No Man is an Island: the Inter-personal Determinants of Regional Well-Being in Europe *

Aqib Aslam† Luisa Corrado‡

April 2007

Abstract

There is a strong need to complement the analysis of social well-being at the European regional level to supplement existing, predominantly economic analysis. This work extends the measurement of well-being across the EU-15 regions in several ways. First, we assess the determinants of well-being using a multilevel modelling approach using data at the national, regional and individual levels. Second, we have extended the model to account for the effects of social interactions within each group, as well as intrinsic socio-demographic indicators and higher-level exogenous contextual factors. Empirical findings support the idea that well-being is strongly dependent both on these general forms of social interactions and on more specific individual characteristics. We find that there is some evidence of greater regional effects relative to national effects, but individual well-being continues to be affected most by micro-level phenomena.

Key-Words: Multilevel Modelling, Regional Well-Being, Social Interactions, Social Distance.

JEL: R1, I31, O18, D31, D6.

*We would like to thank Thomas Goldring for excellent research assistance, Gernot Doppelhofer, Takashi Yamagata and Felicia Huppert for suggestions and comments on the work.

†University of Cambridge, Faculty of Economics. maa49@cam.ac.uk, UK.

‡Corresponding Author: Marie-Curie Fellow, University of Cambridge, UK and University of Rome Tor Vergata. E-mail: lc242@econ.cam.ac.uk. Luisa Corrado gratefully acknowledges the Marie-Curie Intra European fellowship 514913.

1 Introduction

Economics is no stranger to jumping across the disciplinary divides and for well over a decade has been bedfellows with psychology as the fields of experimental and behavioural economics have developed. As these areas of the subject expand and with accomplishments in such fields being recognised in the 2002 Nobel prize, economics has sidled off its pedestal as the 'dismal' science and closer to its kindlier and undoubtedly warmer neighbours in the social or 'soft' sciences. And as social indicators for such traditionally private events as happiness and life satisfaction have developed over recent years following the collation of large quantities of survey data, it was only a matter of time before economics sought to understand the diverse drivers of human emotion at the macroeconomic level.

Traditionally economic research has focused on collecting and analysing economic data, while little attention has been paid to measuring attitudes and emotional states. Therefore economists are able to comment extensively on the material conditions of an individual's life but very little can be said about their social experiences and their quality of life, to the extent that Nobel laureate Daniel Kahneman has even called for the establishment of national 'well-being accounts'[1] to supplement existing economic data.

Therefore, there is clearly a strong need to augment the analysis of social well-being in economics, both via the data collected and the analyses implemented. And given the burgeoning literature on the multidimensional aspect of individual well-being, there are numerous findings to start from. Research on the relationship between material circumstances and how people feel about their lives, finds that in developed countries, economic factors account for only about 10% of the variation between individuals in measures of life satisfaction or happiness (Helliwell and Putnam (2005); Lyubomirsky, Sheldon, and Schkade (2005)). What is actually relevant is that individuals' levels of life satisfaction and happiness depend mostly on 'intentional activities', i.e. the way individuals are functioning both personally and inter-personally. Ormel, Lindenberg, Steverink, and Verbrugge (1999) have already attempted to introduce such factors into their work on social production function (SPF) theory.

Existing evidence also suggests that it is relative income not absolute income that matters most (Easterlin (1994), Easterlin (2003); Blanchflower and Oswald (2004)). Easterlin (1974) argues that consumption norms exist in societies as standards against which individuals measure their own achievements - the eponymous "keeping up with the Joneses". As objective conditions change across countries and regions, so do social norms and society benchmarks. This can explain why many commentators have noted recently that more prosperous countries are no happier than poorer countries.

Duncan (1975) also finds that satisfaction with the current standard of living is related to one's relative position in the income distribution, alongside factors which take into account an individual's absolute income level. Campbell, Converse, and Rodgers (1976) similarly

[1] Kahneman D., Krueger A.B., Schkade D., Schwarz N. and Stone A. (2004), 'Toward National Well-Being Accounts', The American Economic Review, Volume 94, Number 2, pp. 429-434(6)

find that well-being is dependent on a person's comparison of what they feel they should be achieving - in this case there is an internal benchmark rather than an external social one.

These studies highlight that individuals use a variety of reference points from their own experiences, beliefs and social imprints to evaluate their current situation. As such, we see various 'structural constraints' on attitudes, beliefs and behaviour which have developed through common social values exerting a measure of influence on well-being. As Fernandez and Kulik (1981) note, "it is not my income that makes me happy, but rather a favourable comparison between my income and others'". This is one explanation for the otherwise astonishing observation that although real per capita incomes have quadrupled in the past 50 years, in most advanced economies, aggregate levels of subjective well-being have remained essentially unchanged. This is the widely-cited empirical observation known as the 'Easterlin paradox'[2], which provides a justification for our multilevel approach since changes in mean covariates such as mean incomes across a region might have little effect on well-being whereas large changes in individual incomes relative to the regional average could contribute much more.

When we consider the nature of the survey data most commonly collected these days, measures for subjective well-being (well-being as defined by the individual herself) using relatively simple self-rating questions about happiness and life satisfaction are significantly more reliable as a means for disclosing an individual's state. Paying greater attention to subjective well-being and to the formalisation of models of healthy behaviour through the lifespan could have significant implications for social intervention and even economic policy, with the motivation behind such scrutiny being the belief that investigating an individual's perceived resources and quality of life can actually help promote their subjective well-being and social integration.

Studies on subjective well-being derive from two main perspectives: *hedonism* and *eudaimonia*. The former is roughly expressed as the human desire for pleasure, while the latter refers to the Aristotelian concept of eudaimonia, which is the human desire for overall fulfillment - covering such themes as self-actualisation and a commitment to socially-shared goals. According to Aristotle, the hierarchy of human purposes aims at *eudaimonia* as the ultimate goal. It constitutes rational activity which manifests the virtues of character, including courage and honesty; the intellectual virtues, such as rationality in judgment; and mutually beneficial relationships and scientific knowledge.

It is this distinction between happiness ('hedonism') and life satisfaction ('eudaimonia') as constituents of well-being, which we seek to explore in this paper. Hedonism is considered a more immediate human response whereas life satisfaction, in the eudaimonic sense, refers to a more collectively motivated mindset, i.e. the individual not only maximises personal pleasure or pain, but also considers social spillovers and a type of social goal congruence in his actions. In addition, self-ratings of 'happiness' tend to reflect short-term, situation-dependent expressions of mood, whereas self-ratings of 'life satisfaction' appear to measure longer-term, more projectual evaluations, indicating the extent to which one's experiences

[2]See, for example, Easterlin (1974) and, on modern data, Blanchflower and Oswald (2004).

match one's expectations (Huppert, Baylis, and Keverne (2005)). Both are broadly consistent measures, but can be considered separately as inputs into subjective well-being.[3]

To further enrich our analysis, this work adopts a set of classifications for the factors influencing well-being beyond the simple binomial system of hedonism and eudaimonia. This classification of explanatory variables begins with the concept of *personal feelings*, i.e. hedonistic aspects of well-being such as pleasure, enjoyment, satisfaction, as well as the eudaimonic aspects of well-being, such as competency, interest or engagement, meaning or purpose in life (Huppert, Baylis, and Keverne (2005)). Twinned with this is the notion of *inter-personal feelings,* which characterises the quality of interactions with others (Elliott and Umberson (2004); Huppert, Baylis, and Keverne (2005)). In addition, we distinguish between a further pair of concepts: *personal functioning,* which describes how much control individuals have over their lives and the extent to which they perceive their activities as having purpose, and *inter-personal functioning* - what the individual does for other people or for their community in terms of pro-social behaviour (Helliwell and Putnam (2005)). Variables that come under the middle two headings are considered central to an overall sense of well-being (Ryff (1989); Ryff and Singer (1998); Seligman (2002)). Collectively, however, we treat these factors as exogenous social interaction effects, as the variables that come under these four groupings are taken from subjective survey data and therefore represent the extent to which individuals consider themselves connected or disconnected from their peers and local institutions, due to a variety of external as well as personal factors. Afterall, as our title suggests, 'no man is an island' and we seek to test whether these social interaction effects are strongest at the regional, national or at the individual level.

In addition to these social interactions, we also consider intrinsic social indicators as possible determinants of well-being, e.g. marital status, monthly earnings, and the level of education. These variables are intrinsic in that they are given for each individual interviewed in the survey and thus represent inherent descriptive elements of their current economic or social status. Finally, we introduce exogenous contextual factors, which account for the external characteristics of the respective regional and national groups that each individual belongs to, e.g. regional GDP and the geographical location of the region.

Therefore, in light of the preceding discussion on the concept of well-being and its translation into a suitable set of variables, the aim of the paper is twofold: more generally, to understand the determinants of well-being across European regions using a set of social and structural indicators at the national and regional level, as well as individual measures, drawn from the data collected by Eurostat and the European Social Survey (ESS); more specifically, to apply multilevel modelling techniques to assess the role of regional factors and social interactions in determining the well-being of individuals across European regions. We hope to model and test for the determinants of happiness and life satisfaction, and therefore ul-

[3]Other well-being researchers have adopted different approaches. For example, Seligman (2002) defines well-being as the combination of an hedonistic component (pleasure) and two eudaimonic components (engagement and meaning). For Ryff (1989) well-being comprises six eudaimonic elements: autonomy, environmental mastery, personal growth, positive relationships, purpose in life and self-acceptance.

timately subjective well-being, and, using this information, arrive at an assessment of the determinants of well-being across the EU-15 regions.

The paper is structured as follows. Section two describes a small socio-geographic model of well-being where the emergence of socio-geographic groupings is the results of individual decisions based on social interactions. This model represents the theoretical *a-priori* justification for the multilevel empirical model introduced in section three. Section four illustrates how the notion of social interactions can be modelled using a multilevel empirical framework. On the basis of these backgrounds, sections five and six discuss the dataset and the empirical results. Section seven concludes.

2 A Socio-Geographic Model of Well-Being

In order to support the material presented in this paper, we develop a small model of regional social clustering which borrows from Akerlof (1997)'s paper on social distance.[4] We aim to use this model to obtain an *a priori* theoretical justification for the analysis which is undertaken in the subsequent sections of the paper, whereby multilevel modelling is used to judge whether regional factors play a significant role in determining the well-being of individuals across the EU. The key concept is that there are certain features of social position (or 'location') that can be expected to affect subjective well-being.[5]

Akerlof began by distributing individuals randomly along the real line in one-dimensional space. Using a simple two-period framework he went on to show that an individual's choice of location on the line was determined by a mixture of both the 'intrinsic' value of the location and the expected benefits from social exchange, i.e. the benefits gained due to proximity with one's neighbours. These two components are captured in a utility function, U, for each individual which is maximised so as to determine the direction of their movement in the following period along the real line.

$$U = \frac{A}{\underbrace{f\left(x_{0,i} - x_{0,j}\right)}_{\text{Inherited Social Position}} \underbrace{f\left(x_{1,i} - x_{0,j}\right)}_{\text{Social Distance}}} + \underbrace{\left[-ax_{1,i}^2 + bx_{1,i} + z + w_i\right]}_{\text{IntrinsicValue}} \tag{1}$$

where A is a constant of proportionality, and a and b are arbitrary constants. The first subscript denotes time zero and one, while the second is an index for the individual, $i = 1, 2, 3, ..., N$.

[4] Akerlof's model developed the ideas of the traditional gravity model from physics, as had been applied to the theory of international trade before. In the latter case, trade between two countries is proportional to their GNPs and inversely proportional to the square of their distance. This implied that the benefits from trade would increase as distance approaches.

[5] We do not pursue Krugman's theories (Krugman (1991a); Krugman (1991b); Krugman (1992); Krugman (1993)) on dynamic spatial economies, as we wish to abstract from the issue of increasing returns to scale as a source for the agglomeration of economic activity. Instead, we wish to concentrate on why social clusters form, and why the composition of such clusters could in turn influence individual well-being. As such we would like to develop a simple model which could suggest why intra-regional factors are more significant on individuals than inter-regional factors.

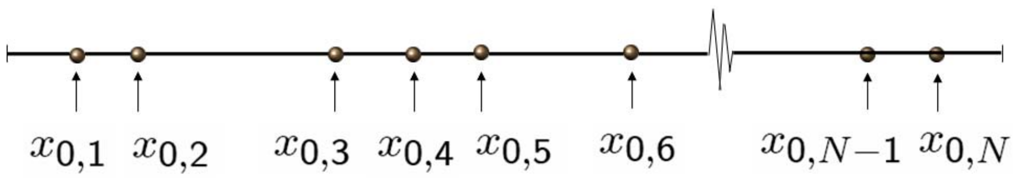


Figure 1: The Emergence of Social Groupings based on Social Interactions in One-Dimensional Space.

The first term is the expected benefit from social interactions and is the key term in emphasising the motive for movement towards a particular neighbour - as the distance between individuals i and j decreases the benefits increase significantly. Therefore this term is paramount in accounting for gains from proximity to others. In terms of the components of well-being - happiness and life satisfaction - this term expresses our hypothesis that social interaction effects, e.g. trust in one's neighbours, exert a significant effect on individual well-being to a greater extent than other exogenous elements. The second term in the utility function (1) is used to represent the 'intrinsic' value of an individual's choice of location. This could depend on the various exogenous geographical, institutional and political factors, z, which we later term as *exogenous contextual factors*, e.g. the quality of government, as well as a set of *individual socio-demographic indicators*, w_i, as described briefly in the introduction. Therefore, the first term in (1) represents a distortion on the choice of location which otherwise would have been purely determined on the basis of exogenous contextual and individual factors - which Akerlof refers to as a 'social optimum' and which we will discuss briefly below.

This model helps us understand the drivers behind regionalisation as it appears to be the most appropriate for justifying the ex post formation of regional groupings by suggesting that they exert relatively stronger effects on their populations. One of the strengths of this model is indeed in allowing for such a rich interpretation of movement across regions. In the case of Akerlof's model, the EU could be envisaged as a horizontal real line (see Figure 1). Along this version of the real line we can imagine that we have N individuals spaced along it at random locations, $x_{0,1}, x_{0,2}, ..., x_{0,N} \in X$ and $X \in \mathbb{R}$, where X is the set of all individuals along the real line.[6]

At some point along the line there is some form of social or continental optimum which serves as the reference point for all social exchanges.[7] In our regional example we could view such an optimum as a reference location where we find that the largest concentration

[6] See Appendix A for similar figure in a two-dimensional context.

[7] Akerlof's notion of an optimum suggests that any proximate movement is sub-optimal. However, given the utility function specified such a move is indeed optimal, assuming that an individual's well-being increases with proximity.

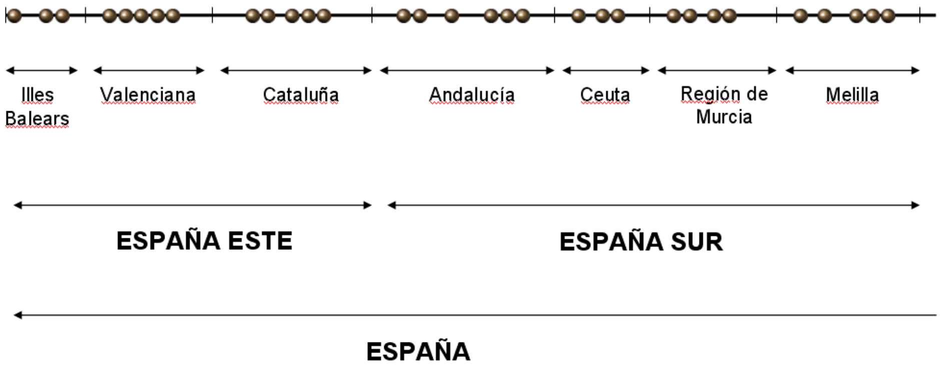


Figure 2: A One-Dimensional Representation of Socio-Geographic Clustering in Spain.

of people settle.[8] For example, certain parts of Europe might be optimally located for access to all regions. However, other than serving as the global optimum to the N optimisation problems, there is no need to provide an interpretation as it will only serve as a point of attraction to those individuals closest to it.

The most important implication of this model is that proximity to certain individuals exerts the greatest effect on social and geographical movement. At the most basic level these movements by individuals are described in Akerlof's model as being motivated by an actual 'need' for greater proximity with one's neighbours, but we take such needs for proximity as a proxy for the socio-economic, cultural and linguistic factors which have driven individuals into the present-day regional groupings we find across Europe. As such we obtain some form of justification that regions (corresponding to the Nomenclature of Territorial Units, or NUTS, classifications) could indeed have stronger effects on the individuals that make up their populations and that strong interaction effects can be identified within those regional populations. We could therefor expect a greater degree of dependence within these groups.

As was noted before, these ideas are supported in certain branches of psychology and suggest that we can therefore expect *a priori* some form of dependence between individuals who share the same regional location. The existence of social and cultural norms can ultimately be seen as the social and geographical confluence of individuals who are collocated in the same region. All this would ultimately be the theoretical translation of the works reviewed in the introduction that individuals are more responsive to those closest to them.

As an example, as Figure 2 shows, we consider two of the regions in Spain, East (Este) and South (Sur), at the NUTS1 level in a linear setting, which is then is subdivided further at the NUTS2 level into distinct principalities. Our model of social and geographic distance can therefore be shown to yield regional groups which over time have lead, if somewhat arbitrarily, to the statistical regions that have been applied to the 25 members of the EU.

Our next step is to extend Akerlof's model into two dimensions. This allows us to capture

[8] In Akerlof's model this was interpreted as some sort of socially optimal level of education.

both the effects of social distance and geography as reasons for individuals clustering into separate regions. Specifically, we augment the objective function for individual i to include two choice variables which represent the individual's choice of spatial coordinates[9]. Thinking in terms of a standard Cartesian coordinate system with both an x-axis and y-axis allows us to imagine individuals choosing their social location in a geographical sense, with their choice being affected once again by issues of social exchange as in Akerlof's one-dimensional case. The objective function is now as follows:

$$\begin{aligned} U_i &= \sum_{i \neq j} \frac{A}{(B + (x_{0,i} - x_{0,j})^2)(C + (x_{1,i} - x_{0,j})^2)} + \left[-ax_{1,i}^2 + bx_{1,i}\right] \\ &+ \sum_{i \neq j} \frac{E}{(F + (y_{0,i} - y_{0,j})^2)(G + (y_{1,i} - y_{0,j})^2)} + \left[-cy_{1,i}^2 + dy_{1,i}\right] + z + w_i \end{aligned} \quad (2)$$

where $(x_{0,i}, y_{0,i})$ denotes the coordinates for individual i in two-dimensional space in period 0. In the above expression, $(x_{0,i} - x_{0,j})$ and $(y_{0,i} - y_{0,j})$ denote inherited social positions in two-dimensional space and $(x_{1,i} - x_{0,j})$ and $(y_{1,i} - y_{0,j})$ are the social distances from one's neighbours. Therefore this allows us to motivate the model as one where a combination of these types of distances help to determine geographical groupings of individuals across Europe.

In addition we extend the model by motivating the idea of individuals being 'close' as being unhindered by significant barriers (e.g. administrative, linguistic or cultural) such that any form of movement towards a particular individual or group is feasible only if such obstacles are absent - this prohibited movement could be represented as some sort of limit for movement in a particular direction (e.g. choice of $x_{1,i}$ not greater than a certain threshold). We are therefore able to account for the cases where individuals who would tend to cluster together geographically are then somehow arbitrarily assigned to certain groups, such as regions, due to administrative and natural barriers. We provide a proof of the results of spatial collocation in a two-dimensional setting and with barriers to movement in Appendix A.

We illustrate our two-dimensional extension with the following diagrams. As we can see in Figure 3, the EU could be envisaged in $x-y$ space with N individuals scattered at random locations with coordinates $(x_{0,1}, y_{0,1}), (x_{0,2}, y_{0,2})..., (x_{0,N}, y_{0,N}) \in (X, Y)$ and $X, Y \in \mathbb{R}$, where X and Y are the sets of all individuals along the x-axis and y-axis respectively.

As we did before, we can also show a multilevel representation of the socio-geographic clusters. Figure 4 shows the three level approach which is taken in this paper and developed in the later sections. We have individuals clustered into regions, which are themselves grouped into countries. In this way we can simply characterise the effects on clustering of the various exogenous social interactions and contextual effects, as well as the social indicators, which were mentioned above.

[9] We have replaced U with W to denote specifically well-being.

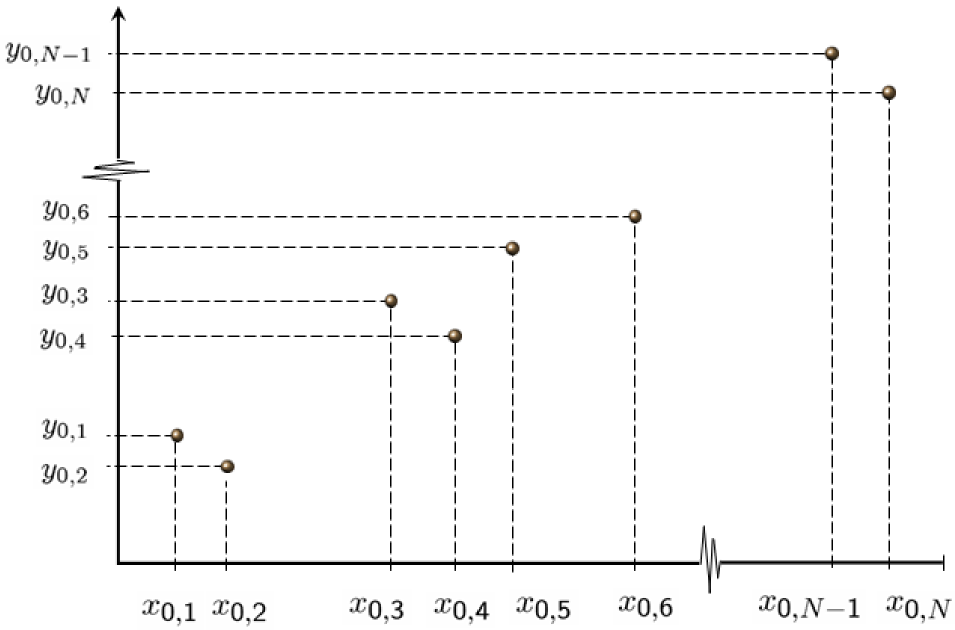


Figure 3: The Emergence of Social Groupings based on Social Interactions in Two-Dimensional Space.

3 A Role for Multilevel Modelling

As we have information at the European sub-regional and individual level we have assessed the determinants of well-being using a multilevel modelling approach. Multilevel modelling originated in the fields of educational research and epidemiology, emphasising the role of *hierarchical* dependence[10]. For example, in a typical household survey dataset, alongside an individual identifier, there may exist a number of additional flags representing household characteristics, residence by census tract, residence by region and so on. The principal issue here is that individual behaviour is determined by a combination of individual characteristics, together with the influence of household structure, residence, peer groups and other groupings effects. This implies that the group and its members can both influence and be influenced by the composition of the group (Goldstein (1998)).

Alongside the nested structure of the hierarchical data, increasing attention has been paid to different forms of interactions and externalities in the hierarchical system (Durlauf (2003), Manski (2000), Brock (2001)). For example, Durlauf (2004) analyses the determinants of poverty traps to show how the persistence in economic status is generated by group-level influences on individuals. The importance of such externalities has led researchers to define

[10] Over the past decade there has been a development of methods which have enabled researchers to model hierarchical data. Examples of these methods include multilevel models (see, for example, Goldstein (1998)), random coefficient models (Longford (1993)) and hierarchical multilevel models proposed by Goldstein (1986) based on iterative generalized least squares (IGLS).

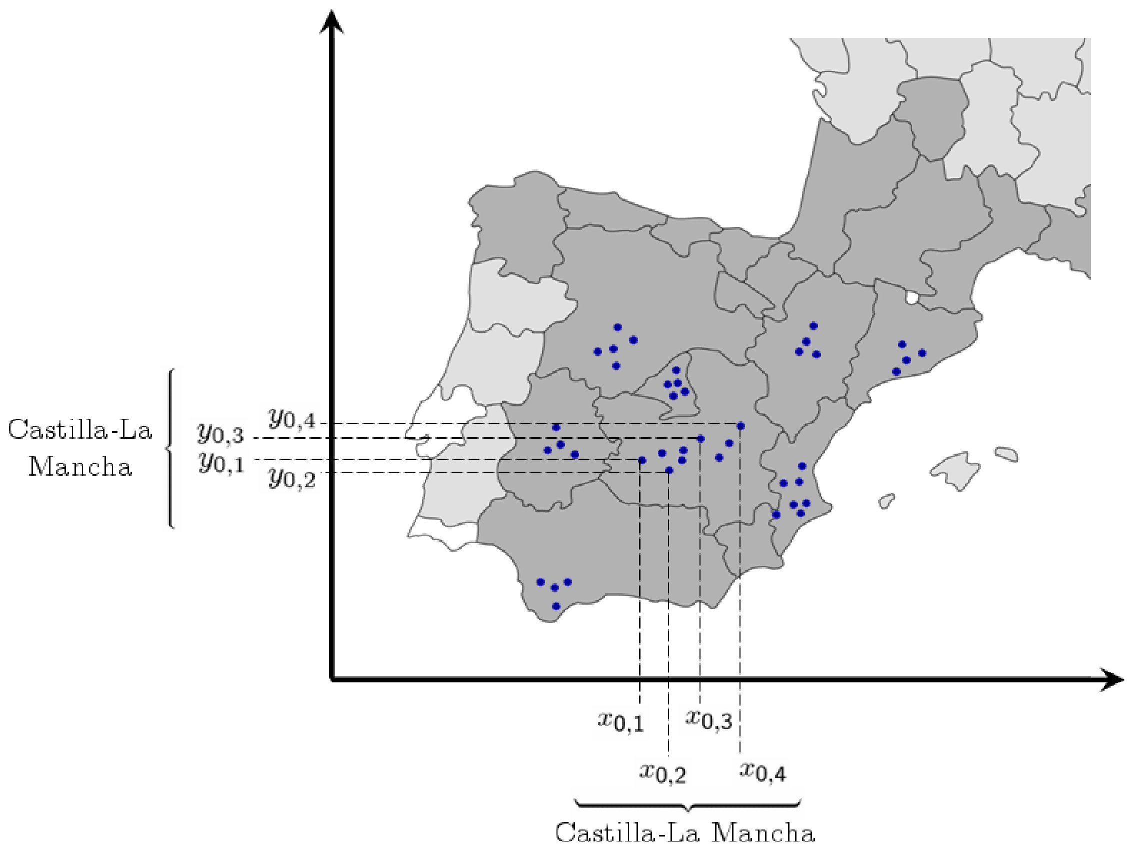


Figure 4: A Two-Dimensional Representation of Socio-Geographic Clustering in Spain showing three levels: Individual, Region and Country.

different concepts of membership and neighbourhood effects as those relying on notions of distance in social space analysed in the previous section (Akerlof (1997)).

There are a number of advantages of this approach. First, in standard unilevel OLS regression the presence of nested groups of observations may be handled by using dummy variables. However, this approach breaks down when there are a large number of levels resulting in a reduction in degrees of freedom. Second, this approach helps to analyse the effect of heterogenous groups in small samples. It is now widely recognised (Gosh and Rao (1994)) that survey estimates for small areas, for example, are likely to yield unacceptably large standard errors due to the smallness of sample size. Given the nature of our data, the proposed hierarchical multilevel method, which is sample size dependent, seems to have a distinct advantage over other methods in solving the bias.

Recognition of the different forms of interaction between variables which affect each individual unit of the system and the groups they belong to has important empirical implications. In fact, the assumption of independence is usually incorrect when data are drawn from a population with a grouped structure. The existence of grouping adds a common element to otherwise independent errors, thereby generating correlated *within*-group errors. Moulton (1986) examines the precision of regression estimates derived from grouped data and finds that it is usually necessary to account for the grouping either in the stochastic structure of the errors or in the specification of the regressors. It is also possible that errors *between*-groups will be correlated. For example, if the groups are geographical regions, as in our dataset, then neighbouring regions display greater similarity than regions that are distant. Moulton (1990) shows that incorrectly using Ordinary Least Squares (OLS) estimation, even with a small level of correlation, will lead to standard errors with substantial downward bias and dangers of spurious findings of statistical significance.[11]

Such group interactions should be properly modeled; they can be modeled either as fixed or random effects and in deciding which is the right approach particular attention should be devoted to the understanding of the survey methods used to generate the sample. A stratification method (Deaton (1997)) effectively converts a sample from one population into a sample from many populations and guarantees that there will be observations to permit estimates for all the geographical subgroups. If the data are grouped by area with all areas represented in the sample then a fixed effects specification is appropriate; this is often modelled in the mean equation through additive or multiplicative dummy variables and mean covariates.[12] We will examine in the next section how such a specification may in fact characterise more complex interaction effects in the hierarchical structure.

[11] Once we adjust for this dependence, the degrees of freedom will fall thereby increasing estimated standard errors, and correcting the downward bias resulting from the use of OLS estimation.

[12] When only some of the areas are represented in the sample or the pattern of dependence is unknown spatial effects can be accommodated in a hierarchical model through the error term by allowing for an unrestricted non-diagonal covariance matrix.

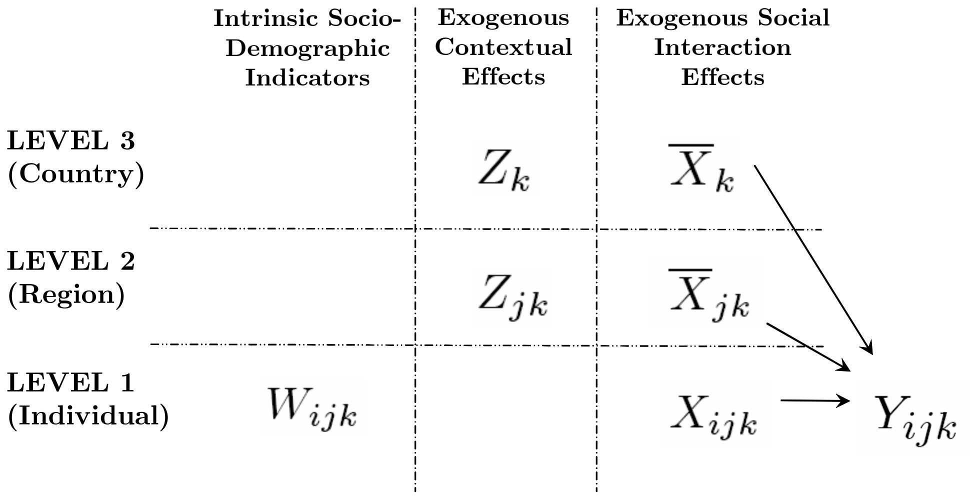


Figure 5: Diagrammatic Representation of a 3-level Model with Exogenous Contextual and Social Interactions Effects and Intrinsic Socio-Demographic Indicators

3.1 A Diagrammatic Representation

In light of the multilevel framework discussed in the preceding section, it is useful to understand the role of multilevel modelling in facilitating our analysis. As we are interested in the effect of individual, regional and national factors on the members of the various sub-populations, this particular branch of statistical testing is an obvious choice. A useful way to think about multilevel sampling for the case of a three-level model is described in Figure 5. The broken horizontal lines delineate the three levels: below the bottom two lines are the 'micro-levels' (regional or level-two and individual or level-one) and above the first line is the 'macro-level' (national or level-three). In this particular paper we are interested in the effect of 'level-three' variables Z_k, and $\overline{X}_k$, and 'level-two' variables, Z_{jk} and $\overline{X}_{jk}$, on the (micro) 'level-one' variable, Y_{ijk}, while controlling for other level-one variables, X_{ijk} and W_{ijk}.

Even if we were interested only in the effects of $\overline{X}_{jk}$ on $\overline{X}_k$ at the lower macro-level (or level-two), as long as these variables are not directly observable we would require two-stage sampling, e.g. $\overline{X}_{jk}$ and $\overline{X}_k$ measured as the average of the micro-level units across clusters, X_{ijk}. This is the the case *a fortiori* for variables which might explicitly be defined as the the aggregate of micro-level data from the outset.

4 Modelling Social Exchange in Multilevel Models

We now show how the effects of individual social exchange within the same group, as highlighted in the model developed in section two, can be modelled econometrically using simple interaction effects in a multilevel framework, as shown in Figure 5. An attractive feature of the proposed method is that it is possible to identify social interactions in a way that is robust to the presence of group-level heterogeneity (Brock (2001))[13].

Specifically, we wish to show how we can translate the theoretical model of section two into a testable empirical relation. In order to transform the theoretical model appropriately, we will concentrate on the two distinct parts of the 'well-being function' and reinforce our earlier discussion that the theoretical basis for considering social distance lends itself naturally to the concepts of multilevel modelling. As a result we will look at the two constituent parts of our well-being function:

$$Y = \textit{Expected Benefits from Social Exchange} + \textit{Intrinsic Value} \tag{3}$$

We choose to denote the different components of well-being (utility) by Y, as they are now the dependent variables in our regressions.

For modelling the expected benefits from social exchange, we would be interested to capture as many of the social interaction effects as possible. Manski (1993) listed three such effects which should ideally be considered when considering individual behaviour in a group context: endogenous effects, exogenous (contextual) effects, and correlated effects. Endogenous effects occur when the behaviour of the individual tends to vary with the prevalence of the behavior in their group. Exogenous effects occur when the individual's behaviour tends to be affected by the underlying characteristics of the group (or region) that they belong to, which have been exogenously determined. Finally, correlated effects occur when individuals in the same group tend to behave similarly because they share similar individual characteristics, e.g. ability, propensity to be happy, etc. Typically the latter effects are unobserved and so the only effects that can be explicitly controlled for are the endogenous and exogenous interactions effects, though we consider measures for excess variance that could account for the unobservable correlated group effects.

Therefore, in order to capture these effects, we would be interested to include in our relation terms which can account for the relative social distances of the individuals within each group, i.e. we would like our dependent variables, happiness and life satisfaction, to depend on their respective group means, as well as the deviations of individuals' characteristics from

[13] Typical examples of interaction-based models are the emergence of social networks or social norms. Other influences are the so called peer influence effects which have been examined in the psychology literature (Brown (1990) and Brown, Clasen, and Eicher (1986)). Other forms of social interactions are role models, in which the aspiration of a student are affected by the observed education/occupation outcomes among adults in his group (Streufert (1991)); imitation effects in consumption preferences where, for example, preferences depend on the observed consumption of neighbors (Bell (1995)). The role of herd behaviour and informational cascades, where agents attempt to learn more by observing the behaviour of others (Bikhchandani and Hirschleifer (1992)).

their respective group means. These deviation scores should capture on average their social positions relative to their neighbours. The first part of the regression relation in a two-level model would ideally therefore take the form

$$Y_{ij} = f\left(\bar{Y}_j, \left(X_{ij} - \bar{X}_j\right)\right) \tag{4}$$

where $\bar{Y}_j$ denotes the average at the group level for the dependent variable and $\bar{X}_j$ denotes the average at the group level for the explanatory variables, which would include the various social interaction effects we are interested in. In a three-level setting we would expect to include means at both the regional level (groups denoted by j) and the national level (groups denoted by k)

$$Y_{ijk} = f\left(\bar{Y}_{jk}, \bar{Y}_k, \left(X_{ijk} - \bar{X}_{jk}\right), \left(\bar{X}_{jk} - \bar{X}_k\right)\right) \tag{5}$$

Therefore a possible regression relation could take a linear form

$$Y_{ijk} = \alpha_0 + \alpha_1 \bar{Y}_{jk} + \alpha_2 \bar{Y}_k + \alpha_3 \left(X_{ijk} - \bar{X}_{jk}\right) + \alpha_4 \left(\bar{X}_{jk} - \bar{X}_k\right) \tag{6}$$

However, as highlighted by Manski (1993), this type of relationship suffers from a reflection problem[14] and a lack of identification (see Appendix B).

Therefore to specify the relation correctly, as well as allowing for some form of endogenous regional and national effects, we take the following steps. Let us consider once again a simple two-level model where level-one, captured by the subscript i, represents individuals and level-two, denoted by j, corresponds to regions. We start with the direct individual relationship as follows

$$Y_{ij} = \delta_0 + \delta_1 X_{ij} + e_{ij} \tag{7}$$

In this equation, e_{ij} represents individual heterogeneity, is assumed to be standard normal, and satisfies $E\left[e_{ij} \mid X_{ij}\right] = 0$. Y_{ij} is our dependent variable, and X_{ij} is a set of individual characteristics. Therefore, this is the simple regression of the relevant dependent variable on the individual's corresponding social attributes. We can transform this relation by taking the average of the micro-level units across groups to obtain the between-group regression

$$\bar{Y}_j = \alpha_0 + \alpha_1 \bar{X}_j + \nu_j \tag{8}$$

where α_1 is the between-group regression coefficient. However, as we are also interested in the regression of the within-group deviation scores, we can also consider the following regression within a particular group:

$$\left(Y_{ij} - \bar{Y}_j\right) = \gamma_1 \left(X_{ij} - \bar{X}_j\right) + \epsilon_{ij} \tag{9}$$

where γ_1 the within-group regression coefficient. This regression, once rearranged, allows for the mean of the dependent variable to enter the relation with a fixed coefficient of 1. In

[14] Manski (1993) writes about the 'reflection problem', whereby researchers are limited in their ability to account for the full range of social interactions by taking account of the endogenous effects of a particular group on the individuals of that group, which are assumed to be modelled by including the mean of the dependent variable as an explanatory variable in the regression equation.

addition, once we substitute in for the group mean, $\overline{Y}_j$, we obtain the following multilevel model in terms of means and deviation scores:

$$\begin{aligned} Y_{ij} &= \bar{Y}_j + \gamma_1 \left(X_{ij} - \bar{X}_j \right) + \epsilon_{ij} \\ Y_{ij} &= \alpha_0 + \alpha_1 \bar{X}_j + \gamma_1 \left(X_{ij} - \overline{X}_j \right) + \nu_j + \epsilon_{ij} \end{aligned} \tag{10}$$

Therefore this would appear to be the relevant relation for capturing as many of the endogenous and exogenous interactions effects as possible.

For the sake of notational ease, we rewrite the above model with a more consistent set of coefficients:

$$Y_{ij} = \beta_{0j} + \beta_1 X_{ij} + \beta_2 \left(X_{ij} - \overline{X}_j \right) + u_{ij} \tag{11}$$

whereby the within-group regression coefficient is now $\beta_1 (= \gamma_1)$, the between-group regression coefficient is now $\beta_2 (= \alpha_1)$, β_{0j} is a group-specific intercept, and $u_{ij} = \nu_j + \epsilon_{ij}$.

Continuing with the development of the model, we chose to introduce random intercepts in addition to the inclusion of means and deviation scores. These intercepts are modelled as follows:

$$\beta_{0j} = \beta_{00} + \beta_{01} Z_j + u_{0j} \tag{12}$$

The random intercept shows that the heterogeneity across the level-two groups is explained by a set of observable group characteristics, Z_j and a random disturbance term, u_{0j}. This would imply that we were attempting to 'explain' the between-group *variability* (as captured by the group-specific intercepts, β_{0j}) by including a level-two covariate, Z_j and a mean intercept, β_{00}, with u_{0j} crucially representing the unobserved level-two effects. In order to complete this particular specification we define $\Sigma_1 \equiv \text{Var}(u_{ij})$ and $\Sigma_2 \equiv \text{Var}(u_{0j})$, and make the following assumptions:

$$\begin{array}{ll} u_{ij} \mid X_{ij} \sim N(0, \Sigma_1) & \text{Cov}(u_{ij}, u_{i'j}) = 0, \ \forall i \neq i' \\ u_{0j} \mid X_{ij} \sim N(0, \Sigma_2) & \text{Cov}(u_{0j}, u_{ij}) = 0 \quad . \end{array}$$

The reduced form of the model therefore becomes

$$Y_{ij} = \beta_{00} + \beta_{01} Z_j + \beta_1 X_{ij} + \beta_2 \left(X_{ij} - \overline{X}_j \right) + u_{0j} + u_{ij} \tag{13}$$

and in terms of social interactions, we can interpret the coefficients as follows: β_1 is a measure of the direct impact of the exogenous individual factors on individual well-being; β_2 is the *social multiplier* associated with the exogenous group social interactions; β_{01} accounts for the exogenous contextual factors, while u_{0j} captures the unobservable correlated effects, with Σ_2 providing a measure for them[15]. The endogenous social interaction effects cannot unfortunately be explicitly disentangled but are implicitly allowed for by the inclusion of the mean covariates, as was highlighted in the derivation above. Equation (13) is the testable counterpart of the first term in the right hand side of equation (3) at the start of this section.

[15] By allowing for the possibility that the conditional mean of group effect and the individual effect vary with group size, Graham (2004) also allows for the possibility that endogenous contextual effects may differ according to group size, being stronger in bigger regions.

For the time being we chose to leave out random coefficients and assumed that the within-group coefficients were the same across groups, at both the national and regional level. In addition, to avoid further problems of endogeneity, we have chosen our exogenous contextual factors, Z_j, to ensure that they are uncorrelated with the explanatory variables, X_{ij}.

Finally we need to capture the intrinsic factors of choice of location on social exchange (the second term on the right hand side of equation (3)). These were individual socio-demographic characteristics which might have somehow determined proximity and corresponding levels of happiness and life satisfaction in such a way as to be uncorrelated with the other explanatory variables, X_{ij}. These additional individual-level factors were included in an additive fashion and denoted W_{ij}. Their inclusion should ideally help to improve the power of our multilevel model in explaining the regional and national determinants of well-being. Therefore the full two-level model we are implementing takes the form

$$Y_{ij} = \beta_{00} + \beta_{01}Z_j + \beta_1 X_{ij} + \beta_2 \left(X_{ij} - \overline{X}_j\right) + \beta_3 W_{ij} + u_{0j} + u_{ij} \tag{14}$$

In the Appendix D we show a generalisation of the two-level model to the three-level case.

Note that one objection to the estimation of (14) is that X_{ij} and $\left(X_{ij} - \overline{X}_j\right)$ may be correlated. As noted also by Hannan and Burstein (1974), when the clusters involve natural groupings, such as neighbourhoods, these groupings may be affected by certain factors (characterised by X_{ij}) that could be correlated with some of the exogenous characteristics of the group, $\left(X_{ij} - \overline{X}_j\right)$. Therefore membership does not tend to be random with respect to those characteristics and a high correlation between X_{ij} and $\left(X_{ij} - \overline{X}_j\right)$ is expected. However, we are considering the case where individuals who cluster together geographically are then assigned to certain groups, such as regions and other administrative units. We expect them to be more homogeneous due to frequent interactions with one other and through sharing relevant 'life experiences'. Hence membership could be random with respect to the exogenous characteristics described by $\left(X_{ij} - \overline{X}_j\right)$ and therefore we expect the unobservable correlated effects to have a greater effect than these characteristics.

To identify such unobservable social interactions across groups we measure the *excess variance*, defined as the ratio of the unconditional between-group and within-group variances at the different levels of the hierarchical structure (Graham (2004)). In the two-level case this measure is:

$$EV = \frac{\Sigma_2}{\Sigma_1} \quad .$$

Without social interactions ($\beta_2 = 0$) the reduced form simplifies to the one-way error component model:

$$Y_{ij} = \beta_{00} + \beta_{01}Z_j + \beta_1 X_{ij} + \beta_3 W_{ij} + u_{0j} + u_{ij} \tag{15}$$

that could represent the social optimum without social-exchange as identified by Akerlof (1997). Therefore the effect of social interactions is to generate excess between-group variance by magnifying the effect of unobservable group-level factors, u_{0j} and mean group characteristics, $\overline{X}_j$, on outcomes through the social multiplier β_2. Within a multilevel model it is then easy to test for the different forms of social exchange identified by Akerlof since we could establish how the level of well-being of an individual is affected by these observable social interactions and by unobservable group-level effects. The latter factors could be easily identified by measuring the excess between-group variance after controlling for the exogenous and individual characteristics of each group member (Graham (2004), Glaeser, Sacerdote, and Scheinkman (2003)).

5 The Determinants of Regional Well-Being across Europe

Our regional dataset is derived primarily from the responses to the European Social Survey (ESS). This survey was developed to conduct a systematic study of changing values, attitudes, attributes and behaviour patterns within European polities. Its aim is to feed into key European policy debates, by measuring and explaining how people's social values, cultural norms and behaviour patterns are distributed, the way in which they differ within and between nations and regions, and the direction and speed at which they are changing. The survey consists of a collection of core[16] and rotating[17] modules. The former is repeated at each round of the survey and remains relatively constant between rounds, while the latter modules are repeated at greater and varying intervals.

The problem, of course, with developing effective indicators for such subjective and notional concepts concerning the European social fabric is the absence of a comprehensive, well-tested and analytically-powerful set of tools for measuring underlying values across nations. Although the Eurobarometer, the European (and World) Values Surveys and the International Social Survey Programme have all made major contributions, even their combined lists of individual items and, more importantly their combined array of validated scales, were not considered comprehensive enough for effective policy formulation. Hence the development of the ESS.

Specifically, the core components of the modules cover:

[16] These three categories make up sections A, B and C of the survey and as they form the core module, responses to these questions are collected at each round. The rotational modules were omitted for the time being. As there had only been two rounds at the time we started to collate the data, the different sets of rotational modules had only been used once on each round and not repeated (some asked in 2002 and others asked in 2004).

[17] There have been three rounds thus far: Round 1 was completed in December 2002, Round 2 in December 2004 and Round 3 in December 2006. The rotating modules in Round 1 covered 'Citizenship, Involvement and Democracy' and 'Immigration'. Those in Round 2 were 'Family, Work, and Well-being'; 'Opinions on Health and Care Seeking', and 'Economic Morality in Europe: Market Society and Citizenship'. The final rotating modules in Round 3 were 'Personal and Social Well-being: Creating indicators for a flourishing Europe' and 'The Timing of Life: The organisation of the life course in Europe'. In addition the questions were altered and rearranged between rounds which required careful matching to ensure the datasets contained consistent and matching categorical variables.

(i) People's value orientations (their world views, including their religiosity, their socio-political values and their moral standpoints)

(ii) People's cultural/national orientations (their sense of national and cultural attachment and their related feelings towards outgroups and cross-national governance)

(iii) The underlying social structure of society (people's social positions, including class, education, degree of social exclusion, plus standard background socio-demographic variables, such as age, household structure, gender, etc., and a few questions about media usage that help to identify the primary sources of people's social and political knowledge).

The ESS data was complemented by a variety of other economic, socio-demographic and geographical regional indicators which were obtained directly from the Study Programme on European Spatial Planning (SPESP). For a complete variable list see Table 1. The individual data extracted is grouped by NUTS[18] regions (161 in 2004 and 182 in 2002).[19] As a result the variables employed in the analysis consist of a number of indicators divided into national, regional and individual level variables. The individual responses are treated as level-one, the NUTS1 and NUTS2[20] regions make up level-two, and the countries (NUTS0) make up level-three, giving us a multilevel model with three levels.

In line with the multilevel structure in equation (14) we have identified a set of explanatory variables for each level. We have divided *national* variables into exogenous contextual effects comprising the country's quality of government and average log per-capita income and exogenous social interaction effects comprising the average degree of religiosity, average trust and the average political interest. All these variables are national averages from level-one micro-units.

At the *regional* level the exogenous contextual factors comprise the geographical location of the region and mean log per-capita income. The exogenous social interaction effects comprise the average degree of religion, average trust and the average political interest. These variables are similarly derived as regional averages from level-one micro-units.

The *individual* level variables have been grouped under the heading of exogenous social interaction and socio-demographic effects and subdivided according to[21]:

(a) personal functioning/feelings comprising self-reported physical and mental health and the degree of religiosity.

(b) inter-personal feelings, i.e. the trust in a country's parliament, the police force, people and the respondent's perception of people's level of altruism.

(c) inter-personal functioning, i.e. the level of social engagement, social intimacy and the level of political involvement.

(d) social exclusion and socio-demographic factors comprising: ethnicity, citizenship, marital status, educational level, victim of crime, employment status, gender and household

[18] For a similar treatment of NUTS data see Corrado, Martin, and Weeks (2005).

[19] The full list of NUTS regions used for the multilevel estimation is available on request.

[20] An important point to note is that due to the nature of the data, the regional breakdown across the NUTS1 and NUTS2 definitions is not consistent, i.e. for some countries the survey data is reported on a NUTS1 basis and for others on a NUTS2 basis.

[21] Following Elliott and Umberson (2004) and Huppert, Baylis, and Keverne (2005).

income.

Table 1 provides a complete specification of the variables included in the regressions. In addition, Table 2 provides a correlation matrix for these variables. Unsurprisingly the trust variables are mildly correlated. However, the political variables ('political interest', 'trust in parliament', etc.) are the most correlated with one another. Oddly, the variables 'trust in police' and 'victim of crime' show close to zero correlation. On the whole the pairwise correlations reveal that the explanatory variables are little correlated which reduces the risk of multicollinearity in our results and helps to identify the composite parameters embedding social interaction effects in the multilevel model.

6 The Geography of Happiness and Life-satisfaction across Europe

Before moving to the estimation section it is instructive to analyse the geography of happiness and life-satisfaction in Europe by looking at some of the statistics from our dataset.

Figures 7, 8, 9 and 10 report the regional average level of happiness and life satisfaction in 2002 and 2004. Regions with a darker shade indicate higher levels of happiness and life satisfaction, measured on a scale between zero and ten. The first thing to note is that in general, the EU nations are relatively happy and satisfied with life, as can be seen from the fact that only values of five and above are reported in the ESS. Looking at the actual distribution, it appears that the levels of both happiness and life satisfaction are higher in Northern Europe than in Central Europe and some of the Mediterranean regions. Traditional stereotypes suggest that Europeans that hail from the warmer Mediterranean regions are more content. However, our inspection of the data reveals that in fact the regions of North and East France, Portugal and South Italy are on average the least satisfied and happy in both waves. With particular regard to life satisfaction, West Germany, reports lower levels in 2004, while in Northern Ireland and Southern France the level of life satisfaction has increased in the intervening period. With respect to happiness, most regions in the UK (except the Eastern regions), Central Spain and Austria report a lower level of happiness between the two waves.

In addition, we have also organised the national averages for happiness and life satisfaction, in the manner of country rankings for both 2002 and 2004 in Figure 11. If we turn to these charts, we find that overall for both 2002 and 2004 Italy, Greece, Portugal, Germany and France report the lowest levels of happiness whereas the Scandinavian Countries, Netherlands and Luxembourg report the highest. A similar trend is found in life satisfaction. We hope to understand what could be driving such relative differences in well-being across the different EU regions.

In fact, we might be able to shed some light on this matter by considering how the exogenous contextual and social interaction effects at the country level ($\mathbf{Z}_k$ and $\bar{\mathbf{X}}_k$ respectively) differ across the EU-15. As Figures 12, 13 and 14 show the Northern EU countries (the

Scandinavian members, Luxembourg, the Netherlands) are attaining the highest scores in terms of quality of government, as measured by the World Bank Governance Indicators, as well as reporting the highest level of per capita income. Respondents also report the highest level of trust in their respective parliaments, peoples, and laws, as enforced by their police forces. Similarly they also report the highest level of political interest and perceived altruism. The worst scoring countries in the majority of these categories are once again the Southern European nations of Italy, Portugal and Greece, which could explain their relatively lower levels of well-being.

In addition, we find that Portugal and the Scandinavian EU members have the highest level of social engagement, while Greece, Italy, and Germany register the lowest. However, these results are reversed when we turn to the reported level of social intimacy where Italy, Belgium and France score the highest while Germany, Netherlands and Denmark the lowest. This inverse relation between social intimacy and engagement, though not obvious, could be explained by the differing nature of social and familial bonds across different parts of Europe. If we consider Hofstede's cultural dimensions, Northern Europe has traditionally been representative of individualism, whereas individuals from Central and Southern Europe are typified by the strong collectivist. Nevertheless, Germany stands out as the one country that reports the lowest level of both social engagement and social intimacy.

In terms of average national religiosity, the Southern European nations dominate, with Greece, Italy and Portugal scoring highly, while the Central and Northern European countries such as Sweden, France and Luxembourg report religious belief to be far less prevalent. One final point to make is that there is no obvious continental divide, as we have found up to this point, when we look at average health across the EU-15. Portugal, Germany and Spain report the highest levels of health, while Denmark, Ireland and Austria report the lowest levels.

Having established a variety of interesting, if somewhat familiar, trends in our data for the EU-15 countries, we now proceed to attempt to disentangle and identify the actual determinants of both life satisfaction and happiness. As we have highlighted before, we want to assess the role played by contextual (e.g. institutional) and social interaction effects at the national and regional level in explaining well-being at the individual level. As an input to this analysis, we also consider the role played by individual intrinsic socio-demographic characteristics in affecting both life satisfaction and happiness. The next section estimates a three-level model and discusses our main findings.

6.1 Estimation and Results

Following the three-level specification of equation (14) we define the following three-level multilevel model:

$$
\begin{aligned}
\mathbf{Y}_{ijk} &= \boldsymbol{\beta}_{000} + \mathbf{Z}_{jk}\boldsymbol{\beta}_{010} + \mathbf{Z}_k\boldsymbol{\beta}_{001} + \left(\mathbf{X}_{ijk} - \bar{\mathbf{X}}_{jk}\right)\boldsymbol{\beta}_{11} && (16)\\
&\quad + \left(\bar{\mathbf{X}}_{jk} - \bar{\mathbf{X}}_k\right)\boldsymbol{\beta}_{12} + \bar{\mathbf{X}}_k\boldsymbol{\beta}_{13} + \mathbf{W}_{ijk}\boldsymbol{\beta}_2 + \mathbf{e}_{0jk} + \mathbf{e}_{00k} + \mathbf{e}_{ijk} && (17)\\
\mathbf{e}_{ijk} &\sim N(0, \Sigma_1),\ \mathbf{e}_{0jk} \sim N(0, \Sigma_2), \text{ and } \mathbf{e}_{00k} \sim N(0, \Sigma_3)
\end{aligned}
$$

where level-one denotes individual, level-two denotes regions and level-three denotes nations.

In equation (16) $\mathbf{Y}_{ijk}$ is our relevant measure of well-being; $\mathbf{X}_{ijk}$ is the set of individual social interactions - comprising personal and inter-personal feelings/functioning; $\mathbf{W}_{ijk}$ is the set of individual socio-demographic indicator variables; $\left(\mathbf{X}_{ijk} - \bar{\mathbf{X}}_{jk}\right)$ and $\left(\bar{\mathbf{X}}_{jk} - \bar{\mathbf{X}}_k\right)$ are the exogenous centred social interaction effects at level-two and level-three, and $\mathbf{Z}_{jk}$ and $\mathbf{Z}_k$ are the exogenous contextual factors at level-two and level-three. The last three terms denote the random components at the individual, $\mathbf{e}_{ijk}$, regional, $\mathbf{e}_{0jk}$ and national, $\mathbf{e}_{00k}$, levels. Finally, $\Sigma_1 = \sigma_1\mathbf{I}$ is the variance-covariance matrix for the level-one random effects, $\mathbf{e}_{ijk}$; $\Sigma_2 = \sigma_2\mathbf{I}$ is the variance-covariance matrix for the level-two random effects, $\mathbf{e}_{0jk}$, and finally $\Sigma_3 = \sigma_3\mathbf{I}$ is the variance-covariance matrix for the level-three random effects, $\mathbf{e}_{00k}$.

For the estimation of (16) we used a programme which has become commonplace across statistics and biostatistics for implementing multilevel modelling, MLWiN, which implements an Iterative Generalized Least Squares (IGLS) method[22]. The procedure is briefly set out in the Appendix E.

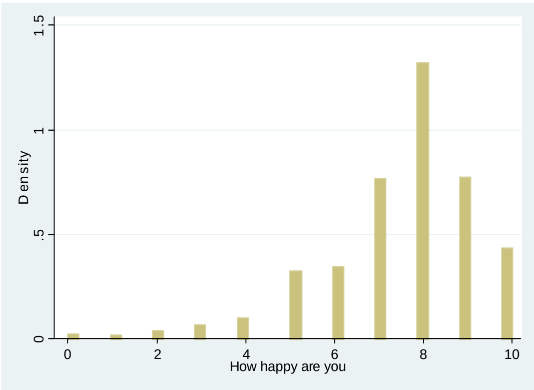
Given the problems highlighted in section four, we are interested in understanding the role played by the exogenous group effects as well as by correlated group effects. The presence of unobservable group effects can be identified by measuring the *excess variance*, defined as the ratio of the unconditional between-group and within-group variances at the different levels of the hierarchical structure (Graham (2004)):

$$
EV_{32} = \frac{\Sigma_3}{\Sigma_2}; \qquad EV_{21} = \frac{\Sigma_2}{\Sigma_1}.
$$

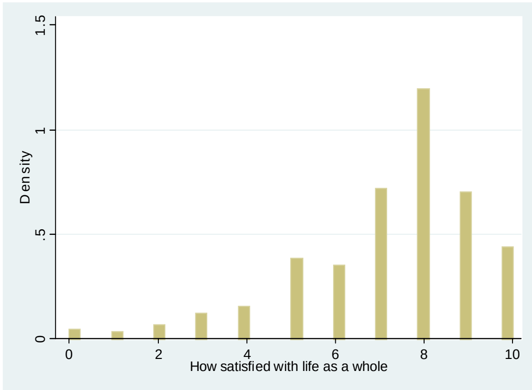
Table 3 reports the multilevel linear estimation results for waves 2002 and 2004 of the ESS for our dependent variables happiness (taken as a proxy for short-term well-being) and life satisfaction (our proxy for long-term well-being). However, it is important to bear in mind, that the choice of specific concepts within the field of well-being, and the choice of specific items to measure these concepts has proved to be a major challenge as a consensus has not yet developed among psychologists about the components of well-being or what would constitute the 'gold standard' for measuring well-being. In addition, as the distributions show, the two categorical dependent variables are approximately normally distributed. So following Helliwell and Putnam (2005) we have transformed both variables to Normal scores

[22]We employ a Restricted Iterative Generalized Least Squares (RIGLS) method. One advantage of the latter method is that, differently from Iterative Generalized Least Squares, estimates of the variance components take into account the loss of the degrees of freedom resulting from the estimation of the regression parameters. Hence, while the IGLS estimates for the variance components have a downward bias, the RIGLS estimates do not.

and use linear multilevel estimates[23]. For each response category, this transformation assigns the value from the inverse of the Standard Normal $(0,1)$ cumulative distribution for the estimated proportion of respondents from the dependent variable's original distribution[24].



Distribution of Happiness



Distribution of Life Satisfaction

Looking first at the estimation results for our short-term measure of well-being, reported subjective happiness, we find that a number of our level-three, level-two and level-one variables are significant, which suggests immediate support for our hypothesis that both regional and national factors have effects on an individual's well-being. In terms of the original aim of the paper, we are interested to verify whether, out of those coefficients that are significant at the national and regional level on the same variables, the following conditions hold

$$\boldsymbol{\beta}_{12} > \boldsymbol{\beta}_{13} \text{ and } \boldsymbol{\beta}_{010} > \boldsymbol{\beta}_{001}$$

where $\boldsymbol{\beta}_{12}$ is the vector of coefficients on the regional social interaction effects, $\bar{\mathbf{X}}_{jk}$; $\boldsymbol{\beta}_{13}$ is the vector of coefficients on the national social interaction effects, $\bar{\mathbf{X}}_{k}$; $\boldsymbol{\beta}_{010}$ is vector of coefficients on the regional contextual effects, $\mathbf{Z}_{jk}$, and $\boldsymbol{\beta}_{001}$ vector of coefficients on the national contextual effects, $\mathbf{Z}_{k}$. That is, do we find that regional factors have a greater effect on an individual's well-being as opposed to the national position, e.g. does regional average trust in people affect well-being more than the corresponding national average?

Out of the regional and national factors we find that in 2002 that there are far more regional contextual and social interaction effects than national effects which are significant.

[23] An alternative way would be to use multinomial ordinal probit estimation, which returns the effects on the underlying latent index. Indeed, the next stage for extending the modelling framework is to incorporate a latent response formulation. There are a number of references in this area which we plan to apply, e.g. Skrondal and Rabe-Hesketh (2004). Nevertheless, as shown by Helliwell and Putnam (2005), linear estimates for well-being were equivalent to such probit estimation and much easier to analyse.

[24] The Normal score transformation is defined as:

$$s = \phi\left(\frac{r}{n+1}\right)$$

where s is the normal score for an observation, r is the rank for that observation, n is the sample size and $\phi(p)$ is the p-th quantile from the standard normal distribution (see Conover (1999)).

In 2004, we find that the number of regional factors that are significant are only marginally greater than those at the national level. For example, in 2002, we find that the level of happiness is affected by one of our key economic indicators - regional GDP - as well as the regional averages of trust in the local police force, of political interest, religiosity and altruism, whereas only average political interest is significant at the national. There is only one variable for which both the regional and national averages are significant, which is in the case of average political interest in the multilevel regression of happiness in 2002. In this instance, we do find that $\beta_{12} > \beta_{13}$, which implies that the degree of political interest at the regional level exerts a stronger effect and is therefore more important in determining the level of happiness as compared with the national average political interest. In 2004, we find that individual happiness is affected by a wider mix of national and regional variables, which overlap only in the case of two variables (social intimacy and altruism). In the case of social intimacy we actually find that $\beta_{12} < \beta_{13}$, i.e. the regional effect is weaker than the national effect, but the desired relation is confirmed when we look at regional and national altruism. However, despite these findings at the second and third levels of the data, the results suggest that, overall, regional and national forces are far less important relative to each individual's intrinsic attributes and social position.

In fact, the 2004 results suggest that the hypothesis that regional factors are more important to individual short-run well-being than national aggregates, as typified by the majority of data collated by national governments, is not as robust as we would have expected. However, they do hold strongly for 2002. This could suggest the presence of some form of shift in behaviour between the two surveys, something which requires further analysis. We are currently integrating the 2006 data to see whether it replicate the trends of 2002 or 2004 or whether that data has its own unique results. Nevertheless, in 2004, despite national averages not being less important than regional variables, it is certainly *not* the case that they are any more important. In fact, as was expected, the estimation results suggest that it is indeed primarily individual intrinsic effects and a certain degree of unobservable correlated group effects that drive the levels of well-being across the EU-15. To further analyse this point we move to the analysis of level-one variables.

Looking at exogenous social interaction effects the degree of religiosity raises happiness. Inter-personal feelings, i.e. the quality of interactions with others, as measured by social intimacy and engagement, are also important. In addition, individuals who are relatively more confident in their country's institutions, who trust people more and who perceive society to be altruistic, report a higher level of happiness.

The intrinsic socio-demographic indicators included in the estimation were household income and a variety of dichotomous variables. These variables were introduced to establish whether descriptive factors such as the respondents' level of schooling or marital status were also significant in explaining happiness. The effects of income are significant only for the highest income levels and these income levels have a positive effect, as would be expected. We also found negative effects from being unemployed. However, the level of education did not appear to be significant, whereas marital status was significant only for those individuals who were in some form of relationship (cohabiting or married). With respect to the effect of the respondent's age we appear to recover the familiar inverted U-shaped relation between happiness and age. We find that younger and older individuals are happier than those who are middle-aged.

The socio-demographic variables related to social inclusion are also important. We find that respondents who are citizens of the country where they reside and who do not belong to ethnic minorities, report higher levels of happiness. Also, for 2002, respondents who have been a victim of crime report lower levels of happiness, which would be expected.

When we turn to the estimates for our longer-term measure of well-being, reported subjective life satisfaction, we find that, in 2002, significant national exogenous factors, both contextual and interactions-based, are completely absent. However, in 2004 we find that more national and regional averages are significant though we retain the same degree of relative importance between them. In addition, the significance of the regional averages is once again marginally more prevalent. For 2004, this greater level of significance (relative to the 2004 happiness estimation results) could be explained by viewing national and regional averages as a useful indicator of not only current moods and mindsets, but also future trends, since social and cultural 'norms', as captured by these averages, take time to develop and eventually to diminish. Therefore they could exert a slightly greater influence on our projectual measure of well-being, i.e. life satisfaction. If we were to include data from surveys of expectations, we might find far more significance among such variables - this is indeed one extension we would like to carry out. For example in 2004, we find that the regional averages such as those for health, the level of social intimacy and trust in police are significant. And at the national level, we find that the level of religious belief and the levels of social intimacies and engagement are all significant.

We also find that location becomes significant for life satisfaction in 2004, whereas it is not relevant in the determination of happiness. This could be justified by appealing to the case of geographical immobility as a long-term factor, such that individuals who are unable to move easily out of particular locations could find that their location can indeed affect their long-term well-being. For example, looking at the size and magnitudes of coefficients, we find in 2004 that living on the Northern or Mediterranean regions of your country can affect one's life satisfaction relatively more than if that individual lived elsewhere in their country. This partially supports the idea that Mediterraneans should be happier, but as we have seen this locational variable is not enough.

In addition, we find that the level-one variables are just as significant as in the case of happiness. The magnitudes of certain level-one coefficients are greater, such as that on health, which is understandable given the longer-term effects of the state of one's health. However, on balance, the level-one variables exert a much greater effect on happiness than life satisfaction, which supports our idea that the snapshots of society captured in each wave have greater short-run implications than long-run.

If we look at the individual social interaction effects, we find that several of these are significant and share similar signs and magnitudes across 2002 and 2004. We see that the level of trust in parliament, police and people, as well as the perceived level of altruism, has a strong positive effect on life satisfaction. Similarly, individuals who report a high level of social engagement and who are socially intimate with at least one other person are positively affected. Also respondents who report higher levels of health share strong positive effects on their life satisfaction. Moving on to the individual intrinsic socio-demographic factors, we find again that younger and older individuals are more satisfied with life than those who are middle aged.

As would be expected, the change in life satisfaction increased across income levels. On the other hand, income was only significant for the highest income brackets and was is more relevant in determining happiness than life satisfaction. There were also significant negative effects from being unemployed. Marital status was significant once again only for those individuals who are cohabiting or married. Finally respondents who are citizens of their country of residence and who do not belong to ethnic minorities once again report higher levels of life satisfaction.

It is also worth briefly mentioning some of the particular differences in the regression of happiness and life satisfaction. As certain inter-personal feelings and functioning have a more projectual perspective, we would expect that they will affect life satisfaction more than happiness, which is what we find in the case of social engagement and certain trust variables, e.g. trust in parliament and people. This can be seen by the relative size of the coefficients.

An important point to note is that the relative magnitudes of those explanatory variables that are significant do not change over the two waves. This suggests that the importance of certain factors affecting happiness and life satisfaction have not changed in the intervening years, which would be expected given only a two year break between the waves of the survey and the fact that the EU-15 are firmly entrenched in a period of relative stability and prosperity.

Having controlled for all the social interactions effects and intrinsic socio-demographic factors, we now consider the presence of the unobservable correlated effects that we discussed in the previous sections. In general, as Table 3 shows, there is evidence of both between-country variation, Σ_3, and between-region variation, Σ_2, suggesting the presence of spatially proximate spillover effects generated by unobservable effects operating at the regional and national levels.

There is also evidence of a relevant within-region heterogeneity across individuals, Σ_1 in all the four estimations reported. The high excess variance given by the ratio of between-region variance, Σ_2, and within-region variance, Σ_1, shows that well-being is strongly dependent on unobservable social interactions among individuals belonging to the same group. And more importantly life satisfaction, i.e. the way each individual perceives his whole life experience, is substantially more dependent on this latent measures of social exchange than happiness is in 2002.

7 Conclusion

With the emergence of social interactions and social exchange models (Manski (2000), Brock (2001), Akerlof (1997)) research has gradually moved from a pure spatial definition of neighbourhood towards a multidimensional measure based on different forms of social distance and interactions. Neglecting such interactions is likely to create problems of inference since this adds a common element to otherwise independent errors.

This work extends the measurement of well-being across European regions in several ways. First, using hierarchical data on well-being its adopts a three-stage multilevel model incorporating a variety of effects: exogenous contextual effects at the regional and national levels; a number of exogenous social interactions effects at all three levels of the model, and finally a series of intrinsic socio-demographic indicators at the individual micro-level.

Together with the excess between-group variance generated by social exchange and interactions, these effects are used to determine the face of individual well-being across the EU-15 nations. One of the main contributions of our empirical methodology is to show how social interactions can be easily modelled in a multilevel setting. Empirical findings support the idea that happiness and life satisfaction are strongly dependent on a variety of observable social interactions within each group as well as on other contextual and intrinsic factors.

As one of the main focus in the European policy arena is social cohesion, this paper clearly shows that social interactions are important for understanding well-being and subsequently must be considered when developing adequate social and economic policies. For example, among some of the more specific factors that we found to affect well-being, e.g. different forms of trust, the implication was that it is important for policies somehow to engender such trust if social cohesion is to be achieved.

The results in this paper are necessarily preliminary. Our aim is for this paper to form the start of a much larger body of research which seeks to explore this growing area of 'lifestyle analysis', namely the use of economic and social indicators to understand the role and effects of current social, political and economic policies and institutions on individuals across Europe. The third round of the ESS has recently been completed in December 2006 and once it is available we plan to integrate its findings into our work.

References

Aitkin, M., and N. T. Longford (1986): "Statistical Modelling Issues in School Effectiveness Studies (with Discussion)," *Journal of the Royal Statistical Society A*, 149(1), 1–43.

Akerlof, G. A. (1997): "Social Distance and Social Decisions," *Econometrica*, 65(5), 1005–1028.

Bell, A. (1995): "Dynamically Interdependent Preferences in a General Equilibrium Environment," Mimeo (Department of Economics, Vanderbilt University).

Bikhchandani, S., and I. Hirschleifer, D. and Welch (1992): "A Theory of Fads, Fashion, Custom and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, 992–1026.

Blanchflower, D., and A. Oswald (2004): "Wellbeing over Time in Britain and the USA," *Journal of Public Economics*, 88, 1359–1386.

Brock, W. A. and Durlauf, S. N. (2001): "Interactions-Based Models," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5, pp. 3297–3380, Amsterdam. Elsevier Science, Volume 5.

Brown, B. (1990): "Peer Groups and Peer Cultures," in *At the Threshold*, ed. by S. Feldman, and G. Elliott, Cambridge, MA. Harvard University Press.

Brown, B., D. Clasen, and S. Eicher (1986): "Perceptions of Peer Pressure, Peer Conformity Dispositions and Self-Reported Behavior Among Adolescents," *Developmental Psychology*, 22, 521–530.

Campbell, A., P. Converse, and W. Rodgers (1976): *The Quality of American Life*. New York: Russell Sage Foundation.

Conover, W. (1999): *Practical Nonparametric Statistics (3rd Edition)*. Wiley.

Corrado, L., R. Martin, and M. Weeks (2005): "Identifying and Interpreting Regional Convergence Clusters Across Europe," *The Economic Journal*, 115, C133–C160.

Deaton (1997): *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Johns Hopkins University Press for the World Bank, Baltimore.

Duncan, O. D. (1975): "Does Money Buy Satisfaction?," *Social Indicators Research*, 2, 267–274.

Durlauf, S. N. (2003): "Neighborhood Effects," in *Hanbook of Regional and Urban Economics*, ed. by J. F. Henderson, J. V. and Thisse. North Holland, Vol. 4, Economics.

Durlauf, S. N. (2004): "Groups, Social Influences, and Inequality: A Memberships Theory Perspective on Poverty Traps," in *Poverty Traps*, ed. by S. Bowles, S. N. Durlauf, and K. Hoff, Princeton. Princeton University Press.

Easterlin, R. A. (1974): "Does Economic Growth Improve the Human Lot?," in *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*, ed. by P. A. David, and M. W. Reder. New York: Academic Press, Inc.

——— (1994): "The Birth Dearth, Aging, and the Economy: Where Have We Been and Where Are We Going?," in *Human Capital and Economic Development*, ed. by S. Asefa, and W. C. Huang. Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

——— (2003): "Explaining Happiness," *Proceedings of the National Academy of Sciences*, 100, 11176–11183.

Elliott, S., and D. Umberson (2004): "Recent Demographic Trends in the US and Implications for Wellbeing.," in *The Blackwell Companion to the Sociology of Families*, ed. by J. T. JL Scott, and M. Richards. MA: Blackwell.

Fernandez, R. M., and J. C. Kulik (1981): "A Multilevel Model of Life Satisfaction: Effects of Individual Characteristics and Neighbourhood Composition," *American Sociological Review*, 46, 840–850.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (2003): "The Social Multiplier," *Journal of the European Economic Association*, 1(2-3), 345–353.

Goldstein, H. (1986): "Multilevel Mixed Linear Model Analysis Using Iterative Generalised Least Squares," *Biometrika*, 73, 46–56.

——— (1989): "Restricted Unbiased Iterative Generalized Least-Squares Estimation," *Biometrika*, 76(3), 622–623.

Goldstein, H. (1998): "Multilevel Models for Analysing Social Data," *Encyclopedia of Social Research Methods*.

Goldtsein, H., and J. Rasbash (1992): "Efficient Computational Procedures for the Estimation of Parameters in Multilevel Models Based on Iterative Least Squares," *Computational Statistics and Data Analysis*, 13, 63–71.

Gosh, M., and J. N. Rao (1994): "Small Area Estimation: An Appraisal," *Statistical Science*, 9(1), 55–93.

Graham, B. S. (2004): "Social Interactions and Excess Variance Contrasts: Identification, Estimation, and Inference with an Application on the Role of Peer Group Effects in Academic Achievement," mimeo, Department of Economics, Harvard University.

Hannan, M. T., and L. Burstein (1974): "Estimation from Grouped Observations," *American Sociological Review*, 39, 374–92.

Hausman, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(6), 1251–71.

Helliwell, J., and R. D. Putnam (2005): "The Social Context of Well-Being," in *The Science of Well-Being*, ed. by A. Huppert, B. Keverne, and N. Baylis. Oxford University Press, Oxford.

Huppert, F. A., N. Baylis, and B. Keverne (2005): *The Science of Well-Being*. Oxford University Press, Oxford.

Krugman, P. (1991a): *Geography and Trade*. MIT Press, Cambridge, MA.

——— (1991b): "Increasing Returns and Economic Geography," *Journal of Political Economy*, 99(3), 483–499.

——— (1992): "A Dynamic Spatial Model," *NBER Working Paper No. 4219.*

——— (1993): "First Nature, Second Nature, and Metropolitan Location," *Journal of Regional Science*, 33, 129–144.

Longford, N. T. (1993): *Random Coefficient Models.* Claredon Press, London.

Lyubomirsky, S., K. Sheldon, and D. Schkade (2005): "Pursuing Happiness: The Architecture of Sustainable Change," *Review of General Psychology*, 9, 111–131.

Manski, C. F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60(3), 531–542.

——— (2000): "Economic Analysis of Social Interactions," National Bureau of Economic Research Working Paper 7580.

Moulton, B. R. (1986): "Random Group Effects and the Precision of Regression Estimates," *Journal of Econometrics*, 32, 385–397.

——— (1990): "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit," *The Review of Economics and Statistics*, 72(2), 334–338.

Ormel, J., S. Lindenberg, N. Steverink, and L. Verbrugge (1999): "Subjective Wellbeing and Social Production Functions," *Social Indicators Research*, 46, 61–90.

Ryff, C. (1989): "Happiness is Everything, or is It? Explorations on the Meaning of Psychological Well-Being.," *Journal of Personality and Social Psychology*, 57, 1069–1081.

Ryff, C., and B. Singer (1998): "The Contours of Positive Human Health," *Psychological Inquiry*, 9(1), 1–28.

Seligman, M. (2002): *Authentic Happiness: Using the New Positive Psychology to Realize Your Potential for Lasting Fulfilment.* Free Press., New York.

Skrondal, A., and S. Rabe-Hesketh (2004): *Generalised Latent Variable Modelling: Multilevel, Longitudinal, and Structural Equation Models.* Chapman Hall, CRC Press.

Streufert, P. (1991): "The Effect of Underclass Isolation on Schooling Choice," Mimeo (University of Wisconsin at Madison).

A Derivation of Results for the Socio-Geographic Model

In this Appendix we provide a simple extension of Akerlof's model to the two-dimensional case. We show how individuals move closer to one another in two-dimensional space, which provides a better analogy for the socio-spatial regional clustering in Europe. This is in apposition to Akerlof's model which is based simply on the real line, i.e. one-dimensional space. We follow Akerlof in

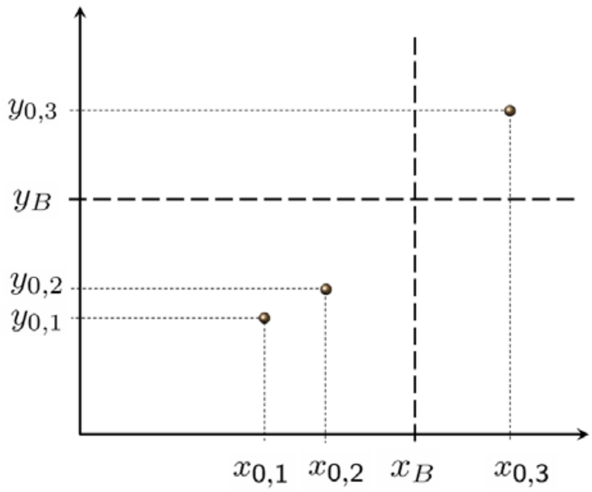


Figure 6: Simple 3-person Representation of a Two-Dimensional Extension to Akerlof's Model with Barriers to Movement.

assuming that there are three individuals randomly distributed in this two-dimensional space, at three distinct sets of coordinates in $x - y$ space (see Figure 6 below).

Our objective function summarises the 2 inputs into well-being, Y_i for individual i: (1) benefits from proximity and (2) the intrinsic benefits from one's choice of location along both the x-axis and y-axis, $x_{1,i}$ and $y_{1,i}$.

$$\begin{aligned} W_i &= \sum_{i \neq j} \frac{A}{(B + (x_{0,i} - x_{0,j})^2)\,(C + (x_{1,i} - x_{0,j})^2)} + \left[-a x_{1,i}^2 + b x_{1,i}\right] \\ &\quad + \sum_{i \neq j} \frac{E}{(F + (y_{0,i} - y_{0,j})^2)\,(G + (y_{1,i} - y_{0,j})^2)} + \left[-c y_{1,i}^2 + d y_{1,i}\right] + z + w_i \end{aligned} \tag{A.1}$$

To simplify the model we have imposed a limit on the range of movement by adding a natural barrier on the two axes at points x_B and y_B. In a model of 3 individuals this would mean that for individual 1 we would only consider the ranges $x_{1,1} < x_{0,2}$ and $x_{0,2} < x_{1,1} < x_B$ on the x-axis and $y_{1,1} < y_{0,2}$ and $y_{0,2} < y_{1,1} < y_B$ on the y-axis, when considering the parts of the line that individual 1 is able to move to. Similarly individual 3 is restricted in his/her movement within the range $x_{1,3} > x_B$ and $y_{1,3} > y_B$ since the barrier prevents any significant collocation with either of the other individuals.

To determine the position of individual 1 along the x-axis, we differentiate with respect to individual 1's location in the following period, $x_{1,1}$.

$$\begin{aligned} \frac{\partial W_1}{\partial x_{1,1}} &= \frac{A}{(B + (x_{0,1} - x_{0,2})^2)} \left(-\frac{2(x_{1,1} - x_{0,2})}{(C + (x_{1,1} - x_{0,2})^2)^2}\right) \\ &\quad + \frac{A}{(B + (x_{0,1} - x_B)^2)} \left(-\frac{2(x_{1,1} - x_B)}{(C + (x_{1,1} - x_B)^2)^2}\right) - 2a x_{1,1} + b = 0 \end{aligned} \tag{A.2}$$

Similarly, to determine the position of individual 1 along the y-axis, we differentiate with respect to individual 1's location in the following period, $y_{1,1}$.

$$\begin{aligned}\frac{\partial W_1}{\partial y_{1,1}} &= \frac{E}{(F+(y_{0,1}-y_{0,2})^2)}\left(-\frac{2(y_{1,1}-y_{0,2})}{(G+(y_{1,1}-y_{0,2})^2)^2}\right) \\ &+ \frac{E}{(F+(y_{0,1}-y_B)^2)}\left(-\frac{2(y_{1,1}-y_B)}{(G+(y_{1,1}-y_B)^2)^2}\right) - 2cy_{1,1} + d = 0\end{aligned} \tag{A.3}$$

We repeat these calculations for individual 2 to obtain:

$$\begin{aligned}\frac{\partial W_2}{\partial x_{1,2}} &= \frac{A}{(B+(x_{0,2}-x_{0,1})^2)}\left(-\frac{2(x_{1,2}-x_{0,1})}{(C+(x_{1,2}-x_{0,1})^2)^2}\right) \\ &+ \frac{A}{(B+(x_{0,2}-x_B)^2)}\left(-\frac{2(x_{1,2}-x_B)}{(C+(x_{1,2}-x_B)^2)^2}\right) - 2ax_{1,2} + b = 0\end{aligned} \tag{A.4}$$

$$\begin{aligned}\frac{\partial W_2}{\partial y_{1,1}} &= \frac{E}{(F+(y_{0,2}-y_{0,1})^2)}\left(-\frac{2(y_{1,2}-y_{0,1})}{(G+(y_{1,2}-y_{0,1})^2)^2}\right) \\ &+ \frac{E}{(F+(y_{0,2}-y_B)^2)}\left(-\frac{2(y_{1,2}-y_B)}{(G+(y_{1,2}-y_B)^2)^2}\right) - 2cy_{1,2} + d = 0\end{aligned} \tag{A.5}$$

Looking at individual 1 in the range $x_{1,1} < x_{0,2}$, we find that $\frac{\partial W_1}{\partial x_{1,1}} > 0$ as both the first and the second term will be positive. In the range $x_{0,2} < x_{1,1} < x_B$ the derivative $\frac{\partial W_1}{\partial x_{1,1}} < 0$; this occurs if the first term dominates the second. i.e. if $x_{1,1}$ is much larger than $x_{0,2}$. These results are repeated precisely for the y-coordinates: in the range $y_{1,1} < y_{0,2}$, we find that $\frac{\partial W_1}{\partial y_{1,1}} > 0$ and in the range $y_{0,2} < y_{1,1} < y_B$ the derivative $\frac{\partial W_1}{\partial y_{1,1}} < 0$; this occurs if the first term dominates the second. i.e. if $y_{1,1}$ is much larger than $y_{0,2}$. Therefore as utility is increasing when both $x_{1,1} < x_{0,2}$ and $y_{1,1} < y_{0,2}$ and decreasing when both $x_{0,2} < x_{1,1} < x_B$ and $y_{0,2} < y_{1,1} < y_B$ then individual 1 will switch his position with individual 2.

Moving on to individual 2, in the range $x_{1,2} < x_{0,1}$ we find that $\frac{\partial W_2}{\partial x_{1,2}} > 0$ as both the first and the second terms will be positive. In the range $x_{0,1} < x_{1,2} < x_B$ the derivative $\frac{\partial W_2}{\partial x_{1,2}} < 0$; this happens once again if the first term dominates the second, i.e. if $x_{1,2}$ is much larger than $x_{0,1}$. And similar results hold for the y-coordinates. Therefore as utility is increasing when both $x_{1,2} < x_{0,1}$ and $y_{1,2} < y_{0,1}$ and decreasing when both $x_{0,1} < x_{1,2} < x_B$ and $y_{0,1} < y_{1,2} < y_B$ individual 2 will switch his position with individual 1.

In more general terms (considering this problem as a dynamic one which is repeated every two periods) we could differentiate with respect to individual i's location in the following period.

$$\frac{\partial W_i}{\partial x_{1,i}} = \sum_{i \neq j} \frac{A}{(B+(x_{0,i}-x_{0,j})^2)}\left(-\frac{2(x_{1,i}-x_{0,j})}{(C+(x_{1,i}-x_{0,j})^2)^2}\right) - 2ax_{1,i} + b = 0 \tag{A.6}$$

$$\frac{\partial W_i}{\partial y_{1,i}} = \sum_{i \neq j} \frac{E}{(F+(y_{0,i}-x_{0,j})^2)}\left(-\frac{2(y_{1,i}-y_{0,j})}{(G+(y_{1,i}-y_{0,j})^2)^2}\right) - 2ay_{1,i} + b = 0 \tag{A.7}$$

For individual i in the range $x_{1,i} < x_{0,j}$ and $y_{1,i} < y_{0,j}$ we will find that $\frac{\partial W_i}{\partial x_{1,i}}, \frac{\partial W_i}{\partial y_{1,i}} > 0$ while in the range $x_{0,j} < x_{1,i} < x_B$ and $y_{0,j} < y_{1,i} < y_B$ the derivatives $\frac{\partial W_i}{\partial x_{1,i}}$ and $\frac{\partial W_i}{\partial y_{1,i}}$ will be negative if the social distance between individual i in the second period and individual j's position in the first period is increasing.

B The Reflection Problem and Endogeneity

In a two level model, we would be interested in taking into account the three social interaction effects discussed by Manski (1993) using the following type of multilevel regression with random intercepts

$$Y_{ij} = \beta_0 + \beta_1 \bar{Y}_j + \beta_2 X_{ij} + \beta_3 \bar{X}_j + e_{ij} \tag{B1}$$

where

$$\beta_0 = \beta_{00} + \beta_{01} Z_j + u_{0j}$$

The random intercept shows that the heterogeneity across the level two groups is explained by a set of observable group characteristics, Z_j and a random disturbance term, u_{0j} which denotes the unobserved group effects. Therefore having included all forms of social interactions, we end up with

$$Y_{ij} = \beta_{00} + \beta_{01} Z_j + u_{0j} + \beta_1 \bar{Y}_j + \beta_2 X_{ij} + \beta_3 \bar{X}_j + e_{ij} \tag{B2}$$

and we can interpret the coefficients as follows: β_2 and β_3 measure the effects of exogenous contextual effect and β_1 as measuring the effect of endogenous social interactions. The random intercept shows that the heterogeneity across the level-two groups is explained by a set of observable group characteristics, Z_j and a random disturbance term, u_{0j}.

One main problem with relationship (B2) is proper identification of the parameters. The main question is whether the two social effects measured by the parameters β_1 and β_3 can be distinguished from one another and from the non-social effects. This is known as Manski's reflection problem and can arise when a researcher observing the distribution of behaviour in the population tries to infer whether the average behaviour in a group influences the behaviour of the individuals that comprise that group. To establish whether we can properly identify the parameters of (B2) we pursue the following identification strategy.

Taking group means of both sides of (B2) and solving for $\overline{Y}_j$, we can see that $\bar{Y}_j$ is correlated with the error term since

$$\begin{aligned} \bar{Y}_j &= \beta_0 + \beta_1 \bar{Y}_j + (\beta_2 + \beta_3) \bar{X}_j + \upsilon_j \\ \bar{Y}_j &= \frac{\beta_0}{1-\beta_1} + \frac{(\beta_2+\beta_3)}{1-\beta_1} \bar{X}_j + \frac{1}{1-\beta_1} \upsilon_j \end{aligned} \tag{B3}$$

Substituting back into the original relationship

$$
\begin{aligned}
Y_{ij} &= \beta_0 + \beta_1 \left(\frac{\beta_0}{1-\beta_1} + \frac{(\beta_2+\beta_3)}{1-\beta_1} \bar{X}_j + \frac{1}{1-\beta_1} \bar{e}_j \right) + \beta_2 X_{ij} + \beta_3 \bar{X}_j + e_{ij} \qquad \text{(B4)} \\
Y_{ij} &= \left(\frac{\beta_1 \beta_0}{1-\beta_1} + \beta_0 \right) + \beta_2 X_{ij} + \left(\frac{\beta_1 (\beta_2+\beta_3)}{1-\beta_1} + \beta_3 \right) \bar{X}_j + \frac{\beta_1}{1-\beta_1} \upsilon_j + e_{ij} \\
Y_{ij} &= \frac{\beta_0}{1-\beta_1} + \beta_2 X_{ij} + \frac{\beta_1 \beta_2 + \beta_3}{1-\beta_1} \bar{X}_j + \frac{\beta_1}{1-\beta_1} \upsilon_j + e_{ij}
\end{aligned}
$$

Assuming $\beta_1 \neq 1$, this equation shows that we cannot identify and therefore recover all the coefficients of the original model, except for β_2.

As highlighted in Manski (1993) the linear model where $\beta_1 \neq 1$, the composite parameters $\frac{\beta_0}{1-\beta_1}$, $\frac{\beta_1\beta_2+\beta_3}{1-\beta_1}$, $\frac{\beta_1}{1-\beta_1}$, and β_2 are identified if the regressors $[1, \overline{X}_j, X_{ij}, Z_j]$ are linearly independent in the population. However, we can only recover β_2 from the reduced form regression. The reduced form shows that the effect of the social multiplier β_3 and of the peer group effect β_1 is to amplify the natural variations arising from individual and group-level heterogeneity.

Therefore the only way to safely include any form of endogenous social interaction in a multilevel model, it is necessary to follow the sequence of steps laid out in the main body of the paper, which implicitly includes $\bar{Y}_j$ in the deviation form

$$
Y_{ij} = \alpha_0 + \alpha_1 X_{ij} + (\beta_1 - \gamma_1) \overline{X}_j + \bar{e}_j + \epsilon_{ij} \qquad \text{(B5)}
$$

where the within-group regression coefficient is still given by α_1, but the between-group regression coefficient is now $(\beta_1 - \gamma_1)$.

C Understanding the Equivalence of Using $(X_{ij} - \bar{X}_j)$ and X_{ij} in a Multilevel Regression

Raudenbush (1989) gave statistical reasons for preferring a centred regression over the uncentred version, i.e. a preference for

$$
Y_{ij} = \tilde{\beta}_{00} + \tilde{\beta}_{11} \left(X_{ij} - \bar{X}_j \right) + \tilde{\beta}_{12} \bar{X}_j + u_{0j} + e_{ij} \qquad \text{(C1)}
$$

over

$$
Y_{ij} = \beta_0 + \beta_{11} X_{ij} + \beta_{12} \bar{X}_j + e_{ij} \qquad \text{(C2)}
$$

This was because the model was meant to suffer from high collinearity (Aitkin and Longford (1986)). We can show that the former equation is simply a reparameterisation of the latter.

We start with the latter equation where the constant is in fact a random intercept is defined as: $\beta_0 = \beta_{00} + u_{0j}$. This becomes

$$
Y_{ij} = \beta_{00} + \beta_{11} X_{ij} + \beta_{12} \bar{X}_j + u_{0j} + e_{ij} \qquad \text{(C3)}
$$

If we look at this relationship within a particular group, j, and rearrange it as follows

$$Y_{ij} = \left(\beta_{00} + \beta_{12}\bar{X}_j + u_{0j}\right) + \beta_{11}X_{ij} + e_{ij} \tag{C4}$$

where $\left(\beta_{00} + \beta_{12}\bar{X}_j + u_{0j}\right)$ is the random intercept for this group and therefore $\beta_{11}$is the within-group regression coefficient, i.e. the coefficient within this particular group.

If we take averages of (C3) we get the between-group regression (i.e. a regression between the group means)

$$\begin{aligned} \bar{Y}_j &= \beta_{00} + \beta_{11}\bar{X}_j + \beta_{12}\bar{X}_j + \upsilon_j \\ \bar{Y}_j &= \beta_{00} + (\beta_{11} + \beta_{12})\bar{X}_j + \upsilon_j \end{aligned} \tag{C5}$$

where the between-group coefficient is $\left(\beta_{11} + \beta_{12}\right)$. A Wald test of the equality of the between-group and within-group regression coefficients (i.e. a test of the null that $\beta_{12} > 0$) is identical to the Hausmann specification test for the random intercept model (see Hausman (1978)).

If the within-group and between-group coefficients are different we can replace X_{ij} by the *within-group deviation scores*: $\left(X_{ij} - \bar{X}_j\right)$:

$$Y_{ij} = \tilde{\beta}_{00} + \tilde{\beta}_{11}\left(X_{ij} - \bar{X}_j\right) + \tilde{\beta}_{12}\bar{X}_j + u_{0j} + e_{ij} \tag{C6}$$

which is statistically equivalent to (C3) but has a more convenient parameterisation since the sets of covariates, $\left(X_{ij} - \bar{X}_j\right)$ and $\bar{X}_j$, are orthogonal and therefore uncorrelated. We get the following coefficients

$$\begin{aligned} \text{Within-Group Coefficients} \quad &: \quad \tilde{\beta}_{11} = \beta_{11} \\ \text{Between-Group Coefficients} \quad &: \quad \tilde{\beta}_{12} = \beta_{11} + \beta_{12} \end{aligned}$$

D Generalising to a Three-Level Model

Looking at the three-level model we wish to end up with a testable relation of the form

$$Y_{ijk} = \beta_0 + \beta_{11}X_{ijk} + \beta_{12}\bar{X}_{jk} + \beta_{13}\bar{X}_k + e_{ijk} \tag{D1}$$

Once again within a particular group at level two, we can rearrange to find the within-group regression coefficient

$$Y_{ijk} = \left(\beta_0 + \beta_{12}\bar{X}_{jk} + \beta_{13}\bar{X}_k\right) + \beta_{11}X_{ijk} + e_{ijk} \tag{D2}$$

where $\left(\beta_0 + \beta_2\bar{X}_{jk} + \beta_3\bar{X}_k\right)$ is the intercept term and β_{11} is the required coefficient. Taking means at the second level we can obtain the within-group regression at level two

$$\begin{aligned} \bar{Y}_{jk} &= \beta_0 + \beta_{11}\bar{X}_{jk} + \beta_{12}\bar{X}_{jk} + \beta_{13}\bar{X}_k + \upsilon_{jk} \\ \bar{Y}_{jk} &= \left(\beta_0 + \beta_{13}\bar{X}_k\right) + (\beta_{11} + \beta_{12})\bar{X}_{jk} + \upsilon_{jk} \end{aligned} \tag{D3}$$

the intercept for which is $\left(\beta_0 + \beta_{13}\bar{X}_k\right)$ and the within-group regression coefficient at level two is $(\beta_{11} + \beta_{12})$. Finally we can also obtain the between-group regression coefficient at level three by averaging the original relation over the third level groups, denoted by k

$$\bar{Y}_{jk} = \beta_0 + (\beta_{11} + \beta_{12} + \beta_{13})\,\bar{X}_k + \upsilon_{jk} \tag{D4}$$

which shows that our constant in the original regression, β_0, is in fact the constant for the level three between group regression and that within-group regression coefficient at level three is $(\beta_{11} + \beta_{12} + \beta_{13})$. Assuming once again that the within-group regression coefficients at each level are different (using the Hausmann specification test) we can replace X_{ijk} and $\bar{X}_{jk}$ by their within-group deviation scores, $\left(X_{ijk} - \bar{X}_{jk}\right)$ and $\left(\bar{X}_{jk} - \bar{X}_k\right)$, respectively

$$Y_{ijk} = \beta_0 + \tilde{\beta}_{11}\left(X_{ijk} - \bar{X}_{jk}\right) + \tilde{\beta}_{12}\left(\bar{X}_{jk} - \bar{X}_k\right) + \tilde{\beta}_{13}\bar{X}_k + e_{ijk} \tag{D5}$$

where now

$$\begin{array}{rl} \text{Within-Group (Level 1) Coefficient:} & \tilde{\beta}_{11} = \beta_{11} \\ \text{Within-Group (Level 2) Coefficient:} & \tilde{\beta}_{12} = (\beta_{11} + \beta_{12}) \\ \text{Between-Group (Level 3) Coefficient:} & \tilde{\beta}_{13} = (\beta_{11} + \beta_{12} + \beta_{13}) \end{array}$$

With reference to our three-level model, a complete specification including random intercepts would be as follows

$$Y_{ijk} = \beta_0 + \beta_{11}X_{ijk} + \beta_{12}\bar{X}_{jk} + \beta_{13}\bar{X}_k + e_{ijk} \tag{D6}$$

where

$$\beta_0 = \beta_{000} + \beta_{010}Z_{jk} + e_{0jk} + \beta_{001}Z_k + e_{00k} \tag{D7}$$

and once again defining $\mathrm{Var}(e_{ijk}) = \Sigma_1$, $\mathrm{Var}(e_{0jk}) = \Sigma_2$, and $\mathrm{Var}(e_{00k}) = \Sigma_3$ we assume

$$\begin{array}{ll} e_{ijk} \mid X_{ijk} \sim N(0, \Sigma_1) & \mathrm{Cov}(e_{ijk}, e_{i'jk}) = 0,\ \forall i \neq i' \\ e_{0jk} \mid X_{ijk} \sim N(0, \Sigma_2) & \mathrm{Cov}(e_{0jk}, e_{ijk}) = 0 \\ e_{00k} \mid X_{ijk} \sim N(0, \Sigma_3) & \mathrm{Cov}(e_{00k}, u_{ijk}) = 0 \\ & \mathrm{Cov}(e_{0jk}, e_{00k}) = 0 \end{array}$$

In addition to the random intercepts, we would also be including the intrinsic choice variables as explanatory factors, W_{ijk} such that the full model becomes

$$Y_{ijk} = \beta_{000} + \beta_{010}Z_{jk} + \beta_{001}Z_k + \beta_{11}X_{ijk} + \beta_{12}\bar{X}_{jk} + \beta_{13}\bar{X}_k + \beta_2 W_{ijk} + e_{0jk} + e_{00k} + e_{ijk} \tag{D8}$$

E Linear Multilevel Modelling Estimation

This Appendix covers linear multilevel modelling estimation. We rewrite (14), as:

$$\mathbf{Y} = \boldsymbol{\beta}_{00} + \mathbf{Z}\boldsymbol{\beta}_{01} + \mathbf{X}\boldsymbol{\beta}_1 + \overline{\mathbf{X}}\boldsymbol{\beta}_2 + \mathbf{W}\boldsymbol{\beta}_3 + \mathbf{u}\boldsymbol{\theta} \tag{E1}$$

$\mathbf{Y} = \{Y_{ij}\}$, $\mathbf{X} = \{X_{ij}\}$, $\overline{\mathbf{X}} = \{\bar{X}_j\}$, $\mathbf{Z} = \{Z_j\}$ and $\mathbf{u} = \{u_{ij}\} + \{u_{0j}\}$.

The vectors $\boldsymbol{\beta}_{00}$, $\boldsymbol{\beta}_{01}$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_3$ denote the fixed coefficients while the vector $\boldsymbol{\theta}$ denotes the random parameters of the model. We first rewrite (E1) in compact form as:

$$\mathbf{Y} = \mathbf{J}\boldsymbol{\lambda} + \mathbf{u}\boldsymbol{\theta} \tag{E2}$$

The hierarchical two-stage method for estimating the fixed and random parameters (the variance and covariances of the random coefficients) originally proposed by Goldstein (1986) and implemented in the software MLWiN, is based upon an Iterative Least Squares (IGLS) method that results in consistent and asymptotically efficient estimates of $\boldsymbol{\lambda}$.

First we obtain starting values for $\boldsymbol{\lambda}$, $\boldsymbol{\lambda}_0$, by performing OLS in a standard single level system assuming the variance at higher level of the model to be zero. Conditioned upon $\boldsymbol{\lambda}_0$, we form the vector of residuals which we use to construct an initial estimate, $\mathbf{V}$, the covariance matrix for the response variable $\mathbf{Y}$. Then we iterate the following procedure first estimating the fixed parameters in a GLS regression as:

$$\overset{\wedge}{\boldsymbol{\lambda}} = (\mathbf{J}^T\mathbf{V}^{-1}\mathbf{J})^{-1}\left(\mathbf{J}^T\mathbf{V}^{-1}\mathbf{Y}\right) \tag{E3}$$

and again calculating residuals $\widehat{\mathbf{H}} = \mathbf{Y} - \mathbf{J}\overset{\wedge}{\boldsymbol{\lambda}}$. We form the matrix product of these residuals and stack them into a vector, i.e. $\mathbf{H}^* = vec(\widehat{\mathbf{H}}\,\widehat{\mathbf{H}}^T)$ which is then used in the next level of estimation to obtain consistent estimates of $\sigma^2_{u_{ij}}$ and $\sigma^2_{u_{0j}}$. Hence, we can estimate the random parameters $\boldsymbol{\theta}$ as:

$$\widehat{\boldsymbol{\theta}} = (\mathbf{u}^{*T}\mathbf{V}^{*-1}\mathbf{u}^*)^{-1}\left(\mathbf{u}^{*T}\mathbf{V}^{*-1}\mathbf{H}^*\right) \tag{E4}$$

where $\mathbf{V}^*$ is the Kronecker product of $\mathbf{V}$, namely $\mathbf{V}^* = \mathbf{V} \otimes \mathbf{V}$ and the covariance matrix is given by $\mathbf{V} = E(\widehat{\mathbf{H}}\,\widehat{\mathbf{H}}^T)$. The matrix $\mathbf{u}^*$ is the design matrix of the random parameters. The estimates of $\sigma^2_{u_{ij}}$ and $\sigma^2_{u_{0j}}$ are used to construct the covariance matrix of the response variable $\mathbf{Y}$ at each iteration using a GLS estimation of the fixed parameters. Once the fixed coefficients are obtained, updated residuals are formed and the random parameters estimated once again. This procedure is repeated until some convergence criteria are met.[25] As Goldstein (1989) has stressed, the IGLS used in the context of random multilevel modelling is equivalent to a maximum likelihood method under multivariate normality which in turn may lead to biased estimates. To produce unbiased estimate we use a Restricted Iterative Generalized Least Squares (RIGLS) method which, after the convergence is achieved, turns to be equivalent to a Restricted Maximum Likelihood Estimate (REML). One advantage of the latter method is that, differently from IGLS, estimates of the variance components take into account the loss of the degrees of freedom resulting from the estimation of the regression parameters. Hence, while the IGLS estimates for the variance components have a downward bias, the RIGLS estimates don't.

[25] Assuming multivariate normality the estimated covariance matrix for the fixed parameter is $\text{cov}(\overset{\wedge}{\boldsymbol{\lambda}}) = (\mathbf{J}^T\mathbf{V}^{-1}\mathbf{J})^{-1}$and for the random parameters (Goldtsein and Rasbash (1992)) is $\text{cov}(\overset{\wedge}{\boldsymbol{\theta}}) = 2(\mathbf{u}^T\mathbf{V}^{*-1}\mathbf{u})^{-1}$.

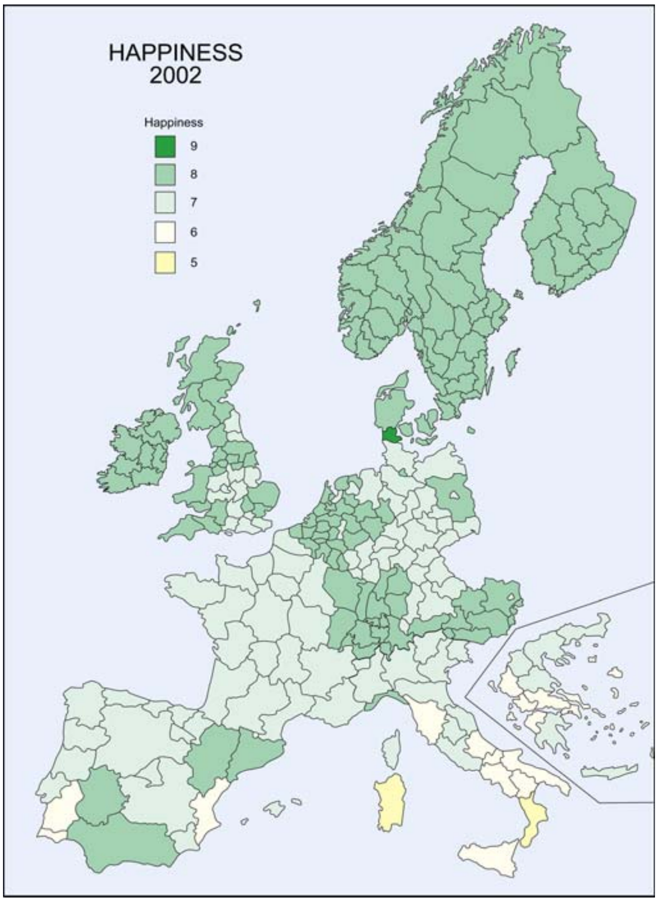


Figure 7: Geography of Happiness 2002

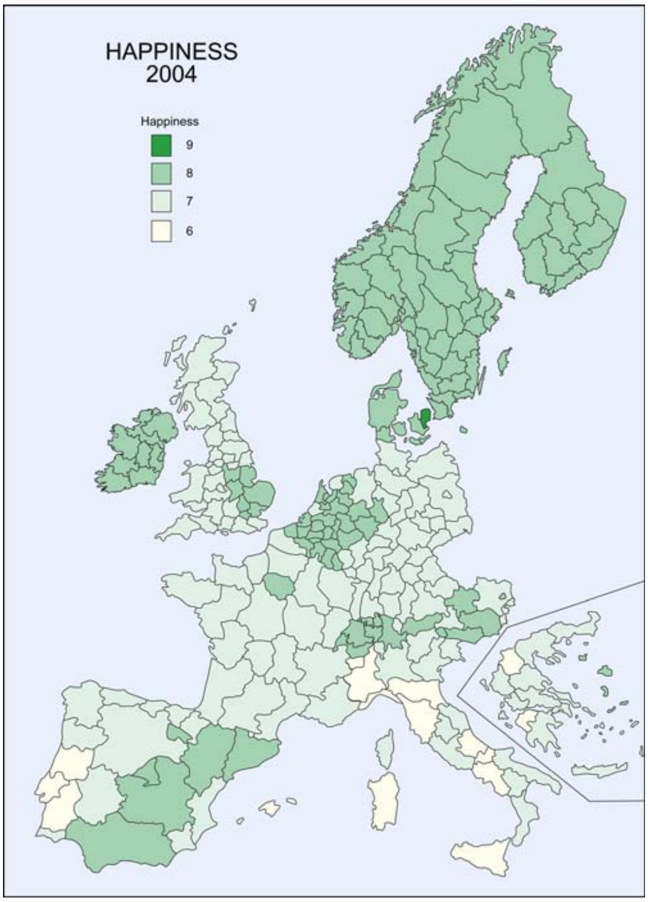


Figure 8: Geography of Life Satisfaction 2002

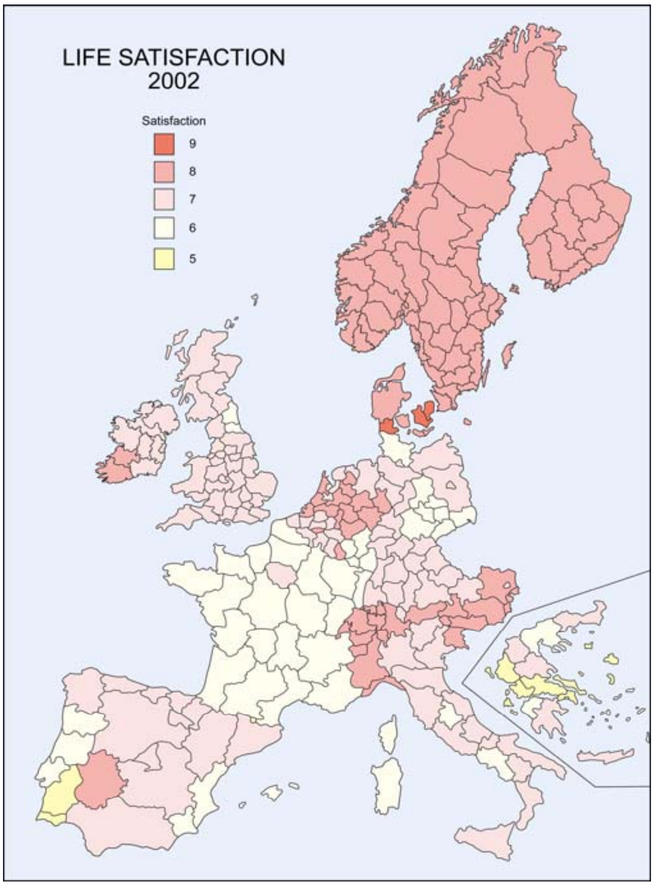


Figure 9: Geography of Happiness 2004

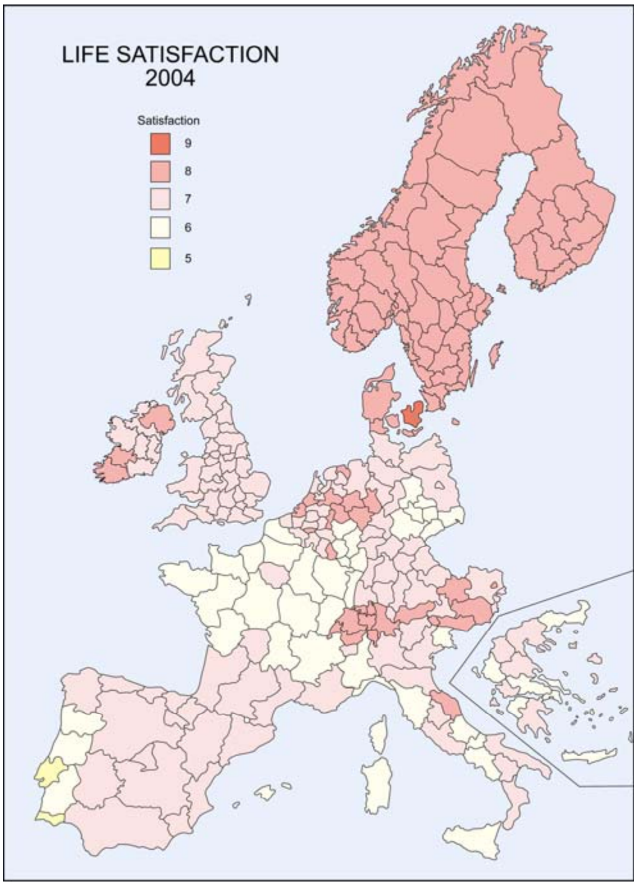


Figure 10: Geography of Life Satisfaction 2004

2002 | 2004

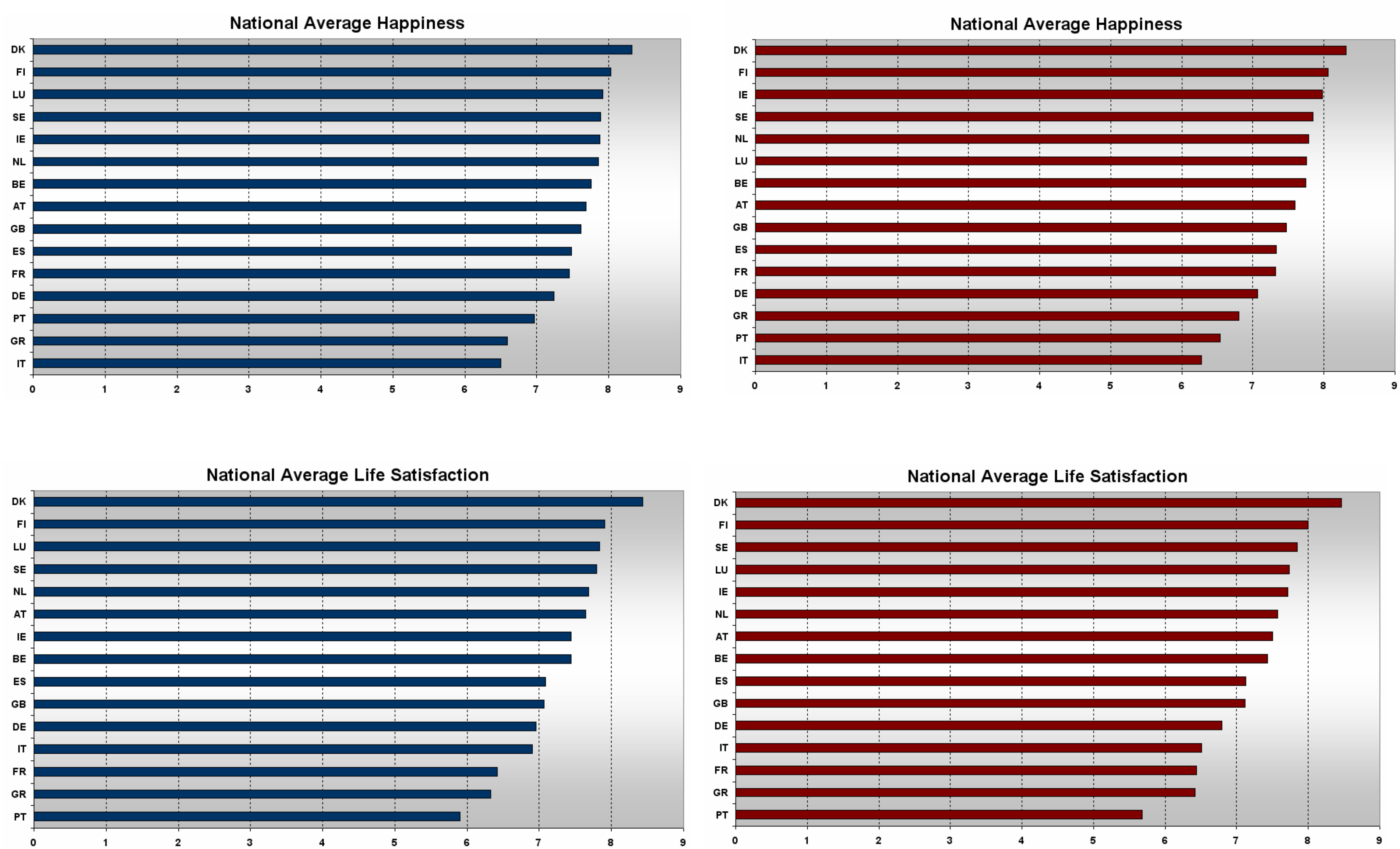


Figure 11: National Averages for Happiness and Life Satisfaction

2002 2004

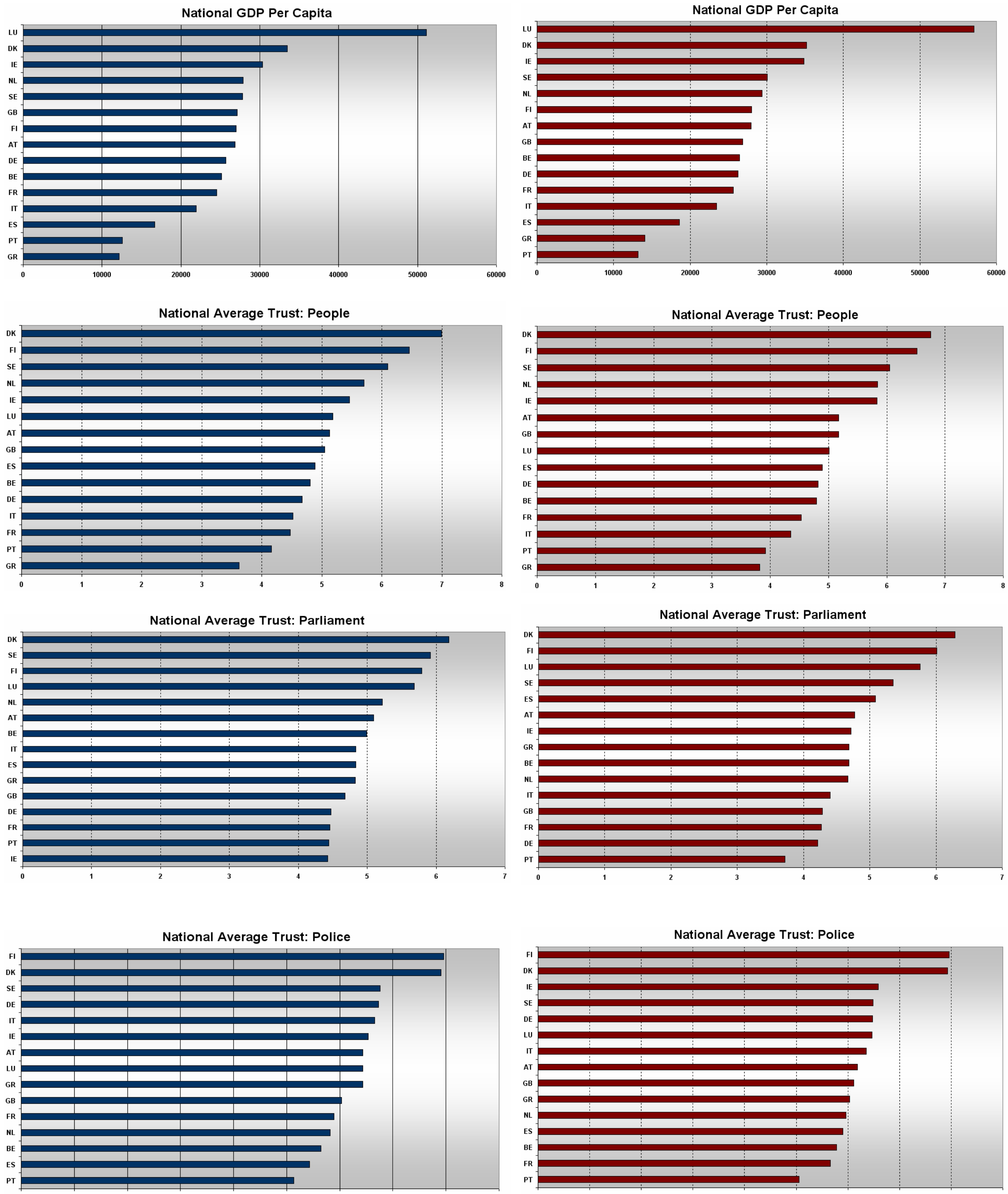


Figure 12: National Averages for Exogenous Variables

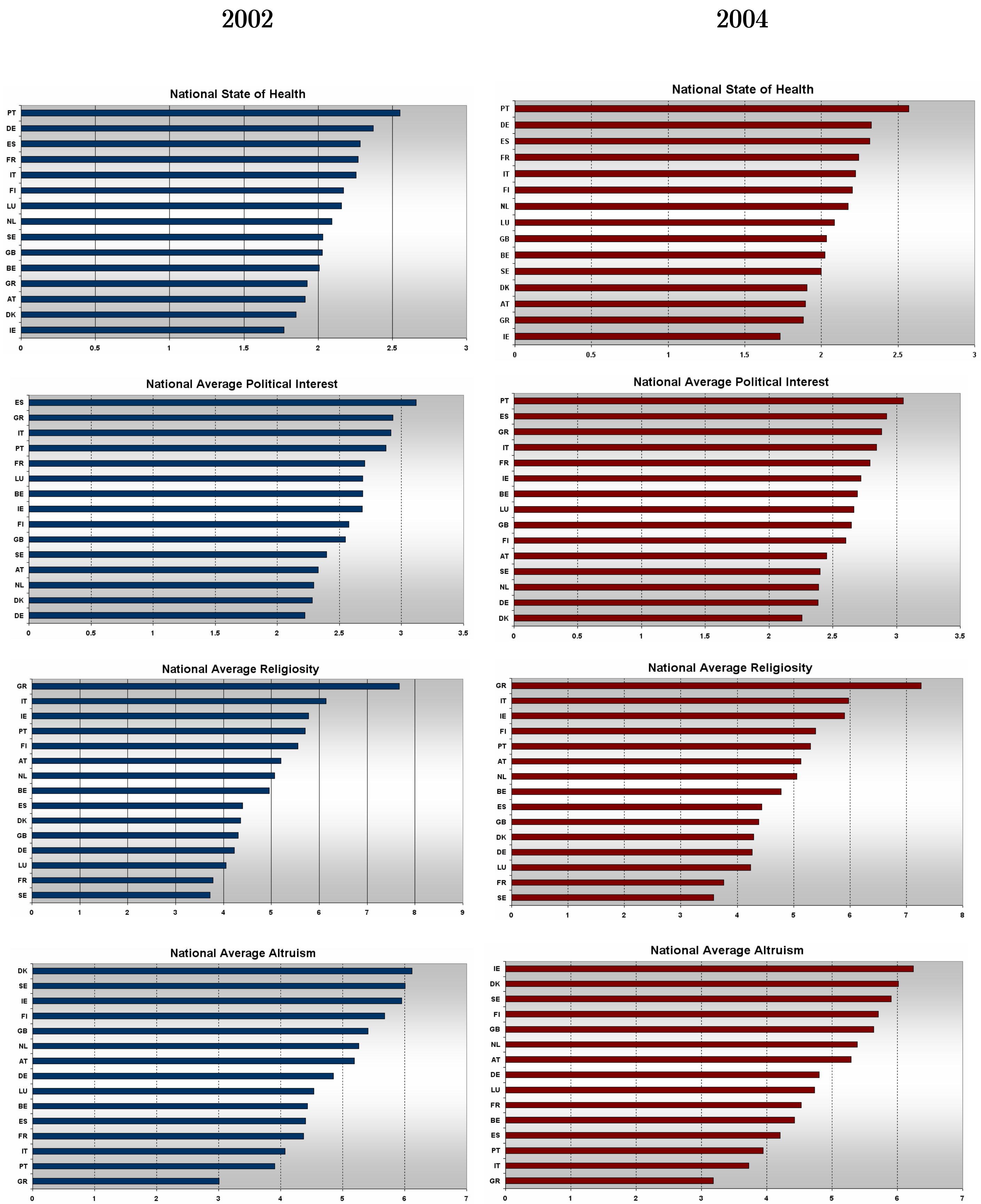


Figure 13: National Averages for Exogenous Variables

2002

2004

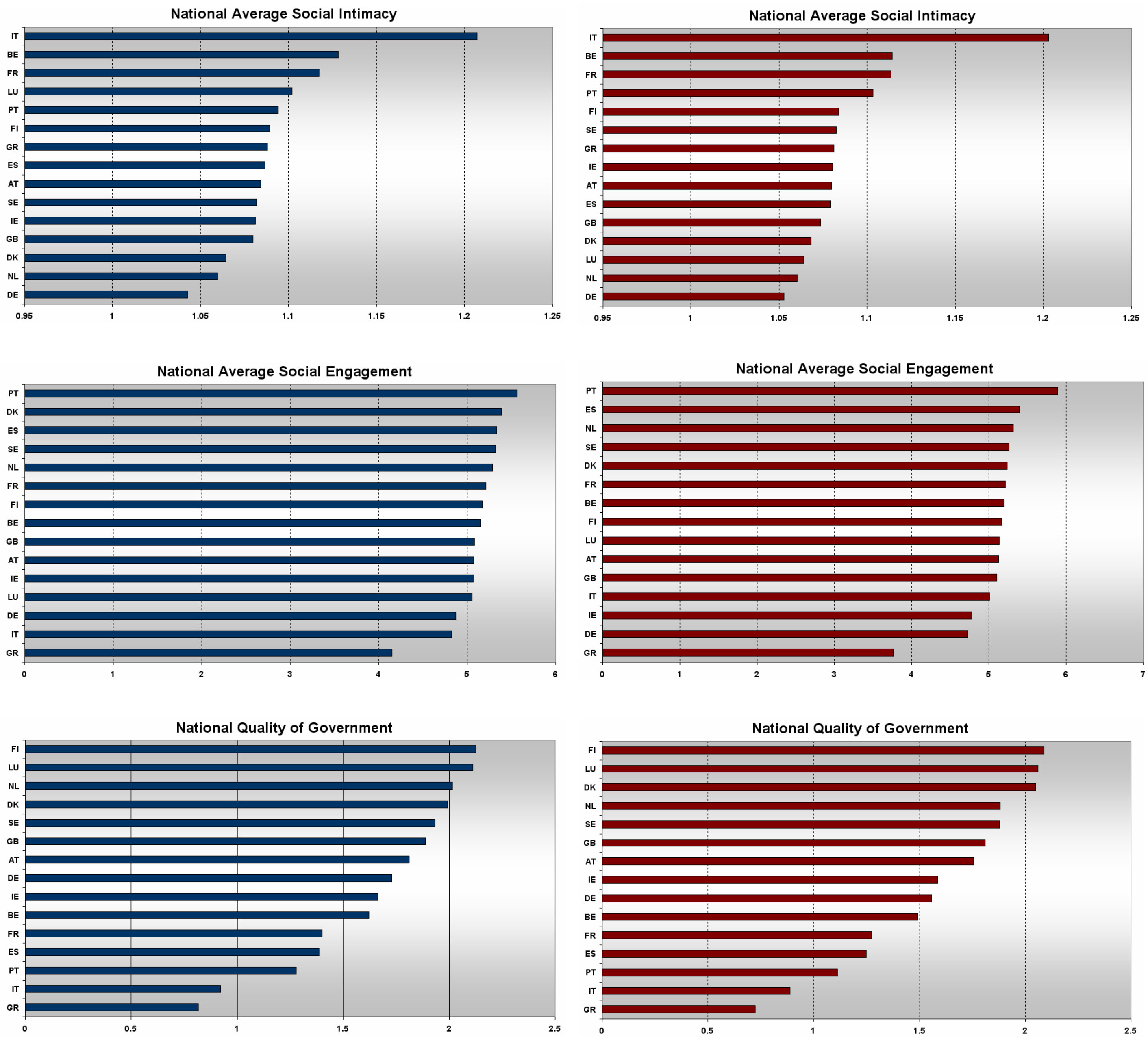


Figure 14: National Averages for Exogenous Variables

Table 1: List of Variables.

Variables	Response Categories	Description
LEVEL 3 - COUNTRY		
Exogenous Contextual Effects (Z_k)		
Quality of Government	Scaled -2.5 to 2.5	Average of: Quality Government Effectiveness, Regulatory Quality, Rule of Law, & Control of Corruption.
National GDP	Continuous	Log of GDP per capita for NUTS0 units.
Exogenous Social Interactions ($\overline{X}_k$)		
Trust: Country's Parliament	National Average	How much do you personally Trust in country's parliament?
Trust: Police	National Average	How much you personally trust in the Police?
Trust: People	National Average	Would you say that most people can be trusted?
Political Interest	National Average	How interested would you say you are in politics?
Religiosity	National Average	How religious are you?
Social Engagement	National Average	How often socially meet with friends, relatives or colleagues?
Social Intimacy	National Average	Do you have anyone with whom you can discuss intimate and personal matters?
Subjective State of Health	National Average	How is your health in general?
LEVEL 2 - REGIONS		
Exogenous Contextual Effects (Z_{jk})		
Regional GDP	Continuous	Log of GDP per capita for NUTS1/NUTS2 regional units
Location of Region	Unordered 1 - 5	Manual classification according to geographic location: 1=North; 2=Atlantic; 3=Mediterranean; 4=Eastern EU-border; 5=Centre
Settlement Structure	Scaled Continuous 0-1	I.1 Agglomerated regions with a centre > 300,000 and a population density > 300 inhabitants/km^2, I.2 Agglomerated regions with a centre > 300,000 or a population density 150,000 - <300/km^2 II.2 Urbanised regions with a centre 150,000 - <300,000 and a population density 150 - <300/km^2 [or a smaller population density (100 - <150/km^2) with a bigger centre (> 300,000)] II.2 Urbanised regions with a centre 150,000 - <300,000 or a population density 100 - <150/km^2 III.1 Rural regions with a population density < 100 inhabitants/km^2 and a centre > 125,000 III.2 Rural regions with a population density < 100 inhabitants/km^2 or a population density < 100/km^2 with a centre < 125,000
Exogenous Social Interactions ($\overline{X}_{jk}$)		
Trust: Country's Parliament	Regional Average	How much do you personally Trust in country's parliament?
Trust: Police	Regional Average	How much you personally trust in the Police?
Trust: People	Regional Average	Would you say that most people can be trusted?
Political Interest	Regional Average	How interested would you say you are in politics?
Religiosity	Regional Average	How religious are you?
Social Engagement	Regional Average	How often socially meet with friends, relatives or colleagues?
Social Intimacy	Regional Average	Do you have anyone with whom you can discuss intimate and personal matters?
Subjective State of Health	Regional Average	How is your health in general?

[1]Kaufmann, Kraay and Mastruzzi (2003), Governance Indicators.
[2]Established by Schmidt-Seiwert, Volker (1997): Landkarten zum Vergleich westlicher Regionen. In: Hradil S., Immerfall S. (eds.), Die westeuropäischen Regionen im Vergleich. Opladen, pp. 603-628. Based on the two criteria population density and size of centres, he distinguishes agglomerations, urbanised areas and rural areas and defines six types of regions

Table 1: List of Variables (contd.).

Variables	Response Categories	Description
LEVEL 1 - INDIVIDUALS		
Exogenous Social Interactions (X_{ijk})		
Trust: Country's Parliament	Scaled	How much do you personally Trust in country's parliament?
Trust: Police	Scaled	How much you personally trust in the Police?
Trust: People	Scaled	Would you say that most people can be trusted?
Social Intimacy	Scaled	Do you have anyone with whom you can discuss intimate and personal matters?
Social Engagement	Scaled	How often socially meet with friends, relatives or colleagues?
Religiosity	Scaled	How religious are you?
Political Interest	Scaled	How interested would you say you are in politics? 1=Very Interested; 2=Quite Interested; 3=Hardly Interested; 4=Not At All Interested.
Subjective State of Health	Scaled	How is your health in general? 1=Very Good; 2=Good; 3=Fair; 4=Bad; 5=Very Bad.
Intrinsic Socio-demographic Indicators (W_{ijk})		
Citizen of Country	Binary	Are you a citizen of [country]? 1=Yes; 0=No.
Victim of Crime	Binary	Have you or a member of your household been the victim of a burglary or assault in the last 5 years? 1=Yes; 0=No
Ethnicity	Binary	Do you belong to a minority ethnic group in [country]? 1=Yes; 0=No.
Age	Binary	Dummy variables for 'Aged less than 25 years'; 'Aged between 25-34 years'; 'Aged between 35-44 years'; 'Aged between 45-54 years'; 'Aged between 55-64 years'.
Marital Status	Binary	Dummy variables for 'Married'; 'Divorced'; 'Living with Partner'; 'Widowed'.
Altruism	Continuous	How Helpful Are People?
Household Income (Euros per month)	Ordinal	If you add up the income from all sources, which letter describes your household's total net income? J=Less than 150; R=150 to under 300; C=300 to under 500; M=500 to under 1000 F=1000 to under 1500; S=1500 to under 2000; K=2000 to under 2500; P=2500 to under 3000; D=3000 to under 5000;H=5000 to under 7500; U=7500 to under 10000; N=10000 or more.
Unemployed	Binary	1=Unemployed; 0=Employed
Gender	Binary	1=Male; 2=Female
Educational level	Binary	Dummy variables for 'High-School'; 'Post-High'; 'University'.

Table 2: Pairwise Correlations for the Variables (Dependent and Explanatory) in the Multilevel Model

		Y1	Y2	X1	X2	X3	X4	X5	X6	X7	X9	X10	X11	X12	X13	X14	X15	X16	X17
Y1	Happiness	1.0000																	
Y2	Life Satisfaction	0.7070	1.0000																
X1	Trust in People	0.2442	0.2869	1.0000															
X2	Altruism	0.2270	0.2546	0.4875	1.0000														
X3	Political Interest	-0.0992	-0.1216	-0.1692	-0.0998	1.0000													
X4	Trust in Parliament	0.2040	0.2623	0.3322	0.2442	-0.2160	1.0000												
X5	Trust in Legal System	0.2109	0.2650	0.3354	0.2575	-0.1501	0.5888	1.0000											
X6	Trust in Police	0.2413	0.2734	0.2978	0.2562	-0.0942	0.4547	0.6257	1.0000										
X7	Trust in Politicians	0.2083	0.2636	0.3374	0.2803	-0.2153	0.7086	0.5449	0.4539	1.0000									
X9	Social Engagement	0.1684	0.1406	0.1079	0.0832	-0.0432	0.0311	0.0229	0.0015	0.0186	1.0000								
X10	Social Intimacy	-0.1751	-0.1382	-0.0667	-0.0643	0.0581	-0.0472	-0.0610	-0.0509	-0.0463	-0.1422	1.0000							
X11	Crime Victim	0.0030	0.0135	0.0054	0.0358	0.0339	-0.0055	0.0227	0.0324	0.0160	-0.0599	0.0096	1.0000						
X12	Health	-0.2941	-0.2905	-0.1436	-0.0858	0.0560	-0.1169	-0.1261	-0.0887	-0.1131	-0.1146	0.1177	0.0095	1.0000					
X13	Religious Degree	0.0517	0.0567	0.0051	0.0387	0.0161	0.1027	0.0820	0.1208	0.1115	-0.1042	0.0112	0.0454	0.0314	1.0000				
X14	Citizenship	-0.0139	-0.0273	-0.0252	-0.0297	0.0429	0.0384	0.0257	-0.0102	0.0401	-0.0128	0.0091	-0.0148	-0.0248	0.0421	1.0000			
X15	Unemployed	-0.1308	-0.1705	-0.0704	-0.0607	0.0392	-0.0763	-0.0662	-0.0789	-0.0670	0.0140	0.0389	-0.0350	0.0254	-0.0456	0.0237	1.0000		
X16	Age	-0.0490	-0.0037	-0.0123	0.0603	-0.0844	0.0013	-0.0309	0.0715	0.0260	-0.2257	0.1223	0.1126	0.3258	0.2218	-0.0887	-0.1220	1.0000	
X17	Level of Education	0.1154	0.1218	0.1845	0.0816	-0.2775	0.1379	0.1527	0.0587	0.1134	0.0701	-0.1087	-0.0892	-0.2088	-0.1149	-0.0073	-0.0044	-0.2110	1.000

Table 3: The Determinants of Life Satisfaction and Happiness.

Variables	Response Categories	2002				2004			
		Happiness		Life Satisfaction		Happiness		Life Satisfaction	
		β	SD	β	SD	β	SD	β	SD
LEVEL 3 - COUNTRY									
Exogenous Contextual Effects (Z_k)									
Quality of Government		0.132	(0.267)	0.153	(0.397)	0.144	(0.163)	0.469	(0.151)**
National GDP		0.049	(0.309)	−0.115	(0.402)	0.495	(0.277)*	−0.284	(0.259)
Exogenous Social Interactions ($\overline{X}_k$)									
National Average Trust: Country's Parliament		−0.306	(0.444)	0.168	(0.583)	−0.599	(0.562)	1.330	(0.523)**
National Average Trust: Police		0.151	(0.191)	0.043	(0.297)	0.437	(0.326)	−0.669	(0.302)**
National Average Trust: People		0.625	(0.569)	0.582	(0.764	0.811	(0.333)**	0.058	(0.312)
National Average Political Interest		0.437	(0.226)**	0.004	(0.309)	0.189	(0.343)	−0.584	(0.320)*
National Average Religiosity		−0.288	(0.195)	−0.113	(0.274)	−0.418	(0.101)**	−0.371	(0.094)**
National Average Social Engagement		0.198	(0.332)	−0.433	(0.477)	0.030	(0.342)	−0.903	(0.318)**
National Average Social Intimacy		−0.520	(0.906)	1.337	(1.314)	2.043	(1.807)**	4.731	(1.683)**
National Average State of Health		−0.450	(0.302)	−0.108	(0.427)	−0.460	(0.443)	0.585	(0.410)
National Altruism		−0.575	(0.439)	−0.100	(0.580)	−0.674	(0.437)*	0.444	(0.406)
LEVEL 2 - REGIONS									
Exogenous Contextual Effects (Z_{jk})									
Regional GDP		0.142	(0.075)**	0.076	(0.065)	−0.045	(0.051)	−0.040	(0.047)
Location of Region									
Ref. Category: Atlantic	North	−0.002	(0.137)	0.056	(0.145)	0.007	(0.069)	0.172	(0.065)**
	Mediterranean	−0.206	(0.144)	−0.087	(0.150)	0.138	(0.080)*	0.217	(0.075)**
	Eastern EU-border	0.066	(0.141)	0.072	(0.145)	−0.092	(0.074)	−0.044	(0.070)
	Centre	−0.073	(0.132)	0.080	(0.134)	−0.029	(0.061)	0.061	(0.058)
Exogenous Social Interactions ($X_{jk} - \overline{X}_k$)									
Regional Average Trust: Country's Parliament		−0.108	(0.075)	−0.121	(0.070)**	−0.067	(0.072)	0.049	(0.068)
Regional Average Trust: Police		0.205	(0.091)**	0.106	(0.085)	0.130	(0.067)*	0.208	(0.063)**
Regional Average Trust: People		0.016	(0.098)	0.062	(0.093)	0.047	(0.072)	0.153	(0.068)**
Regional Average Political Interest		0.719	(0.081)**	0.436	(0.077)**	−0.061	(0.062)	−0.026	(0.058)
Regional Average Religiosity		0.339	(0.074)**	0.260	(0.069)**	0.081	(0.059)	0.132	(0.055)**
Regional Average Social Engagement		0.185	(0.076)	0.123	(0.071)**	−0.060	(0.059)	−0.095	(0.056)*
Regional Average Social Intimacy		0.031	(0.163)	0.001	(0.155)	−0.243	(0.116)**	−0.262	(0.111)**
Regional Average State of Health		0.131	(0.106)	0.026	(0.098)	0.293	(0.067)**	0.340	(0.063)**
Regional Average Altruism		0.240	(0.092)**	0.264	(0.085)**	0.139	(0.071)*	0.072	(0.067)

Table 3: The Determinants of Life Satisfaction and Happiness (contd.).

Variables	**Response Categories**	**2002**				**2004**			
		Happiness		Life Satisfaction		Happiness		Life Satisfaction	
		β	SD	β	SD	β	SD	β	SD
LEVEL 1 - INDIVIDUALS									
Exogenous Social Interactions ($X_{ijk} - \overline{X}_{jk}$)									
Trust: Country's Parliament		0.017	$(0.007)^{**}$	0.035	$(0.007)^{**}$	0.019	$(0.008)^{*}$	0.046	$(0.008)^{**}$
Trust: Police		0.089	$(0.007)^{**}$	0.089	$(0.007)^{**}$	0.119	$(0.008)^{**}$	0.107	$(0.007)^{**}$
Trust: People		0.041	$(0.007)^{**}$	0.060	$(0.007)^{**}$	0.066	$(0.007)^{**}$	0.065	$(0.008)^{**}$
Religiosity		0.050	$(0.007)^{**}$	0.038	$(0.007)^{**}$	0.066	$(0.007)^{**}$	0.048	$(0.007)^{**}$
Altruism		0.046	$(0.007)^{**}$	0.064	$(0.007)^{**}$	0.071	$(0.008)^{**}$	0.067	$(0.007)^{**}$
Social Intimacy		0.275	$(0.022)^{**}$	0.160	$(0.022)^{**}$	0.263	$(0.024)^{**}$	0.171	$(0.023)^{**}$
Social Engagement									
Reference Category: Never	Less Than Once A Month	0.020	(0.052)	0.010	(0.051)	0.174	$(0.055)^{**}$	0.201	$(0.053)^{**}$
	More Than Once A Month	0.128	$(0.051)^{**}$	0.074	(0.050)	0.212	$(0.054)^{**}$	0.287	$(0.052)^{**}$
	Several Times A Month	0.171	$(0.048)^{**}$	0.119	$(0.048)^{**}$	0.324	$(0.052)^{**}$	0.376	$(0.050)^{**}$
	Once A Week	0.221	$(0.048)^{**}$	0.146	$(0.048)^{**}$	0.334	$(0.052)^{**}$	0.388	$(0.050)^{**}$
	Several Times A Week	0.261	$(0.048)^{**}$	0.182	$(0.047)^{**}$	0.423	$(0.051)^{**}$	0.463	$(0.050)^{**}$
	Every Day	0.382	$(0.049)^{**}$	0.268	$(0.049)^{**}$	0.569	$(0.053)^{**}$	0.540	$(0.051)^{**}$
Political Interest									
Ref. Category: Very Interested	Quite Interested	−0.024	(0.025)	−0.014	(0.024)	0.044	$(0.026)^{**}$	0.026	(0.025)
	Hardly Interested	−0.067	$(0.020)^{**}$	-0.035	$(0.019)^{**}$	−0.060	$(0.020)^{**}$	−0.059	$(0.020)^{**}$
	Not At All Interested	−0.083	$(0.019)^{**}$	−0.037	$(0.019)^{**}$	−0.040	$(0.020)^{**}$	−0.038	$(0.019)^{*}$
State of Health									
Ref. Category: Very Good	Good	0.928	$(0.059)^{**}$	1.040	$(0.059)^{**}$	0.995	$(0.067)^{**}$	0.981	$(0.065)^{**}$
	Fair	0.678	$(0.058)^{**}$	0.801	$(0.058)^{**}$	0.713	$(0.066)^{**}$	0.734	$(0.064)^{**}$
	Bad	0.449	$(0.059)^{**}$	0.572	$(0.058)^{**}$	0.520	$(0.066)^{**}$	0.532	$(0.064)^{**}$
	Very Bad	0.205	$(0.063)^{**}$	0.322	$(0.062)^{**}$	0.191	$(0.070)^{**}$	0.135	$(0.068)^{*}$
Intrinsic Socio-demographic Indicators (W_{ijk})									
Citizen of Country		0.112	$(0.034)^{**}$	0.104	$(0.034)^{**}$	0.041	(0.035)	0.115	$(0.034)^{**}$
Victim of Crime		−0.046	$(0.015)^{**}$	−0.049	$(0.014)^{**}$	−0.019	(0.015)	−0.024	$(0.015)^{*}$
Ethnicity		−0.082	$(0.035)^{**}$	−0.063	$(0.035)^{*}$	−0.133	$(0.038)^{**}$	−0.163	$(0.037)^{**}$
Age									
Ref. Category: Aged less than 25	Aged between 25-34 years	−0.095	$(0.025)^{**}$	−0.076	$(0.025)^{**}$	−0.096	$(0.026)^{**}$	−0.115	$(0.026)^{**}$
	Aged between 35-44 years	−0.186	$(0.026)^{**}$	−0.153	$(0.025)^{**}$	−0.216	$(0.027)^{**}$	−0.203	$(0.026)^{**}$
	Aged between 45-54 years	−0.185	$(0.027)^{**}$	−0.154	$(0.027)^{**}$	−0.238	$(0.028)^{**}$	−0.203	$(0.028)^{**}$
	Aged between 55-64 years	−0.064	$(0.029)^{**}$	−0.006	(0.028)	−0.146	$(0.030)^{**}$	−0.048	$(0.029)^{*}$
	Aged Over 65 years	0.068	$(0.030)^{**}$	0.156	$(0.030)^{**}$	0.019	(0.031)	0.133	$(0.030)^{**}$
Marital Status									
	Married	0.334	$(0.018)^{**}$	0.210	$(0.017)^{**}$	0.345	$(0.019)^{**}$	0.237	$(0.019)^{**}$
	Divorced	−0.008	(0.027)	−0.026	(0.027)	0.083	$(0.028)^{**}$	0.067	$(0.027)^{**}$
	Living with Partner	0.223	$(0.020)^{**}$	0.138	$(0.020)^{**}$	0.264	$(0.023)^{**}$	0.166	$(0.022)^{**}$
	Widowed	−0.006	(0.029)	0.053	(0.029)	−0.026	(0.031)	0.069	$(0.030)^{*}$
Educational level									
	High-School	−0.035	$(0.016)^{**}$	−0.008	(0.016)	−0.040	$(0.017)^{**}$	−0.011	(0.016)
	Post-High	−0.050	$(0.024)^{**}$	−0.007	(0.024)	−0.010	(0.030)	0.009	(0.029)
	University	−0.061	$(0.018)^{**}$	−0.025	$(0.018)^{**}$	−0.070	$(0.019)^{**}$	−0.046	$(0.018)^{**}$

Table 3: The Determinants of Life Satisfaction and Happiness (contd.).

Variables	**Response Categories**	**2002**				**2004**			
		Happiness		Life Satisfaction		Happiness		Life Satisfaction	
		β	*SD*	β	*SD*	β	*SD*	β	*SD*
LEVEL 1 - INDIVIDUALS									
Intrinsic Socio-demographic Indicators (W_{ijk})									
Female		0.064	(0.012)**	0.078	(0.012)**	0.055	(0.013)**	0.036	(0.012)**
Unemployed		−0.289	(0.027)**	−0.332	(0.027)**	−0.231	(0.027)**	−0.312	(0.026)**
Household Income (Euros per month)									
Ref. Category: Less than 150	150 to under 300	−0.101	(0.071)	0.007	(0.070)	0.074	(0.083)	−0.072	(0.080)
	300 to under 500	−0.040	(0.067)	0.051	(0.067)	−0.005	(0.076)	−0.168	(0.074)**
	500 to under 1000	0.068	(0.065)	0.174	(0.064)**	−0.003	(0.072)	−0.125	(0.070)*
	1000 to under 1500	0.108	(0.065)*	0.209	(0.064)**	0.051	(0.072)	−0.062	(0.070)
	1500 to under 2000	0.125	(0.065)*	0.236	(0.065)**	0.111	(0.072)	−0.004	(0.070)
	2000 to under 2500	0.153	(0.066)*	0.310	(0.065)**	0.132	(0.073)*	0.035	(0.071)
	2500 to under 3000	0.171	(0.066)**	0.306	(0.066)**	0.221	(0.073)**	0.117	(0.071)*
	3000 to under 5000	0.158	(0.066)**	0.342	(0.066)**	0.197	(0.073)**	0.133	(0.070)*
	5000 to under 7500	0.197	(0.069)**	0.388	(0.069)**	0.262	(0.075)**	0.160	(0.073)*
	7500 to under 10000	0.192	(0.082)**	0.413	(0.081)**	0.026	(0.084)	0.128	(0.082)
	10000 or more	0.196	(0.093)**	0.453	(0.093)**	0.293	(0.089)**	0.212	(0.086)**
VARIANCE COMPONENTS									
	Σ_3	0.006	(0.004)**	0.021	(0.009)**	0.006	(0.003)*	0.005	(0.002)**
	Σ_2	0.026	(0.004)**	0.018	(0.003)**	0.003	(0.001)**	0.002	(0.001)**
	Σ_1	0.679	(0.007)**	0.669	(0.007)**	0.665	(0.007)**	0.629	(0.007)**
INTRACLASS CORRELATION									
	$\frac{\Sigma_1}{\Sigma_1+\Sigma_2+\Sigma_3}$	0.954		0.944		0.986		0.988	
	$\frac{\Sigma_2}{\Sigma_2+\Sigma_3}$	0.619		0.461		0.330		0.280	
EXCESS VARIANCE									
	Σ_3/Σ_2	0.23		1.166		2.000		2.500	
	Σ_2/Σ_1	0.038		0.026		0.005		0.003	
TEST VARIANCE									
$\chi^2(1\ df)$	$H_0 : \Sigma_3 = \Sigma_2$	0.109		2.084		0.658		1.668	
	$H_0 : \Sigma_2 = \Sigma_1$	1645.791**		838.41**		1444.008**		1578.139**	
	$H_0 : \Sigma_3 = \Sigma_1$	1271.293**		1594.11**		760.578**		840.901**	
NUMBER OF OBSERVATIONS		19875		19875		17932		17932	

The comparison group is male, aged less than 25, employed in labour force, whose educational level and household income is in the lowest category.

*: Significance level at 0.05 level, **Significance at 0.01 level.