide a framework of the possible individual-level and contextual-level covariates of such diverse situations. As discussed in Chapter 4, in Latent Class models covariates help to predict the membership to latent classes, thus improving their description. In this sense, covariates differ from the indicators, which conversely are used to define and measure the latent concept.

The individual's own attributes available in the dataset encompass age, gender, and educational level, above all. Some studies investigated the correlation between these traditional covariates and social exclusion situation (e.g. Böhnke 2008, Ogg 2005). The occupational status of the respondents (employed, homemaker, unemployed, retired/unable, student) is also included. The unemployment status has several consequences on the individual, involving not just a lack of financial resources, but also a weakening and a change in social network of individuals (Negri e Saraceno 2000). Another risk factor is the seniority, the effect of the retirement on financial situation being known.

Other events in the individual's life, such as partnership breakdown, loss of job, and interruption of school studies, should be accounted for (Burchardt *et al.* 1999). Unfortunately, given the questionnaire's structure, this information is available only for a subgroup of respondents. Indeed, only whether the respondents said to be in financial difficulties they were asked the occurrence of such "traumatic" events in their life. We thus decided to not to use these information in the empirical application presented in this thesis. This questionnaire strategy confirms the confusion between the concepts of poverty and social exclusion.

As previously discussed, also elements operating at regional level are relevant, for a double reason: a) they allow to describe the cultural context in which individuals are embedded and which may affect their responses; b) accounting for economic and social context, these covariates help in characterizing environment and thus facilitate the identification of the latent variable at second level.

Objective information has been drawn from Eurostat dataset[42], and refers to 2001. Among the socio-economic statistics available at regional level, the potential indicators describing the contextual environment are: the gross domestic product (absolute level and growth rate), the net disposable income, the level of social benefits, social transfers and taxes. We defined an indicator given by the ratio between the amount of taxes, social contribution and transfers paid, and the primary income. The latter indicator could represent a proxy of the social protection expenditure of the region or, even, the amount of expenditure financed using public taxation. Although the effectiveness of social protection policies is not necessarily relied to the overall level of the expenditure, it may be helpful in comparing objective information with the subjective perception of individuals about public services, social assistance and protection measures. Moreover, information about the labour market includes different unemployment rates (total, for women, for young people), long-term unemployment rate, and total and women employment rate. Eurostat gathers information also about the number of beds available in hospitals. Other useful information is not available at regional level. For example

[42] www.ec.europa.eu/eurostat/.

a risk-of-poverty rate, the persistency of poverty risk, or inequality of income distribution, are computed just at national level.

Cultural and moral contexts in which people are embedded represent important elements that may be assumed to influence individual behaviour, perception and attitudes. It is not a trivial issue to introduce in a statistical analysis these elements. The frame of values and attitudes potentially includes a large set of elements, such as religiousness, sense of solidarity, capability of integration or tendency towards marginalization, civilization, family role, and so on. Since our analysis of social exclusion is carried out at individual level, we cannot properly use them as indicators of social exclusion. Nevertheless, we can introduce some elements at regional level. Particularly, starting from individual responses in Eurobarometer, we computed some indicators in order to quantify, for each European region, the diffusion of participation in political parties and voluntary and charitable organizations, the religious involvement, and to measure the percentage of people attributing the responsibility of poverty and social exclusion either to individual or to societal failure. The latter indicator describes whether the prevalent opinion in a population is that poverty and social exclusion are personal responsibility of each individual living in these situations, or are instead a consequence of injustice in society. Solidarity with and willingness to help the poors will probably be more widespread when responsibility is largely ascribed to injustice in society (Böhnke 2008).

The list of the covariates used in the estimated model is reported in Table A.2 (individual covariates) and in Table A.3 (regional covariates) of the Appendix A.

5.2.4 Territorial units

The decision about the most appropriate regional breakdown is not a trivial and without implications decision, the unit of investigation affecting substantially the results (Stewart 2003, European Commission 2005). For a number of substantial and practical reasons, the logical choice falls upon the administrative units defined at European level, specifically NUTS regions. The Nomenclature of Territorial Units for Statistics, (NUTS), is a geocode standard developed for statistical purposes by the European Union, for referencing the administrative divisions of its members countries. The mainly reason supporting this choice lies in data availability: Eurostat already provides statistical information for NUTS regions, which can be used to construct indicators of regional poverty and deprivation, and to obtain covariates that can be used in statistical models. Moreover, Eurobarometer survey 56.1 provides information about the respondents' region of residence. This information essentially corresponds at the NUTS area, at different level depending on countries. Moreover, the NUTS system provides units that are hierarchical and cover the entire population exhaustively, without overlaps. It comprises three levels, namely NUTS-1, NUTS-2, NUTS-3, which attempt to correspond to administrative divisions within each country.

The choice of the most suitable sub-national level must necessarily represent a compromise between significance and homogeneity of the territorial unit and the availability of statistical information about it. Given current survey research, lack of micro-data or problems with robust measurement of many indicators make it clear that a greater level of detail is unachievable. For our purpose, we will consider the NUTS level-1 regions. The choice of the Nuts level-1 as territorial units allows to satisfy once a twofold objective: on the one hand it represents the lowest level for which one disposes of reliable and well-founded data; on the other hand, it approximates well the environment and the socio-economic context in which individuals are embedded. Moreover, it would bear in mind that a number of member States have decentralized major elements of policy to local governments (at different levels), giving them legitimacy and means to manage and make intervention in their local community in several fields of public interest: education, health, social protection, poverty, social exclusion, etc. At the same time, EU recognizes directly the local governments' instances. Although NUTS-1 (referred to as "regions" in subsequent discussion) are not defined all over in the same way[43], and differ in number and size and, more, in their power of intervention in public policies, they enable to take into account a sort of meso-level, between macro social structures and micro-demographic characteristics.

A last critical and practical question is still open, about the ability of existing data to give reliable answers about patterns of regional disparities in the EU. Indeed, most part of European statistics still refer only to the national level, while certain fields are not covered by NUTS-1 or NUTS-2 level information. The decentralization of significant elements of policy to regional level imposes the necessity to establish statistical indicators and other tools in order to measure and assess the outcomes and the effects of these policies, as it happens at national level.

In conclusion, we showed that differences between regions belonging to the same nation already exist, from the points of view of economic performances, social protection system, attitudes, culture and opinions. Thus, it is clear that, even if less than perfect, regional level makes additional nuances to a picture based purely on national averages.

To summarise, the hierarchical structure of our analysis consists of 15,927 observations nested in 77 regions, with minimum and maximum group sizes equal to 7 and 1,001 respondents, respectively (Table 5.2. See also Table A.6 of the Appendix A for the complete list of regions introduced in the analysis). The unbalanced structure is not a problem, as it is efficiently handled by maximum likelihood methods. The number of clusters and their sizes are sufficient to achieve high power and good accuracy of the asymptotic distributions of the estimators (Snijders and Bosker 1999; Maas and Hox 2004).

[43] NUTS level-1 include, for example, 16 territorial units in Germany, 12 United Kingdom, 9 in France, 7 in Spain, 5 in Italy, 4 in the Netherlands and in Greece, 3 in Belgium and Austria, and 2 in Finland. Instead, five European Union countries – Denmark, Ireland, Luxembourg, Sweden and Portugal (excluding the two autonomous regions) – consist of just one NUTS-1 region each, element that could solicit questions about the degree of intra-regional disparity.

Table 5.2 – Respondents, regions and countries, EB sample 56.1-2001

Countries	N. regions	N. respondents	Respondents in Regions	
			minimum	maximum
France	8	1,002	44	196
Belgium	3	1,032	34	590
The Netherland	4	1,006	61	479
Germany	16	2,009	10	303
Italy	5	992	53	284
Luxembourg	1	600	411	600
Denmark	1	1,001	786	1,001
Ireland	2	996	101	718
United Kingdom	12	1,288	18	304
Greece	3	1,004	69	580
Spain	7	1,000	27	273
Portugal	7	1,001	7	343
Finland	4	996	89	645
Sweden	1	1,000	869	1,000
Austria	3	1,000	127	433
Total	*77*	*15,927*	*7*	*1,001*

Source: Our elaboration on Eurobarometer 56.1-2001 data.

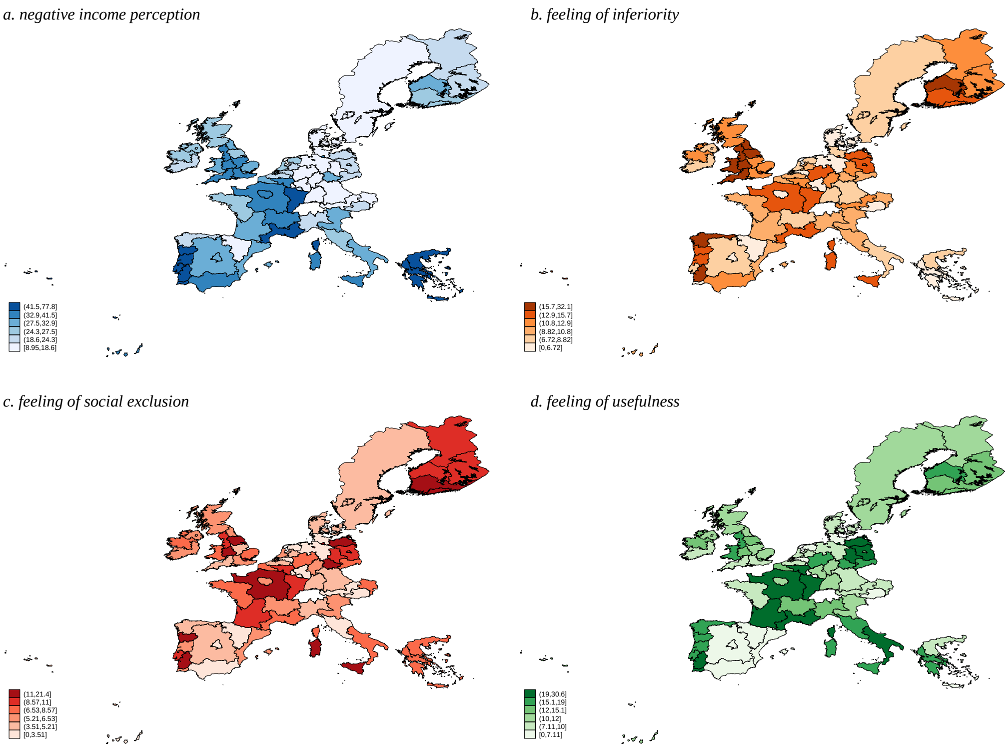
5.3 The European context: indicators and latent variables

In this paragraph we describe some features concerning the variables we will use as indicators of social exclusion and the covariates, presented in § 5.2.2 and 5.2.3. We briefly describe some principal findings about differences among European regions. All these indicators prove that variability is high not only between nations, but also between regions within nations, showing as well as poverty and social exclusion represent a major challenge for all countries in European Union.

Figure 5.1 shows how heterogeneous is, at the regional level, the perception of poverty and of some aspects we relied to social exclusion, according to the indicators used in the analysis. This negative perception is present all over European regions, even if with differentiate intensities. Map *a* of Figure 5.1 depicts the regional distribution of the perception of poverty. Most of Southern regions experience very high levels of subjective poverty, as well as in almost all French regions and in England, where from 27 to 41% of people declare that their income is not sufficient to make ends meet. On the contrary, in Scandinavian countries, The Netherlands, Germany and Austria, the perception to be poor is lower than the EU average.

The other three maps of Figure 5.1 show the geographical distribution of some of the variables used in our analysis to approximate the perception to be integrated in the society. These maps shows that the majority of European citizens perceive themselves as socially integrated, however, there are some areas in which high percentages of people have negative perception about it. Concerning the feeling to be inferior due to one's own income or job situation (map *b*), the worst situation is in Finland, in UK and in continental Europe, while southern European countries register, on average, lower levels. Finland's regions have also high levels of social exclusion perception (map *c*), beside East Germany and French regions, some UK and southern European regions (namely Greece, south of Italy and some Portuguese regions). The map *d* in the Figure 5.1 depicts how the sense of usefulness is high in almost all French regions, Italy, Portugal, East Germany and Finland. Citizens of Spain (with exception for the north), Austria and The Netherlands experience the lowest level for this variable.

Figure 5.1 – Percentage of respondents having negative income perception, feeling of inferiority, feeling of social exclusion, feeling of uselessness, by European regions

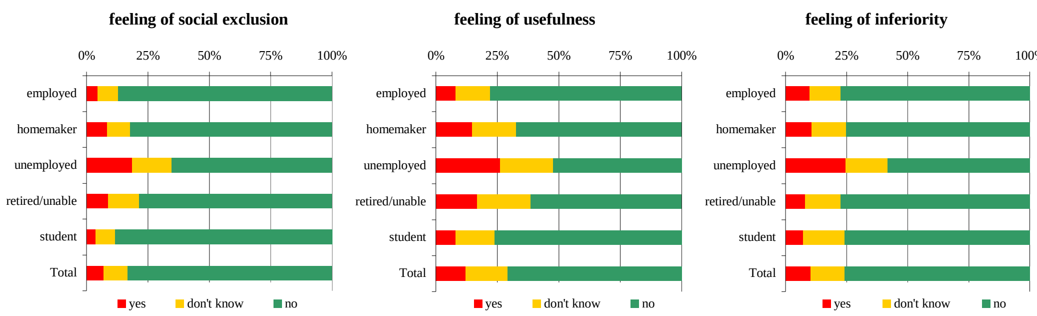


Note: Classes formed with the natural breaks method (Jenks 1963).

Source: Our elaboration on Eurobarometer 56.1-2001 data.

The self-perception of being part of a society is related with the occupational status. For example, unemployed people register the worst situation on all the three variables (Figure 5.2). Referring to retired people, we see that 16.7% (with respect a mean of 12%) perceives a sense of usefulness but, referring on the other variables, they have a level below the average. Both employed people and students perceive themselves as socially integrated.

Figure 5.2 – Feeling of social exclusion, feeling of usefulness and feeling of inferiority by occupational status of respondents



Source: Our elaboration on Eurobarometer 56.1-2001 data.

Table 5.3 shows that the higher levels of social contacts are found for homemakers and retired people. But the 32.6% of retired people declares also to not to have someone, external to their family, to rely on in case of help. This value is high also for unemployed persons. It is interesting to note that the unemployed have levels of social contacts on the average (Table 5.3), but they feel significantly more socially excluded (Figure 5.2). Moreover, their scarce availability of help is probably due to the fact that their social networks are mainly represented by other unemployed people, who are therefore poorly placed to offer significant support. As expected, the participation in social, cultural and sports association is high for students.

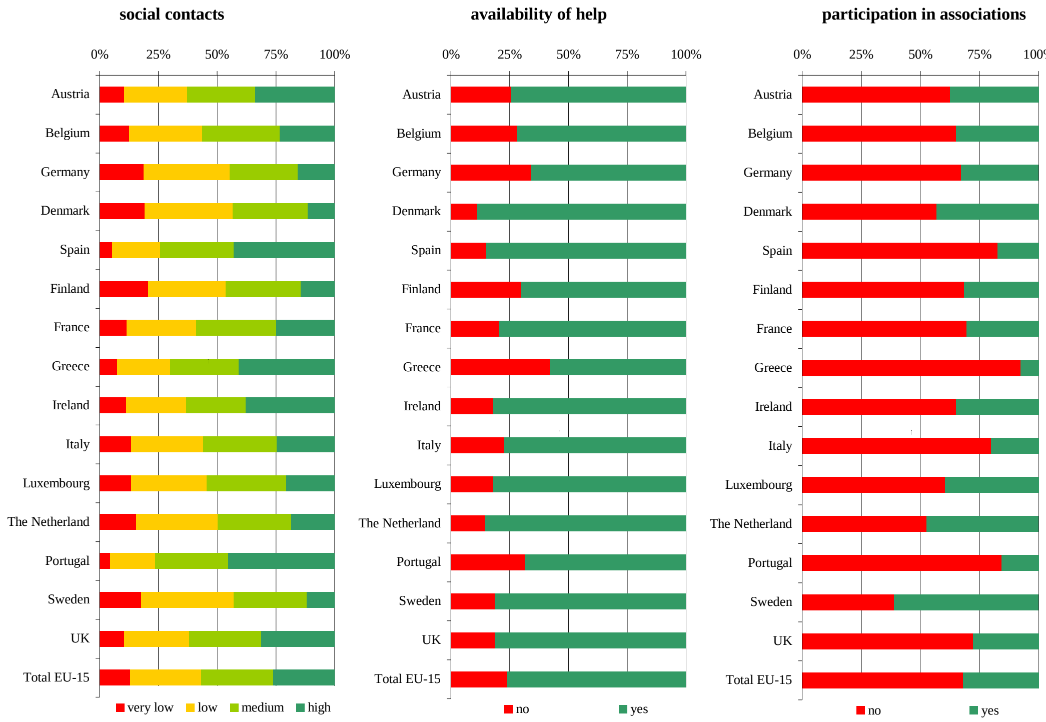
Table 5.3 – Social contacts, availability of help and participation in associations by occupational status of respondents

	social contacts					help availability			participation in associations		
	very low	**low**	**medium**	**high**	**Total**	**no**	**yes**	**Total**	**no**	**yes**	**Total**
Employed	15.7	31.9	30.4	22.1	100.0	20.0	80.0	100.0	63.7	36.3	100.0
Homemaker	9.3	24.4	31.3	35.0	100.0	23.3	76.7	100.0	77.8	22.2	100.0
Unemployed	13.6	30.5	31.1	24.8	100.0	35.7	64.3	100.0	80.8	19.2	100.0
Retired/unable	12.3	27.1	30.0	30.6	100.0	32.6	67.4	100.0	74.2	25.8	100.0
Student	6.3	34.2	34.4	25.1	100.0	16.6	83.4	100.0	55.8	44.2	100.0
Total	13.1	30.0	30.9	26.0	100.0	24.0	76.0	100.0	68.0	32.0	100.0

Source: Our elaboration on Eurobarometer 56.1-2001 data.

Figure 5.3 shows that people in Southern countries (Spain, Portugal, Greece) together with people in Ireland, have higher levels of social contacts with respect the average. Anyway, in same cases, they feel socially isolated (see map *c* in Figure 5.1). This difference should point attention to the importance of both qualitative and quantitative aspects of social relations in explaining the perception of social isolation. In line with the sociability models in European countries, we see that in Northern countries there are the lowest levels of social contacts. However in this countries there are generally high proportions of people with someone to count on, outside their family, in case of need (in case of depression, search for a job or to borrow money). The highest levels of people participating in associations (from 40 to 60%) are found in Denmark, Sweden, The Netherlands and Luxembourg. On the contrary, for Greece, Italy, Spain and Portugal, from 80 to 92% of people do not take part in associations.

Figure 5.3 – Social contacts, availability of help and participation in associations by European countries

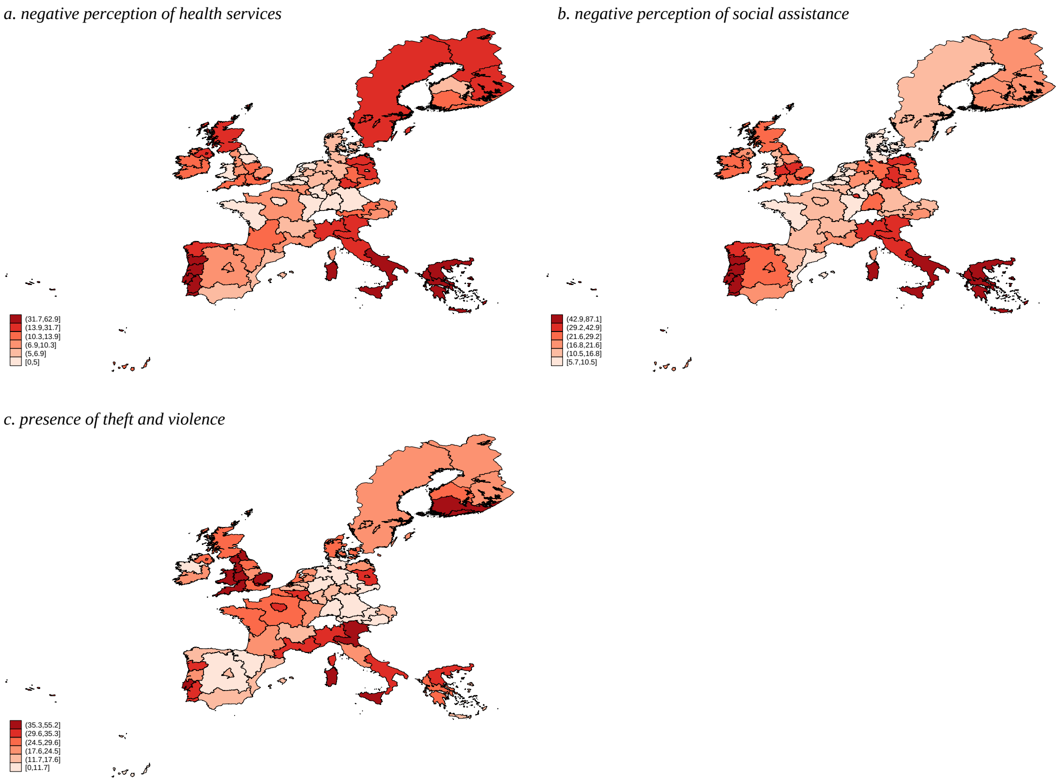


Source: Our elaboration on Eurobarometer 56.1-2001 data.

Also the level of dissatisfaction with the variables we link to the institutional dimension varies across European countries and across European regions (maps in Figure 5.4). In Southern European regions, except Spain, from 25 to 63% of citizens are unsatisfied with the presence of health and medical services in the area where they live (map *a*). Dissatisfaction is present also in Sweden, Finland, Scotland, Northern Ireland and East Germany (10-20%), while continental Europe, together with England, seems to be overall satisfied with respect this aspect of daily life. The social discontent about the social assistance and protection system is depicted in the map *b* of the Figure 5.4. In this case we note a higher homogeneity at country level. Once more, Southern European countries have high level of dissatisfaction in this respect (from 40 to 70%), while the situation in this case is good for Swedish and Finland citizens, and in some continental regions (less than 20%).

Finally, in map *c*, we see that the presence of violence and theft does not represent a problem for Spain, Germany (except the Hamburg region), Austria and Ireland.

Figure 5.4 – Percentage of respondents having negative perception of health services, of social assistance and of the presence of theft and violence, by European regions

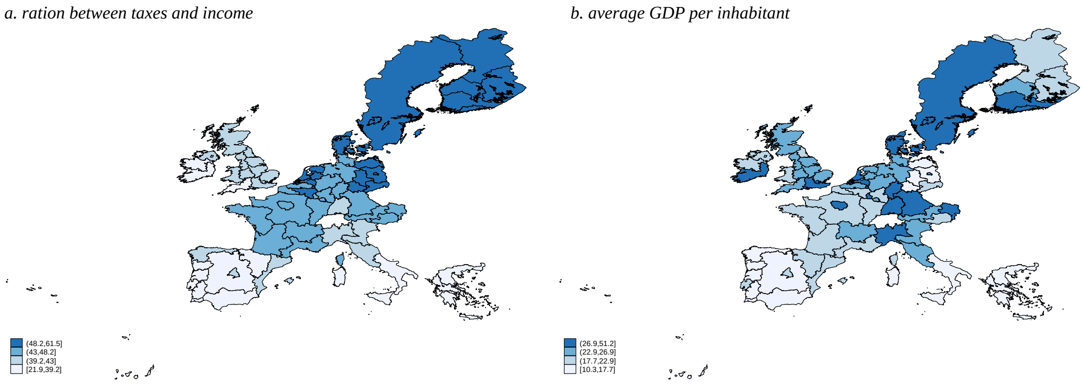


Note: Classes formed with the natural breaks method (Jenks 1963).

Source: Our elaboration on Eurobarometer 56.1-2001 data.

Figure 5.5 shows the average regional values of the two contextual variables used in the analysis: the ratio between taxes and income (map *a*), and the GDP (map *b*). Concerning the first one, we note scarce differences within nations. This is not surprisingly, as well as the fact that the highest level of taxation and social contributions (from 48 to 60%) are for Scandinavian countries (namely Sweden, Finland, The Netherlands, Denmark, and also for Eastern Germany). On the other side, we find southern European countries and Ireland (from 20 to 40%). Major differences among regions, also within nations, are for the mean level of GDP for inhabitants (map *b*). As expected, the richest regions are in the Northern Europe, UK, Germany, The Netherlands, beside Ile de France and North-western Italy (higher than 26,000 of Euro per inhabitant on average). The lowest levels are for south of Italy, Greece, Spain and Portugal, and regions of East Germany (less than 17thousand Euro).

Figure 5.5 – Regional distribution of the ratio between taxes level and the income, and of the mean regional level of GDP per inhabitant (thousands of Euro)

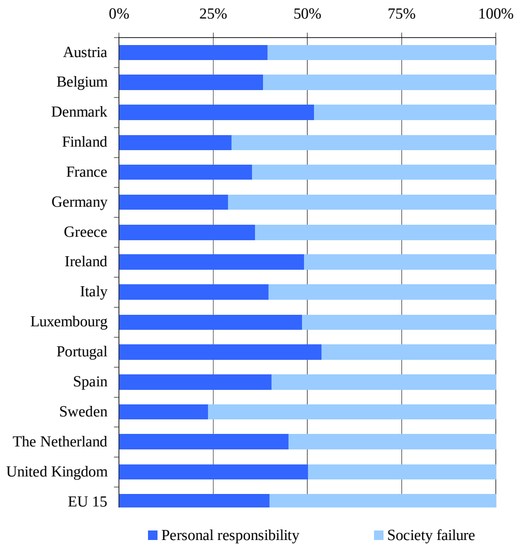


Note: Classes formed with the natural breaks method (Jenks 1963).

Source: Our elaboration on Eurostat data, 2001.

Finally, the distribution of the responses about the causes of poverty and social exclusion is shown in Figure 5.6. Only in Denmark, Portugal and UK personal causes are more important than social causes in explaining poverty. However personal responsibility is over the European average in Ireland, Luxembourg and in The Netherlands. On the contrary, social causes predominate as an explanation of poverty in Sweden, Germany, and Finland, and, in a lower measure, also in France and Belgium. This results poses some questions about the solidity of the European social model based on social justice (Bhalla and Lapeyre 2004). It is worthwhile noting, also, that the injustice explanation varies greatly over time and is related to the overall socioeconomic conditions (European Commission 2004).

Figure 5.6 – Percentage of respondents by poverty and social exclusion as a personal responsibility or as a failure of the society, by European countries



Source: Our elaboration on Eurobarometer 56.1-2001 data.

5.4 Model specification

The model structure implemented to study social exclusion across EU regions, is the following. For each individual i, with $i=1, \ldots, I$, originating from an international sample, we dispose of the responses for a set of indicators denoted by Y_{ik} with $k=1, \ldots, 12$. These indicators refer to the variables previously described. We assume a latent variable X_{ij} that represents the individual's condition of social exclusion. Given their response patterns to the selected indicators, individuals will be classified in a probabilistic way in one of the t latent classes of X_{ij}, with $t=1, \ldots, T$.

This represents the *lower-level part* of the model, that is a standard Latent Class model for the selected indicators with a categorical latent variable.

Fundamentally, we could assume the existence of two most relevant latent classes: "excluded" versus "not excluded". This opposition is certainly clear and not ambiguous, and it enables to identify two clear-cut groups. Anyway, we believe that because of multidimensionality of the concept "social exclusion", it might coexist different sub-groups in the population, each of them characterized by different forms or different degrees of exclusion in each of the identified dimensions. This is the reason why we decided to not limit a priori the number of classes of the latent variable at individual level.

Concerning the *upper-level part* of the model, we consider that individuals are nested in regions, implying that the standard assumption of independent observations does not hold for our data. The 77 European regions considered in the analysis will be identified by the subscript j, that is $j=1, \ldots, 77$. So, Y_{ijk} represent the response of person i coming from region j to item k, whereas $\mathbf{Y}_{ij}$ refers to the full vector of responses of the same individual i, and $\mathbf{Y}_j$ to the full vector of responses of all individuals in region j. Assuming the existence of a latent variable W_j at regional level, with $m=1, \ldots, M$ possible classes, the individual responses are assumed mutually independent given the latent variable at regional level. This latent variable has the role of a random effect in the model for X_{ij}, and it aims to identify latent types of regions for which parameters in the specified model differ. The multilevel component implies that the latent class probabilities vary across regions.

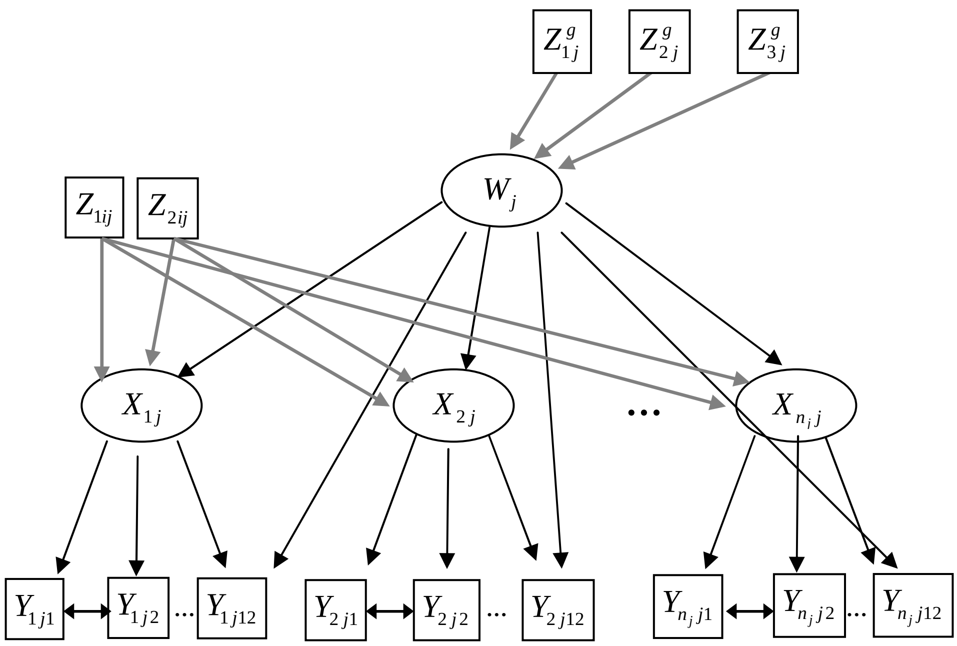
We remark that the latent variables both at individual and at group level are considered as discrete variables. This is not a fundamentally assumption, not even from a substantial point of view. In fact it is possible to think the condition of social exclusion lying on a continuum. Anyway, the discretionalisation of the concept yields to a more meaningful and operational distinction. Working with discrete variables enables to define different degrees of exclusion and analyse their features. Moreover, being one of the objectives of this thesis the mapping of

European regions depending on the similarities in their levels and features of social exclusion, we decided to proceed in this sense.

Finally, the model will be completed with the introduction of a set of covariates both at individual (Z_{pij}) and regional (Z^{g}_{qj}) level, the latter identified using a superscript g. The list of covariates is reported in Table A.2 and in Table A.3 of the Appendix A.

The model structure is depicted in the path diagram of Figure 5.7, which highlight the presence of effects between indicators, between covariates and latent variables, and between latent variables and indicators[44].

Figure 5.7 – Path diagram of the multilevel Latent Class model adopted for the analysis of social exclusion



[44] We modelled the contextual covariates to have direct effects on the latent variable at group level W_j. However, it would be possible to think contextual covariates to affect directly individual latent variable X_{ij}. Four our model, we followed the first alternative, since our aim was to describe and characterize the second level units, i.e. the European regions, by means of these covariates. The second possibility will be further developed in future analysis. We thank Professor Leonardo Grilli for pointing this issue out.

We first derive the probability of the response vector $\mathbf{Y}_{ij}$ conditional on the latent variable at group level:

$$P\left(\mathbf{Y}_{ij}=\mathbf{s}\middle|W_j=m\right)=\sum_{t=1}^{T}P\left(X_{ij}=t\middle|W_j=m\right)P\left(\mathbf{Y}_{ij}=\mathbf{s}\middle|X_{ij}=t\right)$$
$$=\sum_{t=1}^{T}P\left(X_{ij}=t\middle|W_j=m\right)\prod_{k=1}^{K}P\left(Y_{ijk}=s_k\middle|X_{ij}=t\right) \qquad [5.1]$$

The probability associated with all responses of a given region, denoted by $P\left(\mathbf{Y}_j\right)$ can now be obtained by taking the sum over m of the products of $P\left(\mathbf{Y}_{ij}=\mathbf{s}\middle|W_j=m\right)$ over the n_j individual belonging to each region, and multiplying by the probability that region j belongs to a particular class at group level:

$$P\left(\mathbf{Y}_j\right)=\sum_{m=1}^{M}\left[P\left(W_j=m\right)\prod_{i=i}^{n_j}P\left(\mathbf{Y}_{ij}=\mathbf{s}\middle|W_j=m\right)\right] \qquad [5.2]$$

Expliciting the individual response patterns and adding the covariates at individual and at group level, equation [5.2] becomes:

$$P\left(\mathbf{Y}_j\middle|\mathbf{Z}_j\right)=$$
$$=\sum_{m=1}^{M}\left[P\left(W_j=m\middle|\mathbf{Z}_j^g\right)\left[\prod_{i=1}^{n_j}\sum_{t=1}^{T}P\left(X_{ij}=t\middle|W_j,\mathbf{Z}_{ij}\right)\prod_{k=1}^{K}P\left(Y_{ijk}=s_k\middle|X_{ij},W_j\right)\right]\right] \qquad [5.3]$$

that shows the probability structure of the model we adopted.

The right-hand side of equation [5.3] consists of three components:

a) the probability that region j belongs to a particular level of the latent variable W_j, given regional covariates;

b) the probability that respondent i belongs to a particular class of the latent variable at the first level X_{ij}, given regional latent class membership and regional and individual covariates;

c) the joint probability that the i-th respondent following the pattern $\mathbf{s}_i$ given individual and regional latent class membership.

The conditional probabilities that compose equation [5.3] are specified by multinomial logit models:

$$P\left(W_j = m \middle| \mathbf{Z}_j^g\right) = \frac{\exp\left(\alpha_{0m} + \alpha_{1m} Z_{1j}^g + \alpha_{2m} Z_{2j}^g + \alpha_{3m} Z_{3j}^g\right)}{\sum_{m'=1}^{M} \exp\left(\alpha_{0m'} + \alpha_{1m'} Z_j^g + \alpha_{2m'} Z_{2j}^g + \alpha_{3m'} Z_{3j}^g\right)} \qquad [5.4]$$

$$P\left(X_{ij} = t \middle| W_j = m, \mathbf{Z}_{ij}\right) = \frac{\exp\left(\gamma_{0tm} + \gamma_{1t} Z_{1ij} + \gamma_{2t} Z_{2ij}\right)}{\sum_{t'=1}^{T} \exp\left(\gamma_{0t'm} + \gamma_{1t'} Z_{ij} + \gamma_{2t'} Z_{2ij}\right)} \qquad [5.5]$$

$$P\left(Y_{ijk} = s_k \middle| X_{ij} = t, W_j = m\right) = \frac{\exp\left(\beta_{0s_k} + \beta_{1s_k t} + \beta_{2s_k m}\right)}{\sum_{s'=1}^{S_k} \exp\left(\beta_{0s'} + \beta_{1s't} + \beta_{2s'm}\right)} \qquad [5.6]$$

In equation [5.4] we are assuming that the three group level covariates affects level-2 latent class membership, whereas in equation [5.5] the probability of belonging to a certain level-1 latent class depends on the group-level latent variable and on the two level-1 covariates. In equation [5.6] conditional probabilities depend on the individual level latent variable X_{ij}. Moreover, according to the different indicators, we model conditional probabilities in different ways, to account also for direct effects $\beta_{2s_k m}$ of the group-level latent variable or not. It is useful in social sciences, to take into account situations in which individuals belonging to different groups respond to certain items in a different manner. That is, also the conditional response probabilities can be assumed to depend on the latent variable at group level[45].

More precisely, the model specifications [5.4]-[5.6] imply 193 parameters. The three α_{0m} intercepts of [5.4] represent the category effect of the group-level latent variable, while the nine α_{qm} coefficients represent the effects of the contextual covariates on W_j. The 20 γ_{0tm} intercepts in the model for the individual-level latent variable [5.5] are W_j-dependent, thus capturing the differences between the classes of W_j in the category effect of the latent variable at individual-level X_{ij}; instead, the 35 γ_{pt} slopes of [5.5] represent the effects of the categories of the selected nominal covariates on the latent variable X_{ij}. In models for indicators [5.6], the 21 β_{0s_k}, the 85 $\beta_{1s_k t}$ and the 6 $\beta_{2s_k m}$ represent respectively the category effect, the main effects of latent classes at individual-level X_{ij} and the main effects of

[45] In the estimated model we assumed the direct effect of the group-level latent variable only for the indicators: “Social contacts” and “Participation in associations”. Clearly the coefficient $\beta_{2s_k m}$ in equation [5.6] disappears for other indicators not assumed to depend on W_j.

latent classes at regional-level W_j. Note that any covariate has effect on indicators, both at individual and at group level.

Finally, we estimated 14 additional parameters $\beta_{s_k s_h}$ accounting for the direct effects between some pairs of indicators[46], e.g. k and h (cf. equation [4.17]-[4.18]).

In next section, we shall refer mainly to the probabilities given by equations [5.4]-[5.6] rather than directly to the α, γ, and β parameters, that makes the interpretation of the model results much easier.

5.5 Results

The analysis has been performed using the software Latent GOLD 4.5 (Vermunt and Magidson 2005b, 2005c, 2008). Particularly, we used the syntax module[47] that allows a high degree of flexibility in defining model specifications. The parameters of the hierarchical Latent Class model are estimated via an adapted version of the EM algorithm (see section 4.5.2 for further details).

Model estimates can be obtained for a fixed number of classes at group and at individual level, M and T respectively. In order to choose among multi-level Latent Class models for different values of M and T, many models have been estimated, and the relative fit of the alternative model specifications examined by means of the minimum BIC rule. Table A.4 shows the BIC values for different models, provided by Latent GOLD. Moreover, Latent GOLD produces the list of estimated parameters, the bivariate residuals and the profile table associated to the model. For the interpretation of our model results, we shall refer to them. In addition we calculated different predicted probabilities to show the differences between classes and the effects of the latent variable at group level.

The final model we present and discuss here, involves 6 latent classes at individual level, i.e. $T = 6$, and 4 classes at regional level, i.e. $M = 4$ (also subsequently indicated as "clusters"). These classes enable to differentiate rather well the typologies of individuals as regard their deprivation status in all the relevant domains, and to differentiate among regions.

Among the indicators discussed in section 5.2.2 only some of them preserved their significance within a multivariate model. The non-significant indicators have been excluded from the final model, as well as some redundant ones. The complete list of parameter estimates, along with their respective standard errors and *z*-value, is presented in Table A.5 of the Appendix A. The inspection of the bivariate residuals (BVR) represents a useful way to check for the presence of residual association between pair of variables after the model estimation. During the phase of model selection, the analysis of residuals helped in selection of the

[46] For further details about the indicators involved in direct effects specifications, see Appendix B.
[47] The syntax of the final model is reported in Appendix B.

most appropriate indicators, as well as the number of latent classes to retain. In the final model we present, as expected, some associations (high residuals) between certain indicators, and between covariates and indicators, still remain also after controlling for the latent variables. On the other hand, the relaxing of conditional independence assumption for all these cases, would lead to a strong increase of the number of parameters to estimate, thus implying computational problems. Moreover, raising the number of latent classes describing social exclusion does not provide an effective model improvement, both in terms of model fitting and of substantial meaning. Although the evaluation statistics provided useful guidelines to choose the best-fitting model, the final decision was based on the interpretability of the latent classes too.

We first consider the classification of individuals, based on the latent classification or *posterior class membership probability*. This classification gives information on how well one can predict to which latent class cases belong given their observed indicators and covariates patterns. For a given subject $i*$ with a certain covariate pattern, and with response pattern $\mathbf{s}*$, the probability of belonging to the latent class t of X can be obtained by means of the Bayes rule:

$$\hat{P}\left(X_{i*}=t\middle|\mathbf{Y}_{i*}=\mathbf{s}*,\mathbf{Z}_{i*}\right)=\frac{\hat{P}\left(X_{i*}=t\middle|\mathbf{Z}_{i*}\right)\hat{P}\left(\mathbf{Y}_{i*}=\mathbf{s}*\middle|X_{i*}=t,\mathbf{Z}_{i*}\right)}{\hat{P}\left(\mathbf{Y}_{i*}=\mathbf{s}*\middle|\mathbf{Z}_{i*}\right)} \qquad [5.7]$$

where the numerator and the denominator are the ML estimates.

The most common classification rule is modal assignment, which consists to assign each individual to the LC with the highest $\hat{P}\left(X_{i*}=t\middle|\mathbf{Y}_{i*},\mathbf{Z}_{i*}\right)$. This method of assignment is sometimes referred to as empirical bayes modal (EBM) or modal a posteriori (MAP) estimation (Skrondal and Rabe-Hesketh 2004). The classification Table 5.4 cross-tabulates posterior and modal class membership probabilities (Vermunt and Magidson 2005b). More precisely, each entry (t,t') in the table contains the sum of the class t posterior membership probabilities for the cases allocated to modal class t'.

The off-diagonals cases are the misclassified ones, underlying which latent classes are well separated. In addition, Latent GOLD provides the classification error[48]. This index, which ranges from 0 to 1, is not a fit measure, but it is an important measure to evaluate the distinctiveness of different classes. This proportion indicates how well the model can predict latent class membership given the value on indicators and covariates. Concerning our final model the classification error of individuals equals to 0.21.

[48] The proportion of classification error is defined as: $E=\frac{\sum_{i*=1}^{I*} w_i\left[1-\max\hat{P}\left(X_{i*}=t\middle|\mathbf{Y}_{i*},\mathbf{Z}_{i*}\right)\right]}{N}$.

Cf. Vermunt and Magidson (2005b) for further details.

Table 5.4 – Classification table of latent variable X_{ij} based on posterior class membership probabilities $\hat{P}\left(X_{i*}=t \middle| \mathbf{Y}_{i*}, \mathbf{Z}_{i*}\right)$

Posterior class membership probability (t)	Modal classification (t')						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
1	**1,068.1**	316.6	246.8	83.9	50.6	26.5	*1,792.6*
2	143.4	**5,290.8**	386.0	141.3	77.1	1.2	*6,039.9*
3	154.3	448.5	**2,257.0**	76.6	19.5	19.1	*2,974.9*
4	84.4	211.3	101.4	**1,623.1**	39.1	174.6	*2,233.9*
5	40.0	97.5	11.9	41.8	**815.4**	81.0	*1,087.6*
6	25.8	2.3	21.8	194.3	54.4	**1,499.6**	*1,798.1*
Total	*1,516.0*	*6,367.0*	*3,025.0*	*2,161.0*	*1,056.0*	*1,802.0*	***15,927.0***

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

Table 5.4 shows that class 6 is well separated from classes 1, 2 and 3, and class 3 from class 5. Conversely, probabilities of misclassification are present between classes 1, 2 and 3, which are similar for certain aspects.

Examination of the profile Table 5.5 leads to a better understanding of the characteristics of each class, their similarities and their differences. The profile table contains, first, the *estimated marginal latent probabilities* $\hat{P}\left(X=t\right)$ for each t-th class, that is, the class size. Note that whether the model includes covariates,

$$\hat{P}\left(X=t\right)=\sum_{i*=1}^{I*}\frac{w_{i*}}{I}\hat{P}\left(X=t \middle| \mathbf{Z}_{i*}\right) \quad [5.8]$$

where w_{i*} denotes the case weight of the cases with the same covariate pattern $i*$. That is, in models with covariates, the class size $\hat{P}\left(X=t\right)$ is computed by aggregating the model probabilities $\hat{P}\left(X=t \middle| \mathbf{Z}_{i*}\right)$ over covariates values (Vermunt and Magidson 2005b).

Secondly, in the profile table we can read the *class-specific marginal probabilities* associated with each indicator $\hat{P}\left(Y_{ijk}=s_k \middle| X=t\right)$, thus showing how the latent classes are related to the indicator variables. In Latent Class models without associations and direct effects specifications, these are simply the probabilities defining the class-specific distributions. Instead, in Latent Class models that specify direct effects of covariates on indicators and direct association between indicators, probabilities should be obtained by aggregating, for each indicator, over the other variables involved in the effects specification. Moreover, in multilevel models containing group level latent variable that has a direct effect on one or more indicators, the marginal probabilities are obtained summing over the classes of this latent variable (see Vermunt and Magidson 2005b for further details). Clearly, these probabilities sum to 1 within each class (cf. equation [4.6]). Through the examination of profile table, we can characterize each class of the latent variable in term of response probability to each level of the indicators. Our analysis yields to the identification of six distinctly different respondent types.

Looking at Table 5.5, we first remark the presence of two "extreme" profiles of respondents: class number 6 encompasses individuals who have negative and "deprived" responses on all the indicators. Individuals classified in this class have high risk to be in the first two income quartiles, to perceive difficulties to make ends meet with their income, to feel excluded from the society, to have low personal relationships (even if with a probability which is not too high), and to have a negative perception of the institutional system. This class groups the 11.2% of the population. The probability to answer in a "disadvantaged" manner is the highest for almost all the indicators. In the opposite situation, we find more than one-third of the population (37.9%): in class 2 individuals have a positive situation, that is high levels of income, good relationships with family, friends and neighbours, and a solid social network on which they could rely on in case of problems. Moreover, also from a subjective point of view, their situation is not problematic: they don't feel inferior to the others or excluded, and they judge positively their institutional environment in terms of social assistance, health services and security. Class 2, moreover, has the highest probability to participate in social leisure and sport associations. Thus, class 6 raises to be the "excluded class", and class 2 the "not excluded class".

An interesting characteristic in class 3 (size equal to 0.18) is the disagreeing between the objective measure of the income (income quartile) and the perception to get by with that income. Individuals classified in this class, even having a high probability to be in the lowest income quartiles, answer that their income is sufficient to make ends meet. On the other hand, this class has low probability to include people who feel unhelpful, marginalized or excluded, or people who is unsatisfied with the social and security system. It is worthwhile to note that about 50% of elderly people are classified in this class, as well as about one-quarter of students and homemakers. Thus, the low level of the income represents the unique "negative" element of this class. On the other hand, the low level of income seems to do not affect the capability of these individuals to integrate in the mainstream society and to feel overall satisfied.

The profile of class 1 identifies individuals who perceive, in a measure higher than the overall mean, the risk of social exclusion and the difficulty to have a useful role in the society. The economic situation of this class is on the average (both objective and subjective), but it seems that in this case the critical aspect is represented by the social relationships. People in this class have the highest probability to have low or very low social contacts with family, friends and neighbours, and, most important, they answer that they could not rely on anyone in case of problems. It seems to raise a situation in which the risk of marginalization and the feeling of social exclusion is not linked to a lack of economic stability, but rather to a lack of a stable and positive social network.

The two remaining classes detect intermediate situations. Class 4, for which the membership probability equal to 0.14, is characterized by high probability to be in the two lowest income quartiles and to perceive the income as not sufficient. Individuals in this class perceive problematic social relations in the sense that they feel a sense of inferiority with regards the others due to their income or their job situation, and feel to be left out of the society and to not to have an useful role in the society. In this case, the effect, even if not too high, is still relevant. Another critical dimension is represented by the perception to live in a violent and unsafe area, while in this class the health system is positively judged. Conversely, the social contacts of people in this class are medium or high. Finally, class 5 seems to identify mainly a situation of exclusion from what we called "institutional dimension": the probability to be dissatisfied with the social assistance and health care system is the highest, and also the assessment about the presence of violence and theft in the area. Moreover, people in this class tend to don't participate in association activities. Anyway, the other responses, about the economic situation, the social network and the subjective perception, identify not problematic conditions. Particularly, it seems that notwithstanding a negative "institutional dimension", the other areas of life are good. The probability to belong to this typology is the lowest (0.05).

Table 5.5 – Profile table of the latent variable at individual level X_{ij}, for the estimated model: size class $\hat{P}\left(X = t \middle| \mathbf{Z}_i\right)$ and class specific marginal probabilities $\hat{P}\left(Y_{ijk} = s_k \middle| X = t\right)$

	Latent classes for X_{ij}						
t	**1**	**2**	**3**	**4**	**5**	**6**	**Overall**
Class size $\hat{P}\left(X = t \middle| \mathbf{Z}_i\right)$	0.119	0.379	0.183	0.143	0.064	0.112	
Indicators Y_k							
Income perception							
with difficulties	0.043	0.043	0.116	0.794	0.325	0.973	0.286
without difficulties	0.957	0.957	0.884	0.206	0.675	0.027	0.714
Economic difficulties							
++ diff.	0.009	0.004	0.002	0.250	0.071	0.495	0.099
+ diff.	0.086	0.039	0.008	0.452	0.205	0.334	0.142
no diff.	0.904	0.956	0.990	0.298	0.724	0.171	0.759
Income quartiles							
-- (first quartile)	0.216	0.045	0.367	0.338	0.092	0.594	0.231
- (second quartile)	0.328	0.161	0.359	0.358	0.232	0.302	0.266
+ (third quartile)	0.276	0.315	0.193	0.210	0.325	0.085	0.248
++ (fourth quartile)	0.180	0.479	0.081	0.095	0.352	0.019	0.256
Feeling of inferiority							
Yes	0.119	0.049	0.035	0.218	0.081	0.240	0.103
don't know	0.333	0.078	0.084	0.175	0.115	0.215	0.141
No	0.548	0.873	0.881	0.607	0.804	0.545	0.756

(continue)

Table 5.5 (continued)

	Latent classes for X_{ij}						
t	**1**	**2**	**3**	**4**	**5**	**6**	**Overall**
Class size $\hat{P}(X=t\|\mathbf{z}_i)$	0.119	0.379	0.183	0.143	0.064	0.112	
Social contacts							
Very low	0.205	0.131	0.089	0.126	0.092	0.144	0.130
Low	0.359	0.306	0.260	0.302	0.262	0.299	0.300
Medium	0.271	0.311	0.326	0.313	0.328	0.298	0.309
High	0.165	0.252	0.326	0.259	0.318	0.259	0.261
Participation in assoc.							
No	0.790	0.566	0.670	0.685	0.867	0.924	0.688
Yes	0.210	0.435	0.330	0.315	0.133	0.076	0.312
Availability of help							
No	0.414	0.094	0.228	0.233	0.207	0.606	0.241
Yes	0.586	0.906	0.772	0.767	0.793	0.394	0.759
Feeling of social exclusion							
Yes	0.113	0.006	0.031	0.102	0.048	0.260	0.068
don't know	0.311	0.027	0.057	0.117	0.045	0.220	0.102
No	0.576	0.967	0.913	0.782	0.908	0.520	0.830
Feeling of usefulness							
Yes	0.206	0.025	0.106	0.146	0.070	0.378	0.121
don't know	0.407	0.066	0.177	0.189	0.122	0.285	0.173
No	0.387	0.909	0.717	0.666	0.808	0.338	0.707
Health services satisfact.							
Bad	0.165	0.072	0.054	0.109	0.653	0.401	0.159
Good	0.835	0.928	0.946	0.891	0.347	0.599	0.841
Social assistance satisfact.							
Bad	0.274	0.106	0.055	0.265	0.697	0.678	0.241
Good	0.726	0.895	0.945	0.735	0.303	0.322	0.759
Theft and violence							
Yes	0.267	0.158	0.199	0.324	0.336	0.361	0.236
don't know	0.466	0.155	0.130	0.223	0.232	0.275	0.216
No	0.267	0.687	0.671	0.453	0.432	0.364	0.548

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

Summarizing, we identify 6 latent levels of social exclusion, according to different domains of life. Whether we consider only the indicator "perception to be left out from society" as indicator of social exclusion situations, some of them are not properly situation of social exclusion: individuals in class 2, 3 and 5 do not perceive to be socially excluded. The low level of income (class 3) does not represent, per se, an element that influences negatively the perception of social marginalization. Class 5 identifies a typology that is dissatisfied with the social and protection system, but it seems that it does not affect the perception of social exclusion and social usefulness. Classes 4, 1 and 6 have high probabilities to include people that feel excluded (respectively 0.10, 0.11 and 0.26): class 6 identifies a typology of people with all negative indicators and thus excluded from all the dimensions; class 1 refers mainly to relational exclusion and class 4 to economic exclusion.

Let us now move on to the second level of the analysis. The choice of 4 latent levels for the variable W_j seems to operate quite well, providing a clear classification of regions.

As for the latent variable X_{ij}, we can obtain a global synthesis of the characteristics also for the latent classes of W_j. The profile Table 5.6 in this case shows the sizes of the high-level classes $\hat{P}\left(W_j = m\right)$. Since group-level covariates are present in model specification, $\hat{P}\left(W_j = m\right)$ are computed aggregating the model probabilities $\hat{P}\left(W_j = m \middle| \mathbf{Z}_j^g\right)$ over group-level covariates values:

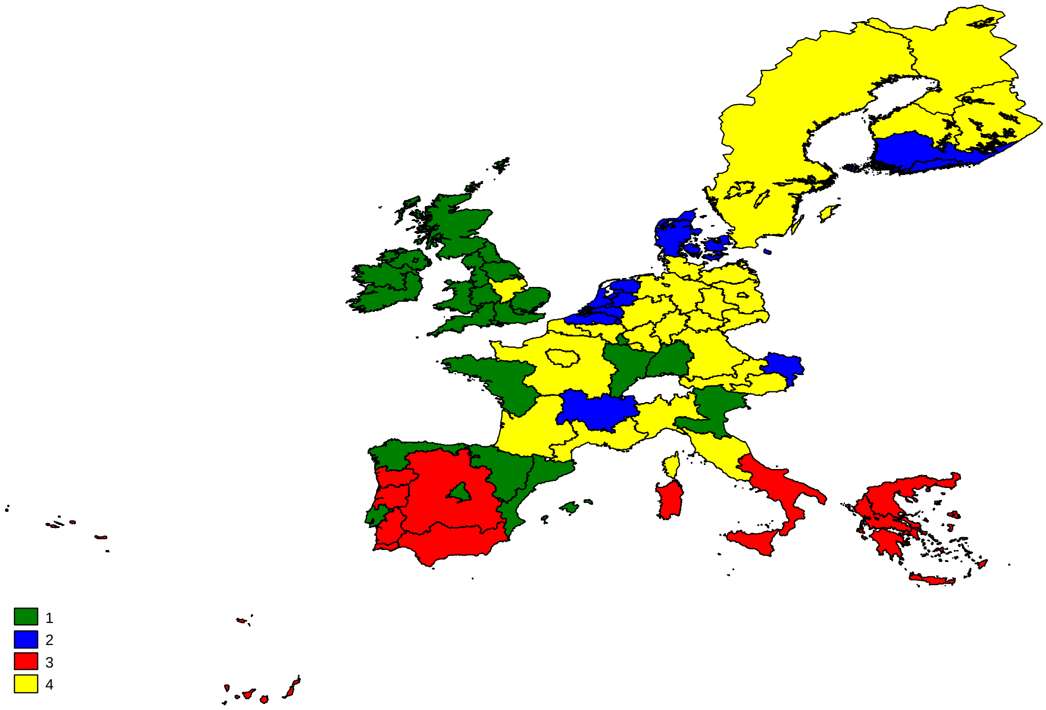
$$\hat{P}\left(W = m\right) = \sum_{j^*=1}^{J^*} \frac{w_{j^*}}{J} \hat{P}\left(W = m \middle| \mathbf{Z}_{j^*}^g\right) \qquad [5.9]$$

where w_{j^*} denotes the weight for a certain covariate pattern j^*. The region classification based on these cluster membership probabilities given only group-level covariates $\hat{P}\left(W_j = m \middle| \mathbf{Z}_j^g\right)$, is depicted in (Figure 5.8). These probabilities are sometimes referred to as *prior or model probabilities*.

Secondly, for each category of the indicators investigated, the profile Table 5.6 shows the group-region specific probabilities given the latent class, $\hat{P}\left(Y_{ijk} = s_k \middle| W_j = m\right)$. This probability is obtained marginalizing $P\left(X_{ij} = t, Y_{ijk} = t \middle| W_j = m\right) = P\left(X_{ij} = t \middle| W_j = m\right) P\left(Y_{ijk} = t \middle| X_{ij} = t\right)$ over X_{ij}, that is $\hat{P}\left(Y_{ijk} = s_k \middle| W_j = m\right) = \sum_{t=1}^{6} \left[P\left(X_{ij} = t \middle| W_j = m\right) P\left(Y_{ijk} = t \middle| X_{ij} = t\right) \right]$.

The first cluster groups together regions for which individuals don't seem to be in a disadvantaged condition. The probability to be in the "advantaged" level of the indicator are rather high for almost all the indicators investigated. The probability to be in the third and fourth income quartile, given the latent class, is higher than the average, as well as the probability to have an high level of social contacts and to feel satisfied with the social and protection system. There seems not to be a relevant presence of strong participation in social activities. In this cluster are classified almost all UK regions, Ireland, two German regions (Bremen and Baden-Wurttemberg), the North-East of Italy, the East and the West of France, and some Spanish regions (see map in Figure 5.8 and Table A.6 in Appendix A).

Figure 5.8 – Classification of the 77 European regions included in the analysis, on the basis of group-level prior membership probabilities given group-level covariates $P\left(W_j = m \middle| \mathbf{Z}_j^g\right)$.



Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

A positive situation is identified also for regions belonging to cluster 2: here people have low probabilities to have high levels of social contacts, but all other dimensions seem to go well. In this cluster, which size equals 0.13, we register the highest probability to participate in leisure, culture or sport activities and associations, and a high probability that people may rely on someone from outside their own household in case of problems. In these regions, we depict a sociability model according to which social contacts and social networks are mainly established via friends and organized activities, rather than to be family-centred. This situation is typical of North and continental European countries (e.g. Böhnke 2008); indeed, the regions classified with the highest probability in this cluster are the Dutch regions, Denmark, South-Finland, East-Austria and Centre-East France.

In the opposite situation it is the latent regional cluster 3, for which the probability to feel left out of society is the highest (0.13). The unique indicator in this class that is likely to assume a "positive" value with high probability is the indicator concerning the presence of high social and personal contacts with family, friends and neighbours. The probability that an individual classified in this class is in the lowest income quartiles is 0.56, and this is the only cluster where the subjective evaluation of the personal income has an high probability to be negative. Relevant probabilities are found also for "negative" responses concerning the institutional dimension. This class identifies regions where the social contacts are important, but the lack of potential support outside the one's own household undermines the individual perception of social integration. Through map in Figure 5.8, one sees that southern European regions (Greece, South and Islands of Italy, most Portuguese and Spanish regions), which are notoriously at lower level of GDP, are most likely to be in this cluster. This classification is consistent also with regard the relevance of social contacts with relatives, friends and neighbourhoods in these areas.

Finally, for the group-level cluster 4, which include most German and French regions, Austria, Sweden, North Finland, and North-West and Centre Italy, the most pronounced feature is that it groups regions where the probability to have low levels of income is higher than the average, but the individuals perceive their income to be sufficient to make ends meet. Except for the subjective evaluation of the social assistance and health system, the other indicators assume negative values with high probabilities, e.g. social contacts, availability of help, and subjective feeling of exclusion, usefulness and inferiority.

Table 5.6 – Profile table of the latent variable at regional level W_j , for the estimated model: size class $\hat{P}\left(W_j = m \middle| \mathbf{Z}_j^g\right)$ and class specific marginal probabilities $\hat{P}\left(Y_k = s_k \middle| W_j = m\right)$

	Latent classes for W_j				
m	**1**	**2**	**3**	**4**	**Overall**
Class size $\hat{P}\left(W_j = m \middle| \mathbf{Z}_j^g\right)$	0.333	0.130	0.158	0.379	
Indicators Y_k					
Income perception					
with difficulties	0.273	0.187	0.532	0.269	0.286
without difficulties	0.727	0.814	0.468	0.731	0.714
Economic difficulties					
++ diff.	0.089	0.048	0.222	0.094	0.099
+ diff.	0.142	0.093	0.232	0.140	0.142
no diff.	0.769	0.859	0.546	0.767	0.759
Income quartiles					
-- (first quartile)	0.201	0.204	0.301	0.246	0.231
- (second quartile)	0.251	0.258	0.268	0.282	0.266
+ (third quartile)	0.258	0.257	0.223	0.244	0.248
++ (fourth quartile)	0.291	0.282	0.209	0.228	0.256
Feeling of inferiority					
Yes	0.099	0.074	0.142	0.109	0.103
don't know	0.124	0.107	0.163	0.170	0.141
No	0.778	0.819	0.695	0.721	0.756

(continue)

Table 5.6 (continued)

	Latent classes for W_j				
m	**1**	**2**	**3**	**4**	**Overall**
Class size $\hat{P}\left(W_j = m \middle| \mathbf{Z}_j^g\right)$	0.333	0.130	0.158	0.379	
Indicators Y_k					
Social contacts					
very low	0.079	0.176	0.058	0.170	0.130
Low	0.249	0.349	0.214	0.342	0.300
Medium	0.332	0.290	0.332	0.293	0.309
High	0.340	0.185	0.397	0.195	0.261
Participation in assoc.					
No	0.719	0.529	0.876	0.694	0.688
Yes	0.281	0.471	0.124	0.306	0.312
Availability of help					
No	0.193	0.186	0.346	0.276	0.241
Yes	0.807	0.814	0.654	0.724	0.759
Feeling of social exclusion					
Yes	0.054	0.037	0.128	0.077	0.068
don't know	0.080	0.064	0.127	0.135	0.102
No	0.866	0.899	0.745	0.788	0.830
Feeling of usefulness					
Yes	0.098	0.082	0.192	0.138	0.121
don't know	0.144	0.136	0.199	0.210	0.173
No	0.758	0.782	0.609	0.652	0.707
Health services satisfact.					
Bad	0.119	0.083	0.437	0.136	0.159
Good	0.882	0.917	0.563	0.864	0.841
Social assistance satisfact.					
Bad	0.193	0.132	0.573	0.226	0.241
Good	0.807	0.868	0.427	0.775	0.759
Theft and violence					
Yes	0.221	0.202	0.320	0.240	0.236
don't know	0.194	0.172	0.250	0.249	0.216
No	0.585	0.626	0.430	0.512	0.548

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

Consider now the *posterior membership probabilities* for the group-level classes, i.e.

$$\hat{P}\left(W_{j*} = m \middle| \mathbf{Y}_{j*} = \mathbf{s}*, \mathbf{Z}_{j*}\right) = \frac{P\left(W_{j*} = m \middle| \mathbf{Z}_{j*}^{g}\right) \prod_{i=1}^{n_j} P\left(Y_{ij*} \middle| W_{j*} = m, \mathbf{Z}_{ij*}\right)}{P\left(\mathbf{Y}_{j*} = \mathbf{s}* \middle| \mathbf{Z}_{j*}\right)} \qquad [5.10]$$

For a given region j^*, with a given full vector of responses of all its individuals $\mathbf{Y}_{j^*}$ and the vector of their covariates $\mathbf{Z}_{j^*}$, equation [5.10] represents the probability of belonging to latent cluster m of W. In this case, the classification error is very low (0.01), and the misclassified cases are negligible (Table 5.7). Indeed, these membership probabilities are virtually indistinguishable from 0 or 1 for almost all regions (see Table A.6 reported in Appendix). The only exceptions are the region of East Anglia (UK) for which the probabilities to be classified in cluster 1, 2 and 4 are respectively 0.24, 0.29 and 0.47, and the region of Nord-Pas-de-Calais (FR) which is classified in cluster 1 with probability 0.46 and in cluster 4 with probability 0.54.

Reporting the modal classification given in Table 5.7 on a cartographic representation (Figure 5.9) may suggest interesting hints.

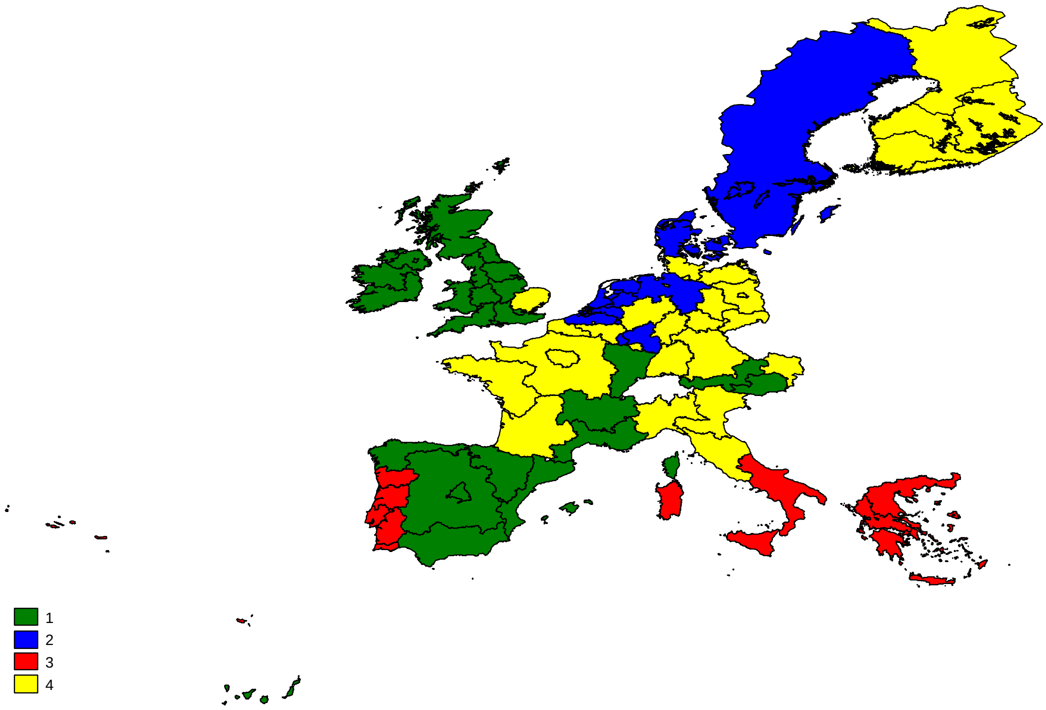
Table 5.7 – Classification table of regional latent variable W_j based on group-level posterior membership probabilities $\hat{P}\left(W_j = m \middle| \mathbf{Y}_j, \mathbf{Z}_j\right)$

Posterior cluster membership probability (m)	**Modal classification (m')**				
	1	**2**	**3**	**4**	**Total**
1	**25.0**	0.0	0.0	0.8	*25.8*
2	0.0	**9.6**	0.0	0.3	*9.9*
3	0.0	0.0	**12.0**	0.0	*12.0*
4	0.0	0.4	0.0	**28.9**	*29.3*
Total	*25.0*	*10.0*	*12.0*	*30.0*	***77.0***

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

Some Southern European regions (south of Italy, Greece and Portugal) are most likely to be classified in regional cluster 3 which, as described above, is characterized by negative conditions with respect to the individual income, the individual perception of social exclusion and low sense of usefulness and high disaffiliation with welfare system. However, it is also a cluster where the probability to have frequent social contacts with friends, family and neighbours is high, a result consistent with previous literature. All Spanish regions, the east, the centre-east and the Mediterranean area of France, Austria, Ireland and almost all regions of Great Britain are classified in cluster 1, which is characterized by high social relationship, as well a better situation for other indicators too. Conversely, we remember that clusters 2 and 4 are characterized by low level of social contacts. Modal posterior cluster membership probabilities place Sweden, Denmark, Luxembourg, The Netherlands and the Belgian Flanders in cluster 2, that is in a cluster characterized also, on average, by high probabilities to be satisfied with the social system, to have high levels of individual income, and low probabilities to feel left out by society. Finally, Finland, almost all German regions, South-East French regions and North-Centre Italy are classified in cluster 4, that seems to be somewhat disadvantaged. These results confirm literature about sociability and relational models across Europe, and allow confirming the validity of the indicators used and of the classification obtained.

Figure 5.9 – Classification of the 77 European regions included in the analysis, on the basis of the group-level posterior membership probabilities $\hat{P}\left(W_j = m \middle| \mathbf{Y}_j, \mathbf{Z}_j\right)$



Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

It is interesting to compare the two region classifications: model classification on the basis of group-level membership (or prior) probabilities $\hat{P}\left(W_j = m \middle| \mathbf{Z}_j^g\right)$ and posterior membership probabilities $\hat{P}\left(W_{j*} = m \middle| \mathbf{Y}_{j*}, \mathbf{Z}_{j*}\right)$ (Figure 5.8 and Figure 5.9, respectively). The former are obtained based on group-level covariates only (see [5.4]), while the latter are the probabilities given individual responses and covariates patterns (equation [5.10]).

We may thus remark that in some cases the individual responses change the probability to belong to a deprived cluster. For example, the Centre and the South of Spain, which are classified in the "deprived" cluster 3 according to prior probabilities, are classified to the cluster 1 according to posterior probabilities. In the same way, also Mediterranean France changes from cluster 4 to 1, and Sweden from 4 to 1. An opposite situation emerges for the region of Lisbon in Portugal (which passes from cluster 1 to cluster 3), and for the North-East of Italy (from cluster 1 to 4). This discrepancy between prior and posterior membership probabilities shows also that the group level covariates – even if useful in order to characterize the contextual environment – are not very good predictors of the cluster membership for the regions concerned.

Regional level clusters reflect, to some extent, individual typologies, even if in this case differences are less marked. Computing the probability of being in a certain latent class of X_{ij} for each level of W_j that is $P\left(X_{ij}=t \middle| W_j=m\right)$, we can quantify the influence of the level-one latent classes across level-two latent classes[49]. Table 5.8 presents model results linking the individual and the regional classes. Considering the relative size of individual-class within a region-cluster, we note that individual latent classes 2 and 3 (which are "not excluded classes") are highly present mainly in region-clusters 1 and 2 (which, in fact, are the less disadvantaged) and, although in a lower measure, also in region-cluster 4. Individual-level class 4 is present in all region-clusters, except the third. Indeed, classes 5 and 6 are prevalent in region-cluster 3.

Table 5.8 – Cross-tabulation of the probability of being in each latent class of X_{ij} for each level of W_j: $\hat{P}\left(X_{ij}=t \middle| W_j=m\right)$

		Latent cluster for W_j				**Marginal probabilities $\hat{P}(X=t)$**
		1	**2**	**3**	**4**	
Latent classes for X_{ij}	**1**	0.060	0.041	0.059	**0.241**	*0.119*
	2	**0.502**	**0.495**	0.085	**0.315**	*0.379*
	3	0.150	**0.309**	0.042	0.181	*0.183*
	4	0.194	0.130	0.035	0.152	*0.143*
	5	0.022	0.000	**0.409**	0.009	*0.064*
	6	0.072	0.025	**0.370**	0.103	*0.112*
		1.000	1.000	1.000	1.000	*1.000*

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

Substantially, Table 5.8 enables to compare latent class probabilities across the latent classes of W_j, highlighting the presence of different structures for the latent variable "social exclusion" across regions, depending on the effect of the discrete latent variable grouping the regions. In fact, while the probability to belong to the individual latent class 6 (the most disadvantaged one) equal 0.07 for regions belonging to cluster 1, it raises to 0.37 for regions belonging to cluster 3. Moreover, on the basis of results previously described (e.g. profile Table 5.6), whereas in certain regions (e.g. cluster 4) social exclusion situations are mainly linked to a lack of social networks, in other regions (e.g. cluster 3) the critical fac-

[49] These probabilities are obtained aggregating over covariates patterns.

tor is represented by poverty and dissatisfaction towards the social protection system.

Membership to individual latent classes, that is characterization of social exclusion situations, is often related to external variables describing the demographic and the socio-economic condition of individuals. Hence, the probability that an individual belongs to a particular latent class has been modelled to depend also on his socio-demographic characteristics (equation [5.3]). In this analysis, we considered the age of individuals (grouped in classes) and their occupational status. The effects of covariates depend on logit parameters of equation [5.5], but, in order to facilitate interpretation we present in Table 5.9 the class membership probability given each level of each covariate, $\hat{P}\left(X = t \middle| Z_p = z\right)$, computed aggregating over the categories of the other covariate, and over the latent variable at group level W_j.

Table 5.9 – Conditional probabilities of X_{ij} for individual level covariates: $\hat{P}\left(X_{ij} = t \middle| Z_p = z\right)$

	Latent classes for X_{ij}					
	1	**2**	**3**	**4**	**5**	**6**
Overall	0.119	0.379	0.183	0.143	0.064	0.112
Age						
15-24 years	0.104	0.395	0.157	0.221	0.076	0.046
25-39 years	0.103	0.483	0.038	0.190	0.101	0.086
40-54 years	0.112	0.507	0.047	0.149	0.058	0.127
55 years and over	0.145	0.187	0.421	0.061	0.031	0.155
Occupational status						
Employed	0.111	0.575	0.026	0.149	0.077	0.063
Homemaker	0.093	0.305	0.210	0.157	0.078	0.157
Unemployed	0.120	0.136	0.109	0.190	0.063	0.383
retired/unable	0.163	0.089	0.486	0.070	0.031	0.161
Student	0.085	0.357	0.251	0.239	0.066	0.002

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

The effect of age is relevant, particularly for some profiles. The class number 2 is overrepresented in the age groups 25-39 and 40-54, whereas the class 3 is overrepresented among people over 55 years. Class 4 is overrepresented among young people (15-24), and adult and elderly people have a higher presence of people belonging to class 6. Also the occupational status helps to predict the class membership probabilities. As expected, employed people present a high proportion of people in class 2, which included a positive individual condition for all the dimensions. Retired people have high probability to belong to class 3, characterized by a low level of income but, on the whole, a global satisfaction about the other aspects investigated. The unemployment rises the probability to be in the classes characterized by the higher risk of social exclusion, mainly in the economic dimension (classes 4 and 6). The status of homemaker has not significant effect on the membership probability to a given latent class.

Considering now also the group-level latent class membership, it is interesting to identify some hypothetical individuals based on covariate patterns, and to compare for each of them the probabilities to belong to a certain individual-level latent class t, given the membership to a certain cluster m. The hypothetical individuals we selected are: a) a retired person with more than 55 years; b) a 25-39 years old unemployed person; c) a 25-39 years old employed person. Figure 5.10 shows the latent class probabilities of such typical individuals for each regional cluster.

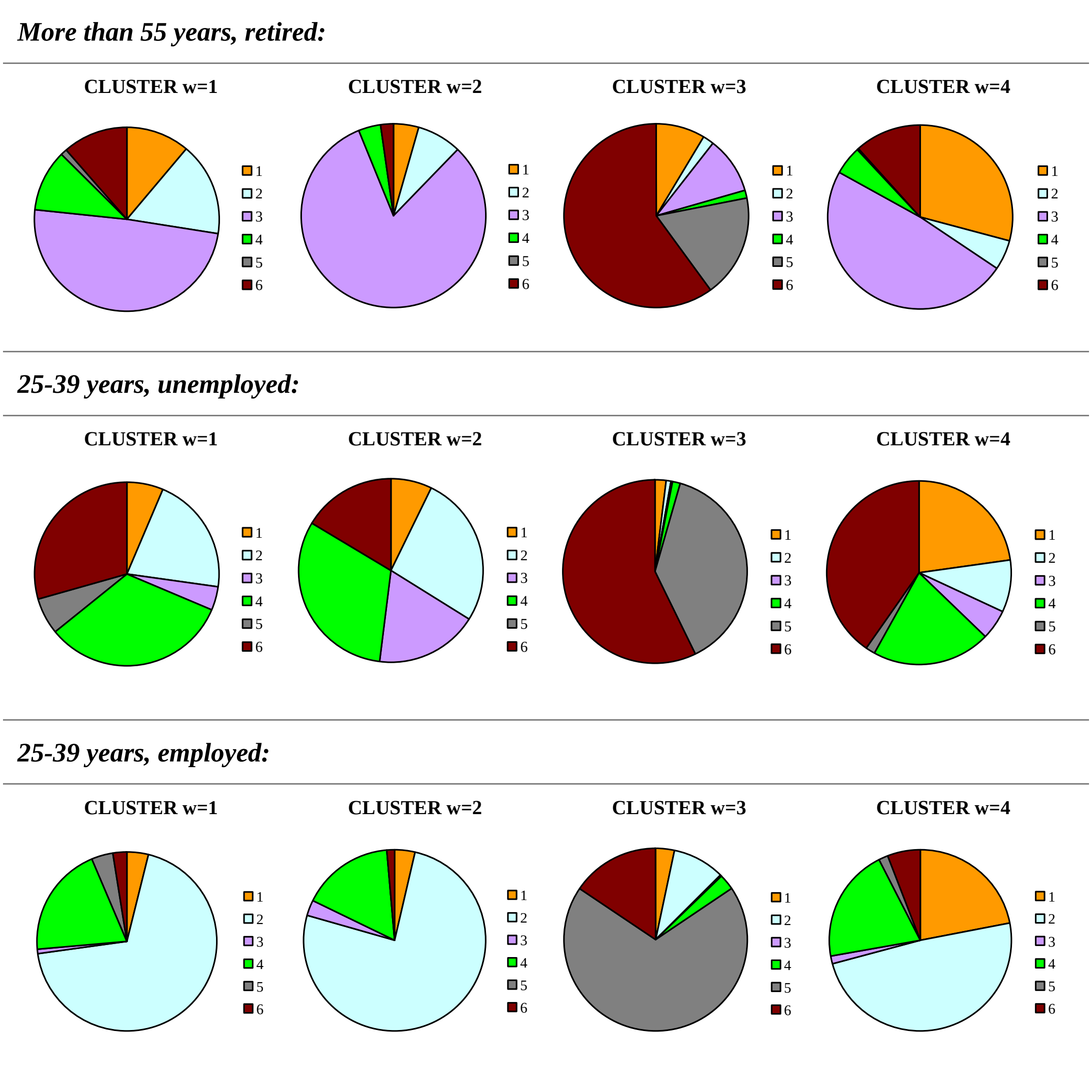
Bearing in mind that class 6 identifies the "excluded class", we see that the probability to belong to this class for a retired person passes from 0.02 for regional cluster 2 to 0.60 for cluster 3. Note that for an average region[50], this probability equal to 0.11. For clusters 1, 2 and 4, this hypothetical individual has high probabilities to belong to class 3 characterized by low levels of income but by a positive subjective evaluation of the overall conditions in all domains of one's own life.

For a young unemployed person the probability to belong to class 6 changes significantly according to the regional cluster membership, and also in this case the highest probability is given membership to cluster 3 (0.57). On the other hand, for these regions, such an individual has high probability to be classified also in class 5 (0.38 versus a probability of 0.03 for an average region), probably due to the presence of a strong attitude towards social relations and contacts that characterize regions in this group. Moreover, it is worthwhile noting that membership to the first two clusters reduces the risk of social exclusion. Except for cluster 4, the probability to belong to class 1 (relational exclusion) is low.

Considering a person of the same age (25-39) but employed, the probability to belong to the excluded class decreases drastically, regardless of the group-region, thus confirming the role of the employment against exclusion. Particularly, we remark an increase of the probability to belong to class 2, the "not excluded" one. Considering cluster 3, it remains a prevalence of individuals classified in class 5 (0.69), in which the lack of economic resources does not represent a problem, but where there is an high dissatisfaction with social and protection system. Conversely, given group-region 4, we remark that the probability to belong to class 1 – where the major element of exclusion is represented by a lack of social relations and a negative subjective evaluation of them – remains rather high (0.22). This feature is pronounced for all the three profiles of individuals identified.

[50] The average region is represented by a region in the second quartile of the ratio between level of taxes and primary income, in the third quartile of GDP, and in which from 30 to 40% of respondents attributes responsibility of poverty and social exclusion to individuals.

Figure 5.10 – Individual latent class probabilities conditional to regional latent class, for selected covariate patterns $\hat{P}\left(X=t \middle| W=m,\mathbf{Z}_{i*}\right)$



Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

In model [5.3] also the probability that a region belongs to a particular latent class is modelled to depend on its socio-economic characteristics. Table 5.10 shows the group-level class membership probabilities by the categories of each contextual covariate: the ratio between the amount of taxes, social contribution, transfers paid, and the primary income (grouped in quartiles); the GDP (in quartiles); a covariate indicating the percentage of people attributing the responsibility of poverty and social exclusion either to individual or to societal failure (cf. paragraph 5.2.3).

Referring to the first covariate in the table, it emerges that region-cluster 2 is overrepresented in high levels of taxes (++). Remember that this cluster groups together Sweden and regions of North Continental Europe, and that it has the lowest probabilities of social exclusion. Conversely, the lowest level of taxation (--), and thus a scarce improvement of social protection system, is mostly present in southern European regions of cluster 3 which, in turn, identifies the excluded typology of regions. Considering the GDP quartiles, the third cluster is strongly present in the lowest quartile of GDP (--), while the cluster number 2 is overrepresented in the higher levels of GDP (++). The medium-low level of GDP (-) presents a high proportion of the first cluster, while the medium-high (+) level of GDP has a high proportion of the fourth cluster. Consider, finally, the third covariate. Conditionally on tendency to consider injustice in society as responsible of poverty and social exclusion situations, cluster 4 is overrepresented, while cluster 1 and 3 are prevalent conditionally on personal responsibility of individuals.

Table 5.10 – Conditional probabilities of W_j for group level covariates: $\hat{P}\left(W = m \middle| Z_q^g = z\right)$

	Latent classes for W_j			
	1	**2**	**3**	**4**
Overall	0.333	0.130	0.158	0.379
Taxes/inc. (quart.)				
--	0.377	0.001	0.590	0.032
-	0.600	0.049	0.013	0.339
+	0.129	0.133	0.000	0.737
++	0.017	0.634	0.000	0.349
GDP (quart.)				
--	0.227	0.009	0.570	0.195
-	0.478	0.047	0.065	0.410
+	0.400	0.172	0.003	0.425
++	0.225	0.289	0.000	0.486
Individ. responsibility				
0-10	0.001	0.001	0.000	0.998
10-20	0.011	0.029	0.000	0.959
20-30	0.047	0.071	0.001	0.882
30-40	0.249	0.102	0.136	0.513
40-50	0.415	0.211	0.222	0.152
50-60	0.691	0.165	0.117	0.027
60-70	0.739	0.017	0.241	0.003
70-80	0.326	0.000	0.674	0.000

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

To further illustrate the effects of these group-level covariates, we computed cluster size probabilities given different patterns of regional covariates, i.e. $\hat{P}\left(W_j = m \middle| \mathbf{Z}_{j*}^g\right)$. Table 5.11 shows that with respect to probabilities for an average region (pattern $*$), a higher attitude to consider individuals as responsible for their situation of poverty and social exclusion (pattern A) increases the probability to belong to cluster 1, while a higher level of GDP (pattern B) increases the probability to be classified in cluster 4. Probability to belong to cluster 4 is high also for those regions having low levels of “individual responsibility” (pattern C). The probability to be classified in cluster 3 increases strongly for poorest regions, regardless of the third covariate (patterns D and E), while regions with high level of GDP and taxation (patterns F and G) are classified with high probability in cluster 2. A low taxation tends instead to increase the probability to belong to cluster 1 for rich regions (patterns H), and the probability to belong to cluster 3 for poor regions (patterns D and E).

Table 5.11 – Group level latent class probabilities $\hat{P}\left(W_j = m \middle| Z^g_{j*}\right)$ *for some patterns of group-level covariates*

	Regional covariate pattern						
	Taxes/inc. (quart.)	*GDP (quart.)*	*Individ. respons.*	**W=1**	**W=2**	**W=3**	**W=4**
pattern							
*	-	+	30-40	0.337	0.027	0.001	0.635
A	-	+	60-70	**0.940**	0.050	0.002	0.008
B	-	++	30-40	0.187	0.047	0.000	**0.766**
C	++	+	10-20	0.120	0.182	0.000	**0.974**
D	--	--	30-40	0.182	0.000	**0.794**	0.024
E	--	--	70-80	0.326	0.000	**0.674**	0.000
F	++	++	30-40	0.008	**0.694**	0.000	0.298
G	++	++	50-60	0.014	**0.971**	0.000	0.015
H	--	++	40-50	**0.809**	0.009	0.005	0.177

Note: $*$ indicates the pattern corresponding to the average region.

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

Chapter 6

Concluding remarks

During the recent years, there has been an increase in the size of the statistical literature on measuring *social exclusion* instances, as well as an increasing emphasis of public policies at European and national level. This is linked to the growing relevance around this problem, which undermines the capacity of citizens to actively participate in social life in its different forms. Social exclusion may imply a progressive lack of social cohesion with serious consequences on the socio-economic development of individuals and of society in its complex.

In this thesis, we focused on the study of social exclusion across European (EU 15) regions, with the ultimate aim of providing an operational conceptual and methodological framework, within which meaningful and useful analysis could be undertaken. We evaluated social exclusion issues from the individual point of view, in a multidimensional perspective and accounting for contextual environment in which people live.

To this purpose, we outlined an operational approach applicable to existing data, in terms of *conceptual model*, *indicators* and *statistical framework*. Starting from previous literature, we proposed a conceptual working model of social exclusion, which encompasses some founding elements of social exclusion notion, namely multidimensionality, subjectivity and relativity. While the relevance of these attributes of social exclusion is widely acknowledged in theoretical literature, we remark that they are somewhat neglected in current empirical research. In our conceptual approach, social exclusion is thus conceived as a multidimensional concept that encompasses different aspects of human life; that is, individuals may be excluded from different activities in their daily life. In this sense, we defined an economic, a social and an institutional dimension from which individuals might be excluded, identifying for each of them some relevant indicators. Social exclusion derives from the accumulation of different forms of deprivation and disadvantage in these dimensions. Clearly the economic, the social and the institutional dimensions are correlated each other.

On the basis of this understanding, and reckoning with current data availability, we selected some indicators referring to each dimension. The phase of indicator selection aimed to conjugate objective and subjective information in all the relevant dimensions.

Finally, we implemented a multilevel Latent Class model, which simultaneously derives regional and individual profiles. Firstly, Latent Class models represent a powerful statistical tool in presence of highly interrelated observed measures of a same underlying multidimensional latent concept. Secondly, the multilevel modelling enable to take into account the hierarchical structure of the population under investigation, and to carry on a comparative perspective. The two levels were modelled as interdependent: regions are grouped on the basis of the similarity between their within-region structure of the individual profiles.

The approach we proposed, which represents a new way to afford the problem of social exclusion, yields a number of interesting substantive insights, that usual analysis fail to pick up on.

The use of LCA allows treating social exclusion as a multidimensional concept thus underlying different types of exclusion, according to the different identified dimensions. It thus emerged that an individual might be excluded from the economic point of view, but not deprived in his social relationships; conversely, situations in which individuals suffer for weak social relations and interactions do not always go with a disadvantaged economic situation. The role of economic conditions in determining social exclusion situations seems thus to be reduced whether one considers in the analysis also the relational dimension. Clearly, our model identified also a profile for individuals who result excluded from all the dimensions, which represents the most serious situation, and a profile of individuals for whom the social exclusion does not represent a concrete threat.

The multilevel extension, particularly the choice to use a non-parametric approach to model the regional level, led to the identification of a typology of regions, underlying a different social exclusion structure for different groups of regions. It being understood the multidimensionality of the concept, it emerges that the importance of the different dimensions varies across regions. For some European areas, the condition of social exclusion is mainly due to the relational sphere, for certain ones to the economic dimension, and for other ones to the detachment with respect the institutions and the public context. Thereby, the different profiles of social exclusion are present within each group of regions in different proportions. Moreover, it seems that the social networks, as well as the effectiveness of the social and protection system, do not have the same impact in reducing the risk of social exclusion in all countries. The profiling of social exclusion situations is strongly related to demographic variables such as age and occupational status. Also in this case, however, the relevance of these covariates changes according to the region to which individual belongs. For example, the probability to be classified in the social excluded class for unemployed people varies across groups of regions. In this sense, the contextualization is fundamental to understand the relations among the risk factors that may trigger social exclusion situations. The comparative perspective of the analysis enabled to characterize various structures of the same latent concept.

Finally, the introduction of subjective elements in all the dimensions highlighted that negative objective situations are not always perceived in the same way. In particular, it is interesting the emerging of an individual profile within which the subjective perception of one's own economic situation does not correspond to the situation evaluated from an objective point of view. The subjective perception, in turn, may be affected by elements referring to other dimensions, and by the socio-economic and cultural context in which individuals live. This represents an interesting result, since at political and institutional level it is usual to consider only objective socio-economic indicators to evaluate and measure social exclusion situations.

Summarising, this thesis outlined and applied a conceptual and statistical methodology for the identification of individuals at high risk of social exclusion in EU member State regions, using the data of the Eurobarometer survey. The results suggest that the problem of social exclusion is not equally important across the European regions. According to different welfare and protection systems, the extent to which different risk factors (namely economic deprivation, social relations, family conditions, institutional involvement, and so on) affect social exclusion situations varies across regions. Moreover, also quite different sociability and familial models present across EU regions appear to influence differently the impact of the risk factors associated with individual social exclusion and the individual perception of these situations.

Therefore, we believe that policies and measures aiming at fighting social exclusion in the EU might be differentiated according to different social, cultural and economic contexts. A same policy is not likely to have a significant impact to reduce the social exclusion risk in all countries. Since major differences exist between some countries, and between some regions of the same country, they need to be accounted for when formulating policy strategies.

To conclude, a word about further potential developments of our research. We think that it could be interesting and relevant to extend this kind of analysis of social exclusion to the new member States of European Union. The recent admission in the EU of countries with different historical, cultural, social, and economic backgrounds is a result of a long work by the European Commission, but the differences cannot be rapidly filled. The socio-economic reality of an enlarged Europe (EU-27) is quite unbalanced today, due to a contrast between the western countries and the countries with a more or less marked situation of social vulnerability. In the EU-27, the pressure of inequalities is remarkably increasing for specific population segments. It is conceivable to hypothesize that the map of social and economic vulnerabilities will become more complex, due to the presence of new elements and profiles. However, a deeper understanding in this sense implies the necessity of most up-to-date and comprehensive datasets, from a point of view both geographical, temporal and substantial. Due to the largely recognized importance of social exclusion issues, we deem that a major effort towards data availability is needed in order to allow a meaningful and concrete monitoring of European regional differences in social exclusion situations.

Appendix A

Table A.1 – Social exclusion indicators used in the estimated model

Indicator	*Categories*	*Frequency*	*% missing*
ECONOMIC SITUATION			
Income perception	with difficulties	28.8	5.4
	without difficulties	71.2	
Economic difficulties	++ difficulties	9.6	6.7
	+ difficulties	13.8	
	no difficulty	76.6	
Income quartiles	-- (first quartile)	23.6	30.3
	- (second quartile)	26.7	
	+ (third quartile)	24.5	
	++ (fourth quartile)	25.2	
Feeling of inferiority	yes	10.2	0.0
	don't know	14.0	
	no	75.9	
SOCIAL RELATIONSHIPS			
Social contacts	very low	13.1	5.1
	low	30.0	
	medium	30.9	
	high	26.0	
Participation in associations	no	68.0	2.9
	yes	32.0	
Availability of help	no	24.0	0.0
	yes	76.0	
Feeling of social exclusion	yes	6.8	0.0
	don't know	10.0	
	no	83.2	
Feeling of usefulness	yes	12.0	0.0
	don't know	17.1	
	no	70.9	
INSTITUTIONAL DIMENSION			
Health services satisfaction	bad	16.2	4.2
	good	83.8	
Social assist. satisfaction	bad	24.5	12.4
	good	75.5	
Theft and violence	yes	23.6	0.0
	don't know	21.4	
	no	55.0	

Source: Our elaboration on Eurobarometer 56.1-2001 data.

Table A.2 – Individual level covariates used in the estimated model

Individual covariates $\mathbf{Z}_{ij}$	***Categories***	***Frequency***	***Source***
Sex	Male	47.7	Eurobarometer
	Female	52.3	
Classes of age	15-24	15.6	Eurobarometer
	25-39	28.0	
	40-54	24.0	
	>55	32.4	
Occupational status	employed	48.4	Eurobarometer
	homemaker	11.3	
	unemployed	6.7	
	retired/unable	23.4	
	student	10.2	

Source: Our elaboration on Eurobarometer 56.1-2001 data.

Table A.3 – Group level covariates used in the estimated model

Regional covariates $\mathbf{Z}_{j}^{g}$	***Description***	***Categories***	***Statistics***		***Source***
Individual responsibility	Why in your opinion are there people who live in need? (% of “because of laziness and lack of will-power”, and “because they have been unlucky”)		mean:	39.9	EB Q13
			sd:	11.1	
			min:	7.4	
			max:	71.4	
Taxes/income	% of taxes, social contribution and transfers paid, on primary income (recoded in quartiles)	-- (<39)	mean:	44.3	Eurostat
		- (40-44)	sd:	9.8	
		+ (45-49)	min:	21.9	
		++ (>50)	max:	61.5	
GDP	Gross Domestic product per inhabitant, in thousands of Euro (recoded in quartiles)	-- (<17.7)	mean:	25.1	Eurostat
		- (17.7-22.7)	sd:	9.0	
		+ (22.7-26.7)	min:	10.3	
		++ (>26.7)	max:	51.2	

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

Table A.4 – BIC values and number of parameters estimated for alternative models with different numbers of regional and individual classes M and T

	Regional level latent clusters													
	1		**2**		**3**		**4**		**5**		**6**		**7**	
Individual level latent classes	BIC(LL)	Npar	BIC(LL)	Npar	BIC(LL)	Npar	BIC(LL)	Npar	BIC(LL)	Npar	BIC(LL)	Npar	BIC(LL)	Npar
1	264,362.3	35	263,035.7	41	262,685.8	47	262,521.6	53	262,489.7	59	262,460.3	65	262,453.7	71
2	252,979.3	60	251,089.1	67	250,515.5	74	250,251.1	81	250,207.9	88	250,095.3	95	250,154.7	102
3	251,279.2	85	249,235.6	93	248,571.8	101	248,260.4	109	248,131.3	117	248,116.5	125	248,121.3	133
4	250,539.4	110	248,514.1	119	247,591.0	128	247,115.1	137	247,106.3	146	247,011.1	155	247,178.1	164
5	250,032.5	135	247,700.8	145	246,851.5	155	246,390.6	165	246,249.7	175	246,212.1	185	246,240.2	195
6	249,996.1	160	247,226.6	171	246,362.9	182	245,834.6	193	245,734.7	204	245,639.0	215	245,762.7	226
7	249,836.7	185	246,841.7	197	246,022.1	209	245,467.1	221	245,396.0	233	245,211.6	245	245,716.9	257
8	249,781.9	210	246,735.6	223	245,816.7	236	245,294.4	249	245,194.9	262	245,250.7	275	245,053.8	288
9	249,831.2	235	246,669.0	249	245,832.1	263	245,184.8	277	245,120.0	291	244,968.2	305	245,384.9	319
10	249,934.1	260	246,776.8	275	245,830.0	290	245,678.8	305	245,112.5	320	245,864.0	335	245,295.7	350

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

Table A.5 – Parameter estimates of the selected model

Regression parameters											
Term			**coef**	**s.e.**	**p-value**	**Wald(0)**	**df**	**p-value**	**Wald(=)**	**df**	**p-value**
Gregion(1)	<- 1		0.24	2.11	0.91	14.58	3	0.00			
Gregion(2)	<- 1		-11.08	3.64	0.00						
Gregion(3)	<- 1		7.33	3.10	0.02						
Gregion(4)	<- 1		3.51	2.34	0.13						
Gregion(1)	<- taxes_inc_quart		-0.33	0.59	0.58	12.05	3	0.007			
Gregion(2)	<- taxes_inc_quart		2.62	0.78	0.00						
Gregion(3)	<- taxes_inc_quart		-3.10	1.42	0.03						
Gregion(4)	<- taxes_inc_quart		0.80	0.61	0.19						
Gregion(1)	<- gdp_quart		0.05	0.37	0.89	7.46	3	0.059			
Gregion(2)	<- gdp_quart		1.19	0.56	0.03						
Gregion(3)	<- gdp_quart		-2.06	0.84	0.01						
Gregion(4)	<- gdp_quart		0.82	0.41	0.05						
Gregion(1)	<- individ_cl		0.53	0.32	0.10	10.64	3	0.014			
Gregion(2)	<- individ_cl		0.40	0.49	0.42						
Gregion(3)	<- individ_cl		0.35	0.43	0.42						
Gregion(4)	<- individ_cl		-1.27	0.42	0.00						
Cluster(1)	<- 1	| Gregion(1)	-0.31	0.15	0.04	1224.50	20	0.00	1040.93	15	0.00
Cluster(2)	<- 1	| Gregion(1)	1.52	0.11	0.00						
Cluster(3)	<- 1	| Gregion(1)	0.01	0.16	0.93						
Cluster(4)	<- 1	| Gregion(1)	1.00	0.12	0.00						
Cluster(5)	<- 1	| Gregion(1)	-1.45	0.28	0.00						
Cluster(6)	<- 1	| Gregion(1)	-0.77	0.32	0.01						
Cluster(1)	<- 1	| Gregion(2)	0.46	0.84	0.59						
Cluster(2)	<- 1	| Gregion(2)	2.42	0.82	0.00						
Cluster(3)	<- 1	| Gregion(2)	2.18	0.82	0.01						
Cluster(4)	<- 1	| Gregion(2)	1.61	0.82	0.05						
Cluster(5)	<- 1	| Gregion(2)	-5.97	4.07	0.14						
Cluster(6)	<- 1	| Gregion(2)	-0.70	0.89	0.43						
Cluster(1)	<- 1	| Gregion(3)	-0.29	0.19	0.13						
Cluster(2)	<- 1	| Gregion(3)	-0.35	0.18	0.06						
Cluster(3)	<- 1	| Gregion(3)	-1.30	0.24	0.00						
Cluster(4)	<- 1	| Gregion(3)	-0.83	0.33	0.01						
Cluster(5)	<- 1	| Gregion(3)	1.61	0.14	0.00						
Cluster(6)	<- 1	| Gregion(3)	1.16	0.29	0.00						
Cluster(1)	<- 1	| Gregion(4)	1.06	0.15	0.00						
Cluster(2)	<- 1	| Gregion(4)	0.81	0.14	0.00						
Cluster(3)	<- 1	| Gregion(4)	0.41	0.17	0.02						
Cluster(4)	<- 1	| Gregion(4)	0.64	0.15	0.00						
Cluster(5)	<- 1	| Gregion(4)	-2.57	0.53	0.00						
Cluster(6)	<- 1	| Gregion(4)	-0.36	0.32	0.26						
Cluster(1)	<- cleta4(15-24)		0.10	0.11	0.35	251.07	15	0.00			
Cluster(2)	<- cleta4(15-24)		-0.04	0.08	0.61						
Cluster(3)	<- cleta4(15-24)		-0.15	0.16	0.35						
Cluster(4)	<- cleta4(15-24)		0.34	0.10	0.00						
Cluster(5)	<- cleta4(15-24)		0.11	0.17	0.50						
Cluster(6)	<- cleta4(15-24)		-0.36	0.12	0.00						
Cluster(1)	<- cleta4(25-39)		-0.13	0.08	0.08						
Cluster(2)	<- cleta4(25-39)		-0.01	0.06	0.90						
Cluster(3)	<- cleta4(25-39)		-0.70	0.14	0.00						
Cluster(4)	<- cleta4(25-39)		0.28	0.07	0.00						
Cluster(5)	<- cleta4(25-39)		0.67	0.11	0.00						
Cluster(6)	<- cleta4(25-39)		-0.11	0.07	0.13						
Cluster(1)	<- cleta4(40-54)		-0.06	0.08	0.40						
Cluster(2)	<- cleta4(40-54)		0.11	0.06	0.05						
Cluster(3)	<- cleta4(40-54)		-0.37	0.15	0.01						
Cluster(4)	<- cleta4(40-54)		0.10	0.07	0.20						

Cluster(5)	<- cleta4(40-54)	-0.02	0.12	0.83			
Cluster(6)	<- cleta4(40-54)	0.25	0.07	0.00			
Cluster(1)	<- cleta4(>55)	0.10	0.09	0.29			
Cluster(2)	<- cleta4(>55)	-0.07	0.07	0.36			
Cluster(3)	<- cleta4(>55)	1.22	0.13	0.00			
Cluster(4)	<- cleta4(>55)	-0.71	0.11	0.00			
Cluster(5)	<- cleta4(>55)	-0.76	0.14	0.00			
Cluster(6)	<- cleta4(>55)	0.23	0.08	0.01			
Cluster(1)	<- occup2(empl.)	0.05	0.10	0.63	783.62	20	0.00
Cluster(2)	<- occup2(empl.)	0.98	0.09	0.00			
Cluster(3)	<- occup2(empl.)	-1.46	0.20	0.00			
Cluster(4)	<- occup2(empl.)	-0.03	0.11	0.82			
Cluster(5)	<- occup2(empl.)	0.35	0.14	0.01			
Cluster(6)	<- occup2(empl.)	0.10	0.29	0.72			
Cluster(1)	<- occup2(homaker)	-0.24	0.13	0.06			
Cluster(2)	<- occup2(homaker)	0.00	0.10	0.99			
Cluster(3)	<- occup2(homaker)	0.08	0.13	0.52			
Cluster(4)	<- occup2(homaker)	-0.15	0.12	0.20			
Cluster(5)	<- occup2(homaker)	-0.16	0.16	0.31			
Cluster(6)	<- occup2(homaker)	0.47	0.29	0.11			
Cluster(1)	<- occup2(disempl.)	-0.19	0.14	0.16			
Cluster(2)	<- occup2(disempl.)	-0.98	0.13	0.00			
Cluster(3)	<- occup2(disempl.)	-0.43	0.18	0.02			
Cluster(4)	<- occup2(disempl.)	-0.29	0.12	0.02			
Cluster(5)	<- occup2(disempl.)	0.13	0.18	0.48			
Cluster(6)	<- occup2(disempl.)	1.77	0.30	0.00			
Cluster(1)	<- occup2(ret./unab.)	0.28	0.12	0.02			
Cluster(2)	<- occup2(ret./unab.)	-0.99	0.14	0.00			
Cluster(3)	<- occup2(ret./unab.)	0.32	0.12	0.01			
Cluster(4)	<- occup2(ret./unab.)	-0.24	0.14	0.09			
Cluster(5)	<- occup2(ret./unab.)	-0.02	0.18	0.92			
Cluster(6)	<- occup2(ret./unab.)	0.65	0.29	0.03			
Cluster(1)	<- occup2(stud.)	0.10	0.27	0.71			
Cluster(2)	<- occup2(stud.)	0.99	0.26	0.00			
Cluster(3)	<- occup2(stud.)	1.49	0.31	0.00			
Cluster(4)	<- occup2(stud.)	0.70	0.27	0.01			
Cluster(5)	<- occup2(stud.)	-0.30	0.31	0.33			
Cluster(6)	<- occup2(stud.)	-2.99	1.13	0.01			
perc_inc(with diff.)	<- 1	-0.34	0.04	0.00	64.10	1	0.00
perc_inc(no diff.)	<- 1	0.34	0.04	0.00			
perc_inc(with diff.)	<- Cluster(1)	-1.21	0.15	0.00	1271.49	5	0.00
perc_inc(no diff.)	<- Cluster(1)	1.21	0.15	0.00			
perc_inc(with diff.)	<- Cluster(2)	-1.21	0.06	0.00			
perc_inc(no diff.)	<- Cluster(2)	1.21	0.06	0.00			
perc_inc(with diff.)	<- Cluster(3)	-0.68	0.06	0.00			
perc_inc(no diff.)	<- Cluster(3)	0.68	0.06	0.00			
perc_inc(with diff.)	<- Cluster(4)	1.01	0.07	0.00			
perc_inc(no diff.)	<- Cluster(4)	-1.01	0.07	0.00			
perc_inc(with diff.)	<- Cluster(5)	-0.03	0.07	0.66			
perc_inc(no diff.)	<- Cluster(5)	0.03	0.07	0.66			
perc_inc(with diff.)	<- Cluster(6)	2.12	0.11	0.00			
perc_inc(no diff.)	<- Cluster(6)	-2.12	0.11	0.00			
eco_diff2(++ diff.)	<- 1	-1.39	0.12	0.00	544.11	2	0.00
eco_diff2(+ diff.)	<- 1	-0.18	0.08	0.02			
eco_diff2(no diff.)	<- 1	1.57	0.07	0.00			
eco_diff2(++ diff.)	<- Cluster(1)	-0.87	0.32	0.01	1553.26	10	7.9e-328
eco_diff2(+ diff.)	<- Cluster(1)	0.14	0.18	0.45			
eco_diff2(no diff.)	<- Cluster(1)	0.73	0.18	0.00			
eco_diff2(++ diff.)	<- Cluster(2)	-1.13	0.23	0.00			
eco_diff2(+ diff.)	<- Cluster(2)	-0.16	0.14	0.26			
eco_diff2(no diff.)	<- Cluster(2)	1.28	0.13	0.00			
eco_diff2(++ diff.)	<- Cluster(3)	-1.27	0.54	0.02			
eco_diff2(+ diff.)	<- Cluster(3)	-0.89	0.37	0.02			

eco_diff2(no diff.)	<- Cluster(3)	2.16	0.31	0.00			
eco_diff2(++ diff.)	<- Cluster(4)	1.14	0.13	0.00			
eco_diff2(+ diff.)	<- Cluster(4)	0.52	0.09	0.00			
eco_diff2(no diff.)	<- Cluster(4)	-1.65	0.09	0.00			
eco_diff2(++ diff.)	<- Cluster(5)	0.26	0.16	0.11			
eco_diff2(+ diff.)	<- Cluster(5)	0.12	0.10	0.27			
eco_diff2(no diff.)	<- Cluster(5)	-0.38	0.11	0.00			
eco_diff2(++ diff.)	<- Cluster(6)	1.88	0.13	0.00			
eco_diff2(+ diff.)	<- Cluster(6)	0.27	0.09	0.00			
eco_diff2(no diff.)	<- Cluster(6)	-2.15	0.09	0.00			
quart_inc(--)	<- 1	0.02	0.03	0.59	298.97	3	0.00
quart_inc(-)	<- 1	0.34	0.02	0.00			
quart_inc(+)	<- 1	0.08	0.02	0.00			
quart_inc(++)	<- 1	-0.44	0.03	0.00			
quart_inc	<- Cluster(1)	0.09	0.05	0.04	1060.96	5	0.00
quart_inc	<- Cluster(2)	0.94	0.03	0.00			
quart_inc	<- Cluster(3)	-0.35	0.05	0.00			
quart_inc	<- Cluster(4)	-0.27	0.04	0.00			
quart_inc	<- Cluster(5)	0.60	0.05	0.00			
quart_inc	<- Cluster(6)	-1.00	0.05	0.00			
inferior(yes)	<- 1	-0.47	0.03	0.00	743.24	2	0.00
inferior(dk)	<- 1	-0.22	0.03	0.00			
inferior(no)	<- 1	0.69	0.03	0.00			
inferior(yes)	<- Cluster(1)	-0.16	0.08	0.05	358.95	10	0.00
inferior(dk)	<- Cluster(1)	0.36	0.07	0.00			
inferior(no)	<- Cluster(1)	-0.19	0.06	0.00			
inferior(yes)	<- Cluster(2)	-0.10	0.06	0.12			
inferior(dk)	<- Cluster(2)	-0.14	0.05	0.01			
inferior(no)	<- Cluster(2)	0.23	0.04	0.00			
inferior(yes)	<- Cluster(3)	-0.51	0.10	0.00			
inferior(dk)	<- Cluster(3)	-0.08	0.08	0.31			
inferior(no)	<- Cluster(3)	0.58	0.06	0.00			
inferior(yes)	<- Cluster(4)	0.54	0.05	0.00			
inferior(dk)	<- Cluster(4)	-0.07	0.05	0.18			
inferior(no)	<- Cluster(4)	-0.47	0.04	0.00			
inferior(yes)	<- Cluster(5)	0.01	0.09	0.89			
inferior(dk)	<- Cluster(5)	-0.05	0.08	0.58			
inferior(no)	<- Cluster(5)	0.03	0.07	0.61			
inferior(yes)	<- Cluster(6)	0.21	0.06	0.00			
inferior(dk)	<- Cluster(6)	-0.03	0.05	0.61			
inferior(no)	<- Cluster(6)	-0.18	0.05	0.00			
rel1(veri low)	<- 1	-0.62	0.02	0.00	1201.35	3	0.00
rel1(low)	<- 1	0.30	0.01	0.00			
rel1(medium)	<- 1	0.31	0.01	0.00			
rel1(high)	<- 1	0.02	0.02	0.43			
rel1	<- Cluster(1)	-0.13	0.04	0.00	163.04	5	0.00
rel1	<- Cluster(2)	0.08	0.03	0.00			
rel1	<- Cluster(3)	0.41	0.03	0.00			
rel1	<- Cluster(4)	0.12	0.04	0.00			
rel1	<- Cluster(5)	-0.35	0.06	0.00			
rel1	<- Cluster(6)	-0.14	0.04	0.00			
rel1	<- Gregion(1)	0.13	0.02	0.00	676.43	3	0.00
rel1	<- Gregion(2)	-0.44	0.02	0.00			
rel1	<- Gregion(3)	0.60	0.04	0.00			
rel1	<- Gregion(4)	-0.28	0.02	0.00			
assoc(no)	<- 1	0.61	0.02	0.00	1181.00	1	0.00
assoc(yes)	<- 1	-0.61	0.02	0.00			
assoc(no)	<- Cluster(1)	0.08	0.05	0.11	290.89	5	0.00
assoc(yes)	<- Cluster(1)	-0.08	0.05	0.11			
assoc(no)	<- Cluster(2)	-0.43	0.03	0.00			
assoc(yes)	<- Cluster(2)	0.43	0.03	0.00			

assoc(no)	<- Cluster(3)	-0.16	0.04	0.00			
assoc(yes)	<- Cluster(3)	0.16	0.04	0.00			
assoc(no)	<- Cluster(4)	-0.19	0.04	0.00			
assoc(yes)	<- Cluster(4)	0.19	0.04	0.00			
assoc(no)	<- Cluster(5)	0.13	0.09	0.17			
assoc(yes)	<- Cluster(5)	-0.13	0.09	0.17			
assoc(no)	<- Cluster(6)	0.57	0.06	0.00			
assoc(yes)	<- Cluster(6)	-0.57	0.06	0.00			
assoc(no)	<- Gregion(1)	0.12	0.02	0.00	269.52	3	0.00
assoc(yes)	<- Gregion(1)	-0.12	0.02	0.00			
assoc(no)	<- Gregion(2)	-0.28	0.02	0.00			
assoc(yes)	<- Gregion(2)	0.28	0.02	0.00			
assoc(no)	<- Gregion(3)	0.22	0.05	0.00			
assoc(yes)	<- Gregion(3)	-0.22	0.05	0.00			
assoc(no)	<- Gregion(4)	-0.05	0.02	0.01			
assoc(yes)	<- Gregion(4)	0.05	0.02	0.01			
help(no)	<- 1	-0.32	0.02	0.00	186.16	1	0.00
help(yes)	<- 1	0.32	0.02	0.00			
help(no)	<- Cluster(1)	0.29	0.04	0.00	1017.37	5	0.00
help(yes)	<- Cluster(1)	-0.29	0.04	0.00			
help(no)	<- Cluster(2)	-0.65	0.03	0.00			
help(yes)	<- Cluster(2)	0.65	0.03	0.00			
help(no)	<- Cluster(3)	-0.10	0.03	0.00			
help(yes)	<- Cluster(3)	0.10	0.03	0.00			
help(no)	<- Cluster(4)	-0.10	0.04	0.01			
help(yes)	<- Cluster(4)	0.10	0.04	0.01			
help(no)	<- Cluster(5)	-0.16	0.05	0.00			
help(yes)	<- Cluster(5)	0.16	0.05	0.00			
help(no)	<- Cluster(6)	0.71	0.03	0.00			
help(yes)	<- Cluster(6)	-0.71	0.03	0.00			
escl(yes)	<- 1	-0.80	0.04	0.00	1448.34	2	3.1e-315
escl(dk)	<- 1	-0.34	0.04	0.00			
escl(no)	<- 1	1.14	0.03	1.3e-316			
escl(yes)	<- Cluster(1)	0.08	0.09	0.36	412.53	10	0.00
escl(dk)	<- Cluster(1)	0.47	0.07	0.00			
escl(no)	<- Cluster(1)	-0.56	0.07	0.00			
escl(yes)	<- Cluster(2)	-0.86	0.16	0.00			
escl(dk)	<- Cluster(2)	-0.02	0.11	0.81			
escl(no)	<- Cluster(2)	0.88	0.09	0.00			
escl(yes)	<- Cluster(3)	-0.18	0.11	0.09			
escl(dk)	<- Cluster(3)	-0.14	0.09	0.12			
escl(no)	<- Cluster(3)	0.32	0.07	0.00			
escl(yes)	<- Cluster(4)	0.18	0.08	0.02			
escl(dk)	<- Cluster(4)	-0.03	0.07	0.65			
escl(no)	<- Cluster(4)	-0.15	0.06	0.01			
escl(yes)	<- Cluster(5)	0.20	0.13	0.12			
escl(dk)	<- Cluster(5)	-0.41	0.13	0.00			
escl(no)	<- Cluster(5)	0.21	0.09	0.02			
escl(yes)	<- Cluster(6)	0.58	0.06	0.00			
escl(dk)	<- Cluster(6)	0.13	0.06	0.02			
escl(no)	<- Cluster(6)	-0.71	0.05	0.00			
useful(yes)	<- 1	-0.26	0.03	0.00	81.86	2	0.00
useful(dk)	<- 1	0.00	0.03	0.95			
useful(no)	<- 1	0.26	0.03	0.00			
useful(yes)	<- Cluster(1)	0.22	0.07	0.00	621.03	10	0.00
useful(dk)	<- Cluster(1)	0.33	0.06	0.00			
useful(no)	<- Cluster(1)	-0.55	0.06	0.00			
useful(yes)	<- Cluster(2)	-0.63	0.08	0.00			
useful(dk)	<- Cluster(2)	-0.23	0.06	0.00			
useful(no)	<- Cluster(2)	0.86	0.05	0.00			
useful(yes)	<- Cluster(3)	0.06	0.06	0.28			
useful(dk)	<- Cluster(3)	0.03	0.05	0.51			

useful(no)	<- Cluster(3)	-0.10	0.04	0.03			
useful(yes)	<- Cluster(4)	-0.03	0.06	0.66			
useful(dk)	<- Cluster(4)	-0.08	0.05	0.16			
useful(no)	<- Cluster(4)	0.10	0.05	0.04			
useful(yes)	<- Cluster(5)	-0.24	0.11	0.02			
useful(dk)	<- Cluster(5)	-0.13	0.09	0.13			
useful(no)	<- Cluster(5)	0.37	0.07	0.00			
useful(yes)	<- Cluster(6)	0.62	0.05	0.00			
useful(dk)	<- Cluster(6)	0.07	0.05	0.13			
useful(no)	<- Cluster(6)	-0.69	0.05	0.00			
health_serv(bad)	<- 1	-0.67	0.02	0.00	1332.84	1	0.00
health_serv(good)	<- 1	0.67	0.02	0.00			
health_serv(bad)	<- Cluster(1)	-0.05	0.05	0.35	387.98	5	0.00
health_serv(good)	<- Cluster(1)	0.05	0.05	0.35			
health_serv(bad)	<- Cluster(2)	-0.34	0.04	0.00			
health_serv(good)	<- Cluster(2)	0.34	0.04	0.00			
health_serv(bad)	<- Cluster(3)	-0.42	0.06	0.00			
health_serv(good)	<- Cluster(3)	0.42	0.06	0.00			
health_serv(bad)	<- Cluster(4)	-0.30	0.05	0.00			
health_serv(good)	<- Cluster(4)	0.30	0.05	0.00			
health_serv(bad)	<- Cluster(5)	0.85	0.05	0.00			
health_serv(good)	<- Cluster(5)	-0.85	0.05	0.00			
health_serv(bad)	<- Cluster(6)	0.26	0.04	0.00			
health_serv(good)	<- Cluster(6)	-0.26	0.04	0.00			
soc_ass(bad)	<- 1	-0.22	0.02	0.00	123.40	1	0.00
soc_ass(good)	<- 1	0.22	0.02	0.00			
soc_ass(bad)	<- Cluster(1)	0.00	0.04	1.00	985.05	5	0.00
soc_ass(good)	<- Cluster(1)	0.00	0.04	1.00			
soc_ass(bad)	<- Cluster(2)	-0.51	0.03	0.00			
soc_ass(good)	<- Cluster(2)	0.51	0.03	0.00			
soc_ass(bad)	<- Cluster(3)	-0.85	0.06	0.00			
soc_ass(good)	<- Cluster(3)	0.85	0.06	0.00			
soc_ass(bad)	<- Cluster(4)	0.04	0.04	0.33			
soc_ass(good)	<- Cluster(4)	-0.04	0.04	0.33			
soc_ass(bad)	<- Cluster(5)	0.56	0.05	0.00			
soc_ass(good)	<- Cluster(5)	-0.56	0.05	0.00			
soc_ass(bad)	<- Cluster(6)	0.76	0.04	0.00			
soc_ass(good)	<- Cluster(6)	-0.76	0.04	0.00			
theft(yes)	<- 1	-0.13	0.02	0.00	576.53	2	0.00
theft(dk)	<- 1	-0.28	0.02	0.00			
theft(no)	<- 1	0.41	0.02	0.00			
theft(yes)	<- Cluster(1)	-0.05	0.06	0.34	922.46	10	0.00
theft(dk)	<- Cluster(1)	0.65	0.05	0.00			
theft(no)	<- Cluster(1)	-0.60	0.06	0.00			
theft(yes)	<- Cluster(2)	-0.35	0.03	0.00			
theft(dk)	<- Cluster(2)	-0.22	0.03	0.00			
theft(no)	<- Cluster(2)	0.57	0.03	0.00			
theft(yes)	<- Cluster(3)	-0.13	0.05	0.00			
theft(dk)	<- Cluster(3)	-0.41	0.05	0.00			
theft(no)	<- Cluster(3)	0.54	0.04	0.00			
theft(yes)	<- Cluster(4)	0.14	0.04	0.00			
theft(dk)	<- Cluster(4)	-0.08	0.05	0.09			
theft(no)	<- Cluster(4)	-0.07	0.04	0.09			
theft(yes)	<- Cluster(5)	0.17	0.05	0.00			
theft(dk)	<- Cluster(5)	-0.05	0.06	0.39			
theft(no)	<- Cluster(5)	-0.12	0.05	0.01			
theft(yes)	<- Cluster(6)	0.22	0.04	0.00			
theft(dk)	<- Cluster(6)	0.10	0.04	0.02			
theft(no)	<- Cluster(6)	-0.32	0.04	0.00			

Associations												
Term			**coef**	**s.e.**	**p-value**	**Wald(0)**	**df**	**p-value**	**Wald(=)**	**df**	**p-value**
inferior(yes)	<->	useful(yes)	0.41	0.03	0.00	569.40	4	0.00			
inferior(yes)	<->	useful(dk)	-0.19	0.03	0.00						
inferior(yes)	<->	useful(no)	-0.22	0.03	0.00						
inferior(dk)	<->	useful(yes)	-0.19	0.03	0.00						
inferior(dk)	<->	useful(dk)	0.41	0.03	0.00						
inferior(dk)	<->	useful(no)	-0.22	0.03	0.00						
inferior(no)	<->	useful(yes)	-0.22	0.03	0.00						
inferior(no)	<->	useful(dk)	-0.22	0.03	0.00						
inferior(no)	<->	useful(no)	0.44	0.02	0.00						
help(no)	<->	rel1	-0.11	0.01	0.00	85.76	1	0.00			
help(yes)	<->	rel1	0.11	0.01	0.00						
escl(yes)	<->	useful(yes)	0.56	0.04	0.00	821.68	4	0.00			
escl(yes)	<->	useful(dk)	-0.14	0.04	0.00						
escl(yes)	<->	useful(no)	-0.42	0.04	0.00						
escl(dk)	<->	useful(yes)	-0.06	0.03	0.08						
escl(dk)	<->	useful(dk)	0.35	0.03	0.00						
escl(dk)	<->	useful(no)	-0.29	0.04	0.00						
escl(no)	<->	useful(yes)	-0.50	0.03	0.00						
escl(no)	<->	useful(dk)	-0.21	0.03	0.00						
escl(no)	<->	useful(no)	0.71	0.03	0.00						
escl(yes)	<->	inferior(yes)	0.49	0.04	0.00	545.55	4	0.00			
escl(dk)	<->	inferior(yes)	-0.13	0.04	0.00						
escl(no)	<->	inferior(yes)	-0.36	0.03	0.00						
escl(yes)	<->	inferior(dk)	-0.19	0.04	0.00						
escl(dk)	<->	inferior(dk)	0.35	0.03	0.00						
escl(no)	<->	inferior(dk)	-0.16	0.03	0.00						
escl(yes)	<->	inferior(no)	-0.30	0.04	0.00						
escl(dk)	<->	inferior(no)	-0.22	0.03	0.00						
escl(no)	<->	inferior(no)	0.52	0.03	0.00						
health_serv(bad)	<->	soc_ass(bad)	0.45	0.02	0.00	686.72	1	0.00			
health_serv(bad)	<->	soc_ass(good)	-0.45	0.02	0.00						
health_serv(good)	<->	soc_ass(bad)	-0.45	0.02	0.00						
health_serv(good)	<->	soc_ass(good)	0.45	0.02	0.00						

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

Table A.6 – Prior cluster membership probabilities $\hat{P}\left(W_j = m \middle| \mathbf{Z}_j\right)$, and posterior cluster membership probabilities $\hat{P}\left(W_j = m \middle| \mathbf{Y}_j, \mathbf{Z}_j\right)$ with the respective modal assignment for the 77 regions included in the analysis

		Prior probabilities $\hat{P}\left(W_j = m \middle| \mathbf{Z}_j\right)$					**Posterior probabilities** $\hat{P}\left(W_j = m \middle| \mathbf{Y}_j, \mathbf{Z}_j\right)$				
European regions	**N. resp.**	**1**	**2**	**3**	**4**	**Modal**	**1**	**2**	**3**	**4**	**Modal**
AUS - east	433	0.14	0.57	0.00	0.29	**2**	0.00	0.00	0.00	1.00	**4**
AUS - south	225	0.12	0.18	0.00	0.70	**4**	1.00	0.00	0.00	0.00	**1**
AUS - west	342	0.19	0.05	0.00	0.77	**4**	1.00	0.00	0.00	0.00	**1**
BEL - Brussels	100	0.01	0.07	0.00	0.92	**4**	0.00	0.00	0.00	1.00	**4**
BEL - north (Flanders)	590	0.30	0.40	0.00	0.29	**2**	0.00	1.00	0.00	0.00	**2**
BEL - south (Walloon)	342	0.24	0.12	0.00	0.64	**4**	0.00	0.00	0.00	1.00	**4**
GER - Baden-Wurttemberg	144	0.53	0.12	0.00	0.36	**1**	0.00	0.00	0.00	1.00	**4**
GER - Bayern	192	0.19	0.05	0.00	0.77	**4**	0.00	0.00	0.00	1.00	**4**
GER - Berlin	128	0.00	0.05	0.00	0.94	**4**	0.00	0.00	0.00	1.00	**4**
GER - Brandenburg	168	0.11	0.02	0.00	0.86	**4**	0.00	0.00	0.00	1.00	**4**
GER - Bremen	11	0.53	0.12	0.00	0.36	**1**	0.00	0.01	0.00	0.99	**4**
GER - Hamburg	28	0.00	0.00	0.00	1.00	**4**	0.00	0.00	0.00	1.00	**4**
GER - Hessen	82	0.05	0.26	0.00	0.69	**4**	0.00	0.00	0.00	1.00	**4**
GER - Mecklenburg-Vorpommern	121	0.11	0.02	0.00	0.86	**4**	0.00	0.00	0.00	1.00	**4**
GER - Niedersachsen	124	0.12	0.18	0.00	0.70	**4**	0.00	1.00	0.00	0.00	**2**
GER - Northrhein-Weastfalen	265	0.34	0.03	0.00	0.64	**4**	0.00	0.00	0.00	1.00	**4**
GER - Rheinland-Pfalz	67	0.24	0.12	0.00	0.64	**4**	0.00	0.58	0.00	0.42	**2**
GER - Saarland	17	0.03	0.05	0.00	0.93	**4**	0.02	0.02	0.00	0.96	**4**
GER - Sachsen	303	0.06	0.03	0.00	0.91	**4**	0.00	0.00	0.00	1.00	**4**
GER - Sachsen-Anhalt	172	0.04	0.12	0.00	0.84	**4**	0.00	0.00	0.00	1.00	**4**
GER - Schleswig-Holstein	34	0.00	0.00	0.00	1.00	**4**	0.00	0.00	0.00	1.00	**4**
GER - Thuringen	153	0.02	0.00	0.00	0.97	**4**	0.00	0.00	0.00	1.00	**4**
Denmark	1001	0.01	0.97	0.00	0.01	**2**	0.00	1.00	0.00	0.00	**2**
SPA - north	112	0.77	0.01	0.17	0.05	**1**	1.00	0.00	0.00	0.00	**1**
SPA - north-east	105	0.67	0.00	0.29	0.03	**1**	1.00	0.00	0.00	0.00	**1**
SPA - Madrid	127	0.84	0.02	0.02	0.12	**1**	1.00	0.00	0.00	0.00	**1**
SPA - centre	136	0.22	0.00	0.78	0.00	**3**	1.00	0.00	0.00	0.00	**1**
SPA - east	273	0.55	0.00	0.29	0.16	**1**	1.00	0.00	0.00	0.00	**1**
SPA - south	209	0.22	0.00	0.78	0.00	**3**	1.00	0.00	0.00	0.00	**1**
SPA - Canary Islands	38	0.25	0.00	0.75	0.00	**3**	1.00	0.00	0.00	0.00	**1**
FIN - east	134	0.06	0.03	0.00	0.91	**4**	0.00	0.00	0.00	1.00	**4**
FIN - south	645	0.01	0.69	0.00	0.30	**2**	0.00	0.00	0.00	1.00	**4**
FIN - west	105	0.03	0.05	0.00	0.93	**4**	0.00	0.00	0.00	1.00	**4**
FIN - north	112	0.06	0.03	0.00	0.91	**4**	0.00	0.00	0.00	1.00	**4**
FRA - Ile de France	196	0.01	0.07	0.00	0.92	**4**	0.00	0.00	0.00	1.00	**4**
FRA - Bassin Parisien	177	0.24	0.12	0.00	0.64	**4**	0.00	0.00	0.00	1.00	**4**
FRA - north-Pas-de-Calais	73	0.24	0.12	0.00	0.64	**4**	0.46	0.00	0.00	0.54	**4**
FRA - east	90	0.84	0.02	0.02	0.12	**1**	0.98	0.00	0.00	0.02	**1**
FRA - west	140	0.52	0.01	0.02	0.45	**1**	0.00	0.00	0.00	1.00	**4**
FRA - south-west	109	0.24	0.12	0.00	0.64	**4**	0.07	0.00	0.00	0.93	**4**
FRA - centre-east	115	0.30	0.40	0.00	0.29	**2**	1.00	0.00	0.00	0.00	**1**
FRA - Mediterranean area	102	0.24	0.12	0.00	0.64	**4**	1.00	0.00	0.00	0.00	**1**

(continue)

Table A.6 (continued)

European regions	N. resp.	Prior probabilities $\hat{P}\left(W_j = m \middle| \mathbf{Z}_j\right)$ 1	2	3	4	Modal	Posterior probabilities $\hat{P}\left(W_j = m \middle| \mathbf{Y}_j, \mathbf{Z}_j\right)$ 1	2	3	4	Modal
GRE - north	324	0.22	0.00	0.78	0.00	**3**	0.00	0.00	1.00	0.00	**3**
GRE - centre	580	0.18	0.00	0.79	0.02	**3**	0.00	0.00	1.00	0.00	**3**
GRE - islands	100	0.18	0.00	0.79	0.02	**3**	0.00	0.00	1.00	0.00	**3**
IRE - north, centre e west	278	0.77	0.00	0.23	0.00	**3**	1.00	0.00	0.00	0.00	**3**
IRE - south and east	718	0.81	0.01	0.01	0.18	**1**	1.00	0.00	0.00	0.00	**1**
ITA - north-west	284	0.19	0.05	0.00	0.77	**1**	0.00	0.00	0.00	1.00	**1**
ITA - north-east	204	0.72	0.05	0.00	0.22	**4**	0.00	0.00	0.00	1.00	**4**
ITA - centre	189	0.34	0.03	0.00	0.64	**1**	0.00	0.00	0.00	1.00	**4**
ITA - south	209	0.22	0.00	0.78	0.00	**4**	0.00	0.00	1.00	0.00	**4**
ITA - islands	106	0.22	0.00	0.78	0.00	**3**	0.00	0.00	1.00	0.00	**3**
Luxembourg	600	0.53	0.12	0.00	0.36	**3**	0.00	1.00	0.00	0.00	**3**
NL - north	93	0.03	0.86	0.00	0.10	**1**	0.00	1.00	0.00	0.00	**2**
NL - east	209	0.03	0.86	0.00	0.10	**2**	0.00	1.00	0.00	0.00	**2**
NL - west	479	0.01	0.91	0.00	0.07	**2**	0.00	1.00	0.00	0.00	**2**
NL - south	225	0.01	0.91	0.00	0.07	**2**	0.00	1.00	0.00	0.00	**2**
POR - north	343	0.29	0.00	0.71	0.00	**2**	0.00	0.00	1.00	0.00	**2**
POR - south (Algarve)	36	0.22	0.00	0.78	0.00	**3**	0.00	0.00	1.00	0.00	**3**
POR - centre	178	0.33	0.00	0.67	0.00	**3**	0.00	0.00	1.00	0.00	**3**
POR - Lisbon	342	0.55	0.00	0.29	0.16	**3**	0.00	0.00	1.00	0.00	**3**
POR - centre-south (Alentejo)	56	0.18	0.00	0.79	0.02	**1**	0.00	0.00	1.00	0.00	**3**
POR - Azores	22	0.33	0.00	0.67	0.00	**3**	0.00	0.00	1.00	0.00	**3**
POR - Madeira	24	0.22	0.00	0.78	0.00	**3**	0.00	0.00	1.00	0.00	**3**
Sweden	1000	0.00	0.30	0.00	0.69	**3**	0.00	1.00	0.00	0.00	**3**
UK - north-east	54	0.96	0.02	0.02	0.00	**4**	1.00	0.00	0.00	0.00	**2**
UK - north-west	141	0.90	0.05	0.00	0.05	**1**	1.00	0.00	0.00	0.00	**1**
UK - Yorkshire e Humber	88	0.94	0.05	0.00	0.01	**1**	1.00	0.00	0.00	0.00	**1**
UK - east Midlands	77	0.34	0.03	0.00	0.64	**1**	1.00	0.00	0.00	0.00	**1**
UK - west Midlands	96	0.90	0.05	0.00	0.05	**4**	1.00	0.00	0.00	0.00	**1**
UK - east Anglia	70	0.72	0.05	0.00	0.22	**1**	0.23	0.29	0.00	0.47	**1**
UK - London	126	0.53	0.12	0.00	0.36	**1**	1.00	0.00	0.00	0.00	**4**
UK - south-east	127	0.53	0.12	0.00	0.36	**1**	0.99	0.00	0.00	0.01	**1**
UK - south-west	81	0.94	0.00	0.04	0.02	**1**	1.00	0.00	0.00	0.00	**1**
UK -Wales	44	0.94	0.02	0.02	0.02	**1**	1.00	0.00	0.00	0.00	**1**
UK - Scotland	80	0.90	0.05	0.00	0.05	**1**	1.00	0.00	0.00	0.00	**1**
UK - Northern Ireland	304	0.84	0.02	0.02	0.12	**1**	1.00	0.00	0.00	0.00	**1**

Source: Our elaboration on Eurobarometer 56.1-2001 and Eurostat data.

Appendix B

The model presented and discussed in Chapter 5 has been defined using the syntax language of the software Latent GOLD 4.5. The syntax module makes the model structure much more transparent, allowing the definition of the technical and output options, of the observed and latent variables, and of the model's equations and residual associations. The syntax implemented is the following:

```
options
   algorithm
      tolerance=1e-008 emtolerance=0,01 emiterations=350 nritertions=100;
   startvalues
      seed=0 sets=10 tolerance=1e-005 iterations=50;
   bayes
      categorical=1 variances=1 latent=1 poisson=1;
   montecarlo
      seed=0 replicates=500 tolerance=1e-008;
   quadrature  nodes=10;
   missing  includedependent;
    output  parameters=effect  standarderrors  bivariateresiduals
profile  probmeans  classification=model classification=posterior ;

variables
     groupid regio1;
     dependent perc_inc nominal, eco_diff2 nominal,  quart_inc ordinal, inferior
nominal,
     rel1 ordinal, assoc nominal, help nominal, escl nominal, useful nominal,
     health_serv nominal, soc_ass nominal, theft nominal;


   independent cleta4 nominal,  occup2 nominal,
   taxes_inc_quart ordinal,  gdp_quart ordinal, individ_cl ordinal;

    latent
      Gregion group nominal 4,
      Cluster nominal 6;

equations
Gregion <- 1 + taxes_inc_quart  + gdp_quart + individ_cl ;
Cluster <- 1 | Gregion +  cleta4 +  occup2;
   perc_inc <- 1 + Cluster;
   eco_diff2 <- 1 + Cluster;
   quart_inc <- 1 + Cluster ;
   inferior <- 1 + Cluster;
   rel1 <- 1 + Cluster + Gregion;
   assoc  <- 1 + Cluster + Gregion;
   help <- 1 + Cluster;
   escl <- 1 + Cluster ;
   useful <- 1 + Cluster;
   health_serv <- 1 + Cluster ;
   soc_ass <- 1 + Cluster ;
   theft <- 1 + Cluster ;
   escl <-> useful;
   inferior <-> useful;
   escl <-> inferior;
   health_serv <-> soc_ass;
   help <-> rel1;
end model
```

Bibliography

Agresti A. (2002), *Categorical Data Analysis,* Wiley and Sons, Hoboken, New Jersey, 2nd Edition.

Agresti A., Booth J.G. and Caffo B. (2000), *Random-effects modelling of categorical response data*, Sociological Methodology, vol. 30, pp. 27-80.

Aitkin M. (1999), *A general maximum likelihood analysis of variance components in generalized linear model*, Biometrics, n. 55, pp. 117-128.

Asparouhov T. and Muthen B.O. (2008), *Multilevel mixture models*, in Hancock G.R. and Samuelsen K.M. (eds) *Advances in latent variable mixture models*, IAP.

Atkinson A.B. (1998), *Social exclusion, poverty and unemployment, in* A.B. Atkinson, J. Hills (eds.) *Exclusion, employment and opportunity,* CASE/paper 4, pp 1–20, Centre for Analysis of Social Exclusion, London, School of Economics.

Atkinson A.B., Marlier E. and Nolan B. (2004), *Indicators and targets for social inclusion in the European Union,* Journal of Common Market Studies, vol. 42, n. 1, pp. 47-75.

Atkinson R. and Davoudi S. (2000), *The concept of social exclusion in the European Union: context, development and possibilities,* Journal of common market studies, vol. 38, n. 3, pp. 427-448.

Atkinson T., Cantillon B., Marlier E. and Nolan B. (2002), *Social Indicators: The EU and Social Inclusion*, Oxford University Press.

Barnes M., Heady C., Middleton S., Millar J. Papadopoulos F. and Tsakloglou P. (2002), *Poverty and social exclusion in Europe*, Edward Elgar Publishing.

Barry B. (1998), *Social Exclusion, Social Isolation and the Distribution of Income,* CASE/12, Centre for Analysis of Social Exclusion, London School of Economics.

Bartholomew D.J. and Knott M. (1999), *Latent Variable Models and Factor Analysis*, London: Arnold.

Bartholomew D.J., Steele F., Moustaki I. and Galbraith J.I. (2002), *The analysis and interpretation of multivariate data for social scientists*, Chapman & Hall.

Baum L.E., Petrie G.S. and Weiss N. (1970), *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains,* Annals of Mathematical Statistics, vol. 41, pp. 164-171.

Berghman J. (1995), *Social exclusion in Europe: policy context and analytical framework*, in Room G. (eds.) *Beyond the threshold. The measurement and analysis of social exclusion*, Bristol: The Policy Press.

Bhalla A.S. and Lapeyre F. (1997), *Social exclusion: towards an analytical and operational framework*, Development and Change, vol. 28, pp. 413-33.

Bhalla A.S. and Lapeyre F. (2004), *Poverty and exclusion in a global world*, London: Macmillan.

Böhnke P. (2008), *Are the poor socially integrated? The link between poverty and social support in different welfare regimes*, Journal of European Social Policy, vol. 18, n. 2, pp. 133-150.

Burchardt T., Le Grand J. and Piachaud D. (1999), *Social exclusion in Britain 1991-1995*, Social Policy and Administration, vol. 33, n. 3, pp. 227-244.

Burchardt T., Le Grand J. and Piachaud D. (2002), *Degrees of exclusion: developing a dynamic multidimensional measure*, in Hills J., Le Grand J. and Piachaud D. (eds) *Understanding social exclusion*, pp 30-43, Oxford: Oxford University Press.

Clogg C.C. (1981), *New developments in latent structure analysis*, in Jackson D.J. and Borgotta E.F. (eds.), *Factor analysis and measurement in sociological research*, Beverly Hills: Sage Publication.

Clogg C.C. (1995), *Latent class models*, in Arminger G., Clogg C.C. and Sobel M.E. (eds.), *Handbook of statistical modelling for the social and behavioural sciences*, Chapter 6, p. 311-359, New York: Plenum Press.

Clogg C.C. and Goodman L.A. (1984), *Latent structure analysis of a set of multidimensional contingency tables*, Journal of the American Statistical Association, vol. 79, pp. 762-771.

Clogg C.C. and Goodman L.A. (1985), *Simultaneous latent structure analysis in several groups*, in Tuma N.B. (ed.) Sociological Methodology, San Francisco: Josey-Bass.

Collins L.M., Findler P.L., Wugalter S.E. and Long J.D. (1993), *Goodness-of-fit testing for latent class models*, Multivariate Behavioural Research, vol. 28, n. 3, pp. 375-389.

Commins P. (1993), *Combating exclusion in Ireland 1990-1994: a midway report*, Brussels: Observatory on National Policies to Combat Social Exclusion, Commission of European Countries.

Croon M. (2002), *Ordering the classes*, in Hagenaars J.A. and McCutcheon A.L. *Applied Latent Class Analysis*, Cambridge University Press.

Day N.E. (1969), *Estimating the components of a mixture of normal distributions*, Biometrika, vol. 56, n. 3, pp. 463-474.

Dayton C.M. and McReady G.B. (1988), *Concomitant-variable latent class models*, Journal of the American Statistical Association, vol. 83, pp. 173-178.

De Boeck P. and Wilson M. (2004), *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*, New York: Springer.

De Haan A. (1998), *Social Exclusion: An Alternative Concept for the Study of Deprivation?*, IDS Bulletin vol. 29, n.1, pp. 10-19.

Democratic Dialogue (1995), *Special Reports: Report 2 – Social Exclusion*, Social Inclusion, http://www.dem-dial.demon.co.uk/index.htm.

Dempster A.P., Laird N.M. and Rubin D.B. (1977), *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society B, vol. 39, pp. 1-38.

Dewilde C. (2004), *The multidimensional measurement of poverty in Belgium and Britain: a categorical approach*, Social Indicator Research, vol. 68, pp. 331-369

Dewilde C. (2008), *Individual and institutional determinants of multidimensional poverty: a European comparison*, Social Indicator Research, vol. 86, pp. 233-256.

Diener E. and Suh E. (1997), *Measuring quality of life: economic, social and subjective indicators*, Social Indicator Research, vol. 40, pp. 189-216.

Easterlin R.A. (2001), *Income and happiness: towards a unified theory*, The Economic Journal, vol. 111, n. 473, pp. 465–484.

Eroglu S. (2007), *Developing an index of deprivation which integrates objective and subjective dimensions: extending the works of Townsend, Mack and Lansley, and Hallerod*, Social indicators research, n. 80, pp. 493-510.

European Commission (1993) *Background Report: Social Exclusion – Poverty and Other Social Problems in the European Community*, ISEC/B11/93, Luxembourg: Office for Official Publications of the European Communities.

European Commission (1997), *First report on Economics and Social Cohesion 1996*, Luxembourg: Office for Official Publications of the European Communities.

European Commission (1998), *Non-monetary indicators of poverty and social exclusion: final report*, European Commission, Brussels.

European Commission (2000), *Fight against Poverty and Social Exclusion – Definition of Appropriate Objectives*, European Commission, Brussels.

European Commission (2003), *Laeken indicators – Detailed calculation methodology*, European Commission, Brussels.

European Commission (2004), *Perceptions of social integration and exclusion in an enlarged Europe,* European Foundation for the improvement of living conditions, Office for Official Publications of the European Communities, Luxembourg.

European Commission (2005), *Regional indicators to reflect social exclusion and poverty*, Final report prepared for Employment and Social Affairs DG, Brussels.

European Union (2001), *Report on Indicators in the field of poverty and social exclusion*, Social Protection Committee.

European Union (2007a), *2010 to be the European Year for Combating Poverty and Social Exclusion*, Press Release IP/07/1905, Brussels.

European Union (2007b), *Treaty of Lisbon*, Community Treaties, agreements and conventions, Official Journal of the European Union C306, 17.12.2007.

Eurostat (2005), *Income, poverty and social exclusion in the EU 25*, Eurostat, Luxembourg.

Foster J., Greer J. and Thorbecke E. (1984), *A class of decomposable poverty measure*, Econometrica, 52 (3), pp. 761-766.

Gallie D. and Paugman S. (2002), *Social Precarity and Social Integration*, European Commission, Employment & social affairs, Brussels.

Goodman L.A. (1974a), *Exploratory latent structure analysis using both identifiable and unidentifiable models*, Biometrika, vol. 61, n. 2, pp. 215-231.

Goodman L.A. (1974b), *The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach*, American Journal of Sociology, vol. 79, pp. 1179-1259.

Goodman L.A. (1979), *On the estimation of parameters in latent structure analysis*, Psychometrika, vol. 44, pp. 123-128.

Goodman L.A. (2002), *Latent class analysis: the empirical study of latent types, latent variables, and latent structures*, in Hagenaars J.A. and McCutcheon A.L. *Applied Latent Class Analysis*, Cambridge University Press

Gordon D. (2005), *Indicators of Social Quality*, European Journal of Social Quality, vol. 5, n. 1-2, pp. 1-7

Gordon D., Adelman L., Ashworth K., Bradshaw J., Levitas R., Middleton S., Pantazis C., Patsios D., Payne S., Townsend P. and Williams J. (2000), *Poverty and Social Exclusion in Britain*, York: Joseph Rowntree Foundation.

Granovetter M. (1985), *Economic action and social structure: the problem of embeddedness*, American Journal of Sociology, vol. 91, n. 3.

Haberman S.J. (1979), *Analysis of Qualitative Data, Vol 2, New Developments*, New York: Academic Press.

Haberman S.J. (1988), *A Stabilized Newton-Raphson Algorithm for Log-Linear Models for Frequency Tables Derived by Indirect Observation*, Sociological Methodology, vol. 18, pp. 193-211.

Hagenaars J.A. (1988), *Latent structure models with direct effects between indicators*, Sociological Methods and Research, vol. 16, pp. 379-405.

Hagenaars J.A. (1990*). Categorical Longitudinal Data - Loglinear Analysis of Panel, Trend and Cohort Data*, Newbury Park: Sage.

Hagenaars J.A. (1993), *Loglinear models with latent variables*, Sage University Papers series on Quantitative Applications in the Social Sciences, 07-094, Newbury Park, CA: Sage.

Hagenaars J.A. and De Vos K. (1988), *The definition and measurement of poverty*, The Journal of Human Resources, vol. 23, n. 2, pp. 211-221.

Hagenaars J.A. and McCutcheon A.L. (2002), *Applied Latent Class Analysis*, Cambridge University Press.

Heidenreich M. (2003), *Regional inequalities in the enlarged Europe*, Journal of European Social Policy, vol. 13, n. 4, pp. 313-333.

Heinen T. (1996), *Latent Class and Discrete Latent Trait Models: Similarities and Differences*, Thousand Oakes: Sage Publications.

Hills J., Le Grand J. and Pichaud D. (eds.) (2002), *Understanding social exclusion*, Oxford University Press.

ILO (2003), *Concepts and strategies for combating social exclusion. An overview*, Strategies and tools against social exclusion and poverty programme, Social Security and development branch, International Labour Office.

Jenks G.F. (1963), *Generalization in statistical mapping*, Annals of the Association of American Geographers, vol. 53, pp. 15-26.

Jordan B. (1996), *A theory of poverty and social exclusion*, Blackwell, Oxford.

Kieselbach T. (2003), *Long-Term Unemployment among Young People: The Risk of Social Exclusion*, American Journal of Community Psychology, vol. 32, n. 1-2.

Kronauer M. (1998), *"Social Exclusion" and "Underclass", New Concepts for the Analysis of Poverty*, in Andres H.J. (ed.) *Empirical Poverty Research in a Comparative Perspective*, pp. 51–76, Aldershot, Ashgate.

Lazarsfeld P.F. (1950), *The logical and mathematical foundation of latent structure analysis and the interpretation and mathematical foundation of latent structure analysis*, in Stouffer S.A. *et al.* (eds.), *Measurement and Prediction*, pp. 362-472, Princeton, NJ: Princeton University Press.

Lazarsfeld P.F. and Henry N.W. (1968), *Latent Structure Analysis*, Boston: Houghton Mifflin.

Little R.J.A. and Rubin D.B. (1987), *The analysis of social science data with missing values*, Sociological Methods and Research, vol. 18, n. 2-3, pp. 292-326.

Lukociene O. and Vermunt J.K. (in press), *Determining the number of components in mixture models for hierarchical data*, Studies in Classification, Data Analysis, and Knowledge Organization, Springer: Berlin-Heidelberg.

Maas C.J.M. and Hox J.J. (2004), *Robustness issues in multilevel regression analysis*, Statistica Neerlandica, vol. 58, n. 2, pp. 27-137.

Magidson J., and Vermunt J.K. (2004), *Latent class models*, in Kaplan D. (ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, pp. 175-198. Thousand Oakes: Sage Publications.

Marshall T.H. (1964), *Class, citizenship and social development*, New York: Doubleday.

Mayes D.J., Berghman J. and Salais R. (2001), *Social exclusion and European policy*, Edward Elgar.

McCutcheon A.L. (1986), *Sexual morality, pro-life values, and attitudes toward abortion: a simultaneous latent structure analysis for 1978-83*, Paper presented at the Annual Meeting of the Eastern Sociological Society, New York, NY, April 1986.

McCutcheon A.L. (1987), *Latent class analysis*, Newbury Park, CA: Sage.

McLachlan G.J and Peel D. (2000), *Finite mixture models*, New York: Wiley.

Michalos A.C. (1997), *Combining social, economic and environmental indicators to measure sustainable human well-being*, Social Indicator Research, vol. 40, pp. 221-258.

Micklewright J. and Stewart K. (2001), *Poverty and social exclusion in Europe: European comparisons and the impact of enlargement*, New Economy, vol. 8, n. 2, pp. 104-109.

Moisio P. (2004), *A Latent Class Application to the Multidimensional Measurement of Poverty*, Quantity and Quality: International Journal of Methodology, vol. 38, pp. 703-717.

Muthén B.O. (1994), *Multilevel covariance structure analysis*, Sociological Methods & Research, vol. 22, pp.376-398.

Muthén B.O. (2002), *Beyond SEM: general latent variable modelling*, Behaviormetrika, vol. 29, n. 1, pp. 81-117.

Muthén L.K. and Muthén B.O. (1998-2007), Mplus User's Guide. Fourth Edition, Muthén and Muthén, Los Angeles, CA.

Negri N. and Saraceno C. (2000), *Povertà, disoccupazione ed esclusione sociale*, Stato e mercato, vol. 59, n. 2, pp. 175-210.

Nylund K.L., Asparouhov T., Muthén B.O. (2007), *Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study*, Structural Equation Modelling, vol. 14, n.4, pp. 535–569.

Ogg J. (2005), *Social exclusion and insecurity among older Europeans: the influence of welfare regimes*, Ageing & Society, n. 25: 69-90.

Peace R. (2001), *Social exclusion: a concept in need of definition?*, Social Policy Journal of New Zealand, vol. 16.

Percy-Smith J. (ed.) (2000), *Policy Responses to Social Exclusion: Towards Inclusion?*, Open University Press, Buckingham, Philadelphia.

Petrucci A. and Schifini S. (2004), *Quality of life in Europe: objective and subjective indicators. A spatial analysis using classification techniques*, Social Indicator Research, vol. 60, pp. 55-88

Pirani E. and Schifini S. (2008), *Differenze regionali nei processi di esclusione sociale nell'Europa mediterranea*, paper presented at the XLV SIEDS Scientific Meeting, Bari, 2008 May 29-31.

Rabe-Hesketh S., Skrondal A. and Pickles A. (2004), *GLLAMM Manual*, Berkeley Division of Biostatistics Working Paper Series, University of California, Berkeley.

Read T.R.C. and Cressie N.A.C. (1988), *Goodness-of-fit statistics for discrete multivariate data*, New York: Springer-Verlag.

Robila M. (2006), *Economic pressure and social exclusion in Europe*, The Social Science Journal, n. 43, pp. 85-97.

Room G. (1995), *Beyond the Threshold: The Measurement and Analysis of Social Exclusion*, The Policy Press, Bristol

Room G. (1998), *Social Quality in Europe: perspectives on social exclusion*, in Beck W., Van der Maesen L. e Walker, A. *The social quality of Europe*, pp. 289-297, Bristol, The Policy Press.

Scharf T., Phillipson C. and Smith A.E. (2005), *Social exclusion of older people in deprived urban communities of England*, European Journal of Ageing, vol. 2, pp. 76-87.

Schifini S. and Petrucci A. (2002), *Quality of life in Europe: objective and subjective indicators. A spatial Analysis Using Classification Techniques*, Social Indicators Research, vol. 60, pp. 55-88.

Schifini S. and Petrucci A. (2007), *Social uneasiness in Europe: a comparison of the situation in 1994 and ten years later*, ISI 2007.

Sen A.K. (1975), *Employment, Technology and Development*, Oxford: Clarendon Press.

Sen A.K. (2000), *Social exclusion: concept, application, and scrutiny*, Social Development Papers, n. 1, Manila, Office of Environment and Social Development, Asian Development Bank.

Silver H. (1994), *Social Exclusion and Social Solidarity: Three Paradigms*, International Labour Review, Vol. 133

Silver H. and Miller S.M. (2003), *Social Exclusion: The European Approach to Social Disadvantage*, Indicators, vol. 2, n. 2.

Skrondal A. and Rabe-Hesketh S. (2004), *Generalized Latent Variable Modelling: Multilevel, Longitudinal, and Structural Equation Models*, Chapman and Hall/CRC, Boca Raton, FL.

Skrondal A. and Rabe-Hesketh S. (2007), *Latent variable modelling: a survey*, Board of the Foundation of the Scandinavian Journal of Statistics, vol. 34, pp. 712-745.

Snijders T.A.B. and Bosker R.J. (1999), *An introduction to basic and advanced multilevel modelling*, London: Sage Publications.

Social Exclusion Unit (1998) *Bringing Britain Together: A National Strategy for Neighbourhood Renewal*, presented to Parliament by the Prime Minister by Command of Her Majesty September 1998, http://www.cabinetoffice.gov.uk/.

Social Exclusion Unit (1999) *Opportunity for All: Tackling Poverty and Social Exclusion*, summary of the UK Labour Government's First Report on Tackling Poverty and Social Exclusion, September 1999, CM 4445, the Stationery Office, London.

Stewart K. (2003), *Monitoring social inclusion in Europe's regions*, Journal of European Social Policy, vol. 13, n. 4, pp. 335-356.

Strobel P. (1996), *From poverty to exclusion: a wage-earning society or a society of human rights?*, International Social Science Journal, vol. 48, n. 2, pp. 173-189.

Testa M.R. and Grilli L. (2006), The influence of childbearing regional contexts on ideal family size in Europe, Population, vol. 61, n.1-2, pp. 109-138.

The Scottish Office (1999) Social Inclusion – Opening the Door to a Better Scotland, www.scotland.gov.uk/library/documents-w7/sist-00.htm .

Townsend P. (1979), *Poverty in the United Kingdom*, London: Penguin Books.

Tsakloglou P. and Papadopoulos F. (2002), *Aggregate level and determining factors of social exclusion in twelve European countries*, Journal of European Social Policy, vol. 12, n. 3, pp. 211-225.

Vermunt J.K. (1997), *Log-linear models for event histories*, Thousand Oakes: Sage Publications.

Vermunt J.K. (2003), *Multilevel latent class models*, Sociological Methodology, vol. 33, n. 1, pp. 213-239.

Vermunt J.K. (2004), *An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models*, Statistica Neerlandica, vol. 58, n. 2, pp. 220-233.

Vermunt J.K. (2007), *Multilevel mixture item response theory models: an application in education testing*, Bulletin of the International Statistical Institute, 56th Session, paper n. 1253, 1-4. ISI 2007: Lisboa, Portugal.

Vermunt J.K. (2008), *Latent class and finite mixture models for multilevel data sets*, Statistical Methods in Medical Research, n. 17, pp. 33-51.

Vermunt J.K. (in press), *Mixture models for multilevel data sets*, in Hox J. and Roberts J.K., *Handbook of Advanced Multilevel Analysis*, Lawrence Erlbaum.

Vermunt J.K. and Magidson J. (2004), *Latent class analysis*, in Lewis-Beck M.S., Bryman A., and Liao T.F. (eds.), The Sage Encyclopaedia of Social Sciences Research Methods, 549-553. Thousand Oakes, CA: Sage Publications.

Vermunt J.K. and Magidson J. (2005a), *Hierarchical mixture models for nested data structures*, in Weihs C. and Gaul W. (eds.), *Classification: the ubiquitous challenge*, pp. 176-183, Heidelberg: Springer.

Vermunt J.K. and Magidson J. (2005b), *Technical guide for Latent GOLD 4.0: basic and advanced*, Belmont Massachusetts: Statistical Innovations Inc.

Vermunt J.K. and Magidson J. (2005c), *Latent GOLD 4.0 User's Guide*, Belmont Massachusetts: Statistical Innovations Inc.

Vermunt J.K. and Magidson J. (2008), *LG-Sintax User's Guide: Manual for Latent GOLD 4.5: Syntax Module*, Belmont Massachusetts: Statistical Innovations Inc.

Vermunt J.K. and Van Dijk L.A. (2001), *A non-parametric random-coefficient approach: the latent class regression model*, Multilevel Modelling Newsletter, vol. 13, pp. 6-13.

Whelan B.J. and Whelan C.T. (1995), *In what sense is poverty multidimensional?*, in Room G. (eds.), *Beyond the Threshold: The Measurement and Analysis of Social Exclusion*, The Policy Press, Bristol.

Whelan C.T. and Maître B. (2005a), *Economic Vulnerability, multidimensional deprivation and social cohesion in an enlarged European community*, International Journal of Comparative Sociology, vol. 46, pp. 215-239.

Whelan C.T. and Maître B. (2005b), *Vulnerability and Multiple Deprivation: Perspectives on Economic Exclusion in Europe: A Latent Class Analysis*, European Societies, vol. 7, pp. 423-450.

Wolfe J.H. (1970), *Pattern clustering by multivariate cluster analysis*, Multivariate Behavioural Research, vol. 5, pp. 329-350.