HEDG
HEALTH, ECONOMETRICS AND DATA GROUP
THE UNIVERSITY of York

WP 11/20

Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases

Viola ANGELINI
Danilo CAVAPOZZI
Luca CORAZZINI
Omar PACCAGNELL

July 2011

Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases*

Viola ANGELINI,†Danilo CAVAPOZZI,‡Luca CORAZZINI,§Omar PACCAGNELLA¶

July 2011

Abstract

Self-reported life satisfaction is highly heterogeneous across similar countries. This phenomenon can be largely explained by the different scales and benchmarks adopted by individuals when evaluating themselves. We use cross-sectional data on the population aged 50 and over in ten European countries to compare estimates from a model in which reporting styles are assumed to be constant across respondents with those from a model in which anchoring vignettes are used to correct for individual-specific scale biases. We find that variations in response scales explain a large part of the differences found in raw data. Moreover, the cross-country ranking in life satisfaction significantly depends on scale biases.

JEL classification: C42, D12, I31, J14.
Keywords: Life satisfaction, scale biases, anchoring vignettes, counterfactuals.

*First version: October 2008. We thank Arthur Van Soest, Rob Alessie, Emanuela D'Angelo, Martin Leroch, Karsten Hank, Manfred J. Holler, Guglielmo Weber, Andrew Clark, Nicolai Kristensen, Kirsten Hager, Dirk Engelmann, Oliver Kirchkamp, Wieland Müller, Werner Güth, Giovanni Porzio, Teresa Bago d'Uva, Luigino Bruni and Benedetto Gui for useful comments and suggestions. We are also grateful to participants at seminars at the University of Groningen, the University of Hamburg, INED (Paris), the Max Planck Institute of Jena, the Paris School of Economics and at the conference "Well-being, Are We Happy with our Standard of Living?" (2008), the Nordic Conference on Behavioral and Experimental Economics (2008), the Netspar Pension Workshop (2009), BOMOPA (2009), the Workshop of Italian Behavioral and Experimental Economists (2009), the Conference "Happiness and Relational Goods" (2009), the 64th ESEM (2009), the AIEL Conference (2009), the ESA European Regional Meeting (2009), the Royal Economic Society Annual Conference (2010), IAREP (2010) and the New Directions in Welfare Congress (2011). Financial support from the European Commission through the STREP project COMPARE n. 028857 is gratefully acknowledged. This paper uses data from SHARE Wave 2. SHARE data collection was primarily funded by the European Commission through its 5th and 6th framework programmes (project numbers QLK6-CT-2001- 00360; RII-CT- 2006-062193; CIT5-CT-2005-028857). Additional funding by the US National Institute on Aging (grant numbers U01 AG09740-13S2; P01 AG005842; P01 AG08291; P30 AG12815; Y1-AG-4553-01; OGHA 04-064; R21 AG025169) as well as by various national sources is gratefully acknowledged (see http://www.share-project.org for a full list of funding institutions).

†Department of Economics, Econometrics and Finance, University of Groningen and Netspar [email: v.angelini@rug.nl].

‡Department of Economics and Management "Marco Fanno", University of Padua [email: danilo.cavapozzi@unipd.it].

§Department of Economics and Management "Marco Fanno", University of Padua; ISLA, Bocconi University [email: luca.corazzini@unipd.it].

¶Department of Statistical Sciences, University of Padua [email: omar.paccagnella@unipd.it].

1 Introduction

In analyzing data from a Eurobarometer survey, Kahneman et al. (2004) observe an implausible large difference in the self-reported level of life satisfaction between Denmark and France. Indeed, while 64 percent of the Danish respondents affirm to be "very satisfied" with their lives, only 16 percent of the French did so. As suggested by the authors, the puzzling differences in life satisfaction across seemingly similar countries can be caused by the fact that individuals, who are similar for both economic and non-economic conditions, use different benchmarks or scales in evaluating themselves (van Praag, 1971; Winkelmann and Winkelmann, 1998; Clark and Oswald, 2002; Ferrer-i-Carbonell and Frijters, 2004; Senik, 2004; Clark et al., 2005).[1] In psychometrics this phenomenon is called *differential item functioning* (DIF), defined as the inter-personal and inter-cultural variation in interpreting and using the response categories for the same question (Holland and Wainer, 1993).

On the one hand, scale differences can depend on group-specific characteristics such as socio-economic conditions (Frey and Stutzer, 2002b) or cultural connotations (Uchida et al., 2004; Diener and Suh, 2000; Inglehart and Klingemann, 2000; Jürges, 2007). For instance, there is evidence showing that seemingly similar European countries are associated with highly dissimilar self-reported levels of life satisfaction (Inglehart and Rabier, 1986). Rather than being caused by the existence of objective differences in the conditions of the inhabitants, this puzzling finding can be explained by "*different interpretations of numerical scales*" across societies (Frey and Luechinger, 2007, page 220) or with "*cultural differences in the norms that govern self-descriptions*" (Kahneman et al., 2004a, page 430). The idea is that different groups of people use systematically different norms for what should be called "very good", "good", etc.

On the other hand, scale differences can be determined by individual-specific characteristics such as the psychological status of the individual (De Neve and Cooper, 1998), her personal interpretations of life satisfaction, or her intellectual skills in understanding questions and processing the information needed to answer (Frey and Stutzer, 2002b).

How to correct for scale differences is still an open question for social scientists. In cross-sectional data it is not usually possible to account for individual heterogeneity. When panel

[1] For instance, Clark et al. (2005, page 118) affirm that "one worry regarding statistical analysis of subjective variables is that some people look at life pessimistically or optimistically, even though there is really no difference in their level of well-being."

data are available, conventional fixed or random effects (Kapteyn et al., 2007) or latent class techniques (Clark et al., 2005) can be used in order to take into account time-invariant DIF bias. However, given the psychological nature of life satisfaction, the hypothesis of individuals' time-invariant scales which these methodologies are based on can be criticized. Indeed, the scale adopted by an individual to evaluate herself can vary over time according to her actual mood (Kahneman et al., 2004b) and her socio-economic conditions.

In this paper, we apply a vignette methodology to cross-sectional data from ten European countries to assess whether differences in self-reported life satisfaction are genuine or they just reflect individual-specific scale differences. Following this approach, individuals are presented with two categories of questions on life satisfaction. First, as in traditional questionnaires, respondents are asked to self-report their level of life satisfaction. Second, they are asked to evaluate, on the same scale on which they evaluate themselves, the level of life satisfaction of one or more hypothetical persons described in given conditions (*anchoring vignettes*). In such a way it is possible to find a standard, that is an anchor to which the response categories of the survey questions will be attached. Therefore, by collecting individuals' evaluation of the anchoring vignettes it is possible to correct the level of self-reported life satisfaction from the DIF bias and enhance the comparability of subjective assessments across individuals.

We find that the ranking across countries strongly depends on differences in the scales used by individuals to report their life satisfaction. When no correction for differences in reporting styles is introduced, Danes and Italians result to be the most and the least satisfied with life, respectively. On the contrary, when we apply the vignette methodology, the differences between Danes and Italians as well as those between Danes and French disappear and the Netherlands and Czech Republic respectively replace Denmark and Italy in the ranking of life satisfaction. Interestingly, correcting for differences in reporting styles makes the cross-country ranking in the level of life satisfaction more in line with those implied by more objective indicators, such as the Human Development Index.

The rest of the paper proceeds as follows. In section 2, we describe the dataset and we present descriptive statistics of the variables used in our analysis placing particular emphasis on the structure of the vignettes. In section 3, we specify the econometric model based on the vignettes (the Hopit model). In section 4, we compare estimates from an Ordered Probit model in which scales are assumed to be constant across individuals with those from the more

general Hopit model and we present formal tests to evaluate their estimation performances. In section 5 we discuss and test the identifying assumptions on which the model is based on. In section 6, by using the Hopit model, we estimate the country-specific scales of life satisfaction and we present results from counterfactual simulations to study how the distribution of life satisfaction in a country changes when the response scale of other countries is imposed to its inhabitants and, in addition, when all respondents are assumed to live in the same country. Finally, in section 7 we discuss our findings and we conclude.

2 Data and Descriptive Statistics

Our sample is drawn from the 2006-2007 wave of the Survey of Health, Ageing and Retirement in Europe (SHARE) that contains information on both the economic and non economic conditions of European citizens aged 50 and over. After the personal interview (CAPI, Computer Assisted Personal Interview), as part of the COMPARE project, a subset of respondents are asked to fill out additional questions on a paper and pencil questionnaire, which focuses on self-assessments and vignette evaluations. In particular, respondents are asked to rate their life satisfaction using the question "How satisfied are you with your life in general?". This self-evaluation is followed by two anchoring vignettes that allow us to implement the methodology proposed by King et al. (2004) to control for individual heterogeneity in reporting styles. In particular, the following vignettes are used in our analysis:

1. John is 63 years old. His wife died 2 years ago and he still spends a lot of time thinking about her. He has 4 children and 10 grandchildren who visit him regularly. John can make ends meet but has no money for extras such as expensive gifts to his grandchildren. He has had to stop working recently due to heart problems. He gets tired easily. Otherwise, he has no serious health conditions. How satisfied with his life do you think John is?

2. Carry is 72 years old and a widow. Her total after tax income is about € 1,100[2] per month. She owns the house she lives in and has a large circle of friends. She plays bridge twice a week and goes on vacation regularly with some friends. Lately she has been suffering from arthritis, which makes working in the house and garden painful. How satisfied with her life do you think Carry is?

[2] The value is PPP-adjusted.

Both for the self-evaluation and the vignettes, respondents answer using the following 5-point scale: "very dissatisfied", "dissatisfied", "neither satisfied, nor dissatisfied", "satisfied", "very satisfied".

Overall, our sample contains 5,606 individuals living in Sweden, Denmark, Germany, The Netherlands, Belgium, France, Spain, Italy, Poland and Czech Republic. Figure 1 reports how respondents rate their own level of life satisfaction.

[Figure 1 about here]

As documented in previous studies (Inglehart and Rabier, 1986; Kahneman et al., 2004), life satisfaction is characterized by clear cross-country heterogeneity. The sharpest differential derives from the comparison between Denmark and Italy. While in the former country the proportion of individuals either satisfied or very satisfied with their life is higher than 90 percent, in the latter it is slightly above 60 percent. In general, the picture shows that Denmark has the highest level of life satisfaction, followed by The Netherlands, Sweden and Germany, while Italy is the country with the lowest self-reported satisfaction.

With the exception of Poland and Czech Republic, these differences are somewhat striking in that they refer to a set of countries comparable in many dimensions, such as quality of life, social and economic inclusion and activism of the welfare state. As anticipated in the introduction, a rationale for this evidence is that individuals are affected by differential item functioning biases in evaluating themselves.

Analyzing the distribution of vignette evaluations at a pure descriptive level may help to understand how they can be used to correct self-assessments from individual heterogeneity in reporting styles. Figure 2 shows how respondents rate the level of life satisfaction of the persons described in the two vignettes (John and Carry) for each country.

[Figure 2 about here]

The figure shows that response scales are not invariant across countries. While in Denmark

and Germany more than 20 percent of respondents describe John as satisfied or very satisfied with his life, the proportion drops to less than 10 percent for France and Italy. We also find that Spain, Poland and Czech Republic adopt reporting styles closer to those of Germany, whereas the remaining countries are more in line with France and Italy. Cross country differences are also found in the assessment of the second vignette. As before, while Danish respondents are more likely to consider Carry as either satisfied or very satisfied with her life, the French are more reluctant to use the most extreme labelling of vignette evaluations. Remarkably, while nearly 80 percent of the Danes find Carry either satisfied or very satisfied with her life, this figure falls to less than 40 percent for the French and lies between 50 and 60 percent for the remaining countries.

To summarize, we find evidence suggesting that individuals living in different countries may adopt different reporting styles in life satisfaction self-assessments. Relying on such evaluations may then lead to misleading conclusions and emphasize differences that are mainly due to incomparable response scales rather than true differences in life satisfaction. As an example, although Figure 1 shows that the Danes are more likely to describe themselves as satisfied of very satisfied with their life than the French, in Figure 2 we find that they are also more likely to use the modalities at the top of the life satisfaction scale when asked to evaluate the well-being of the same hypothetical persons described in the vignettes. Our estimation method will exploit the variability in vignette evaluations to assess to what extent the differences in Figure 1 are genuine or they just reflect differences in the response scales used by respondents.

3 The econometric model

Anchoring vignettes were first introduced by King et al. (2004) for analyzing ordinal survey responses taking into account individual differences in the interpretation of the survey questions. Vignettes are indeed a new tool for enhancing self-report data comparability across individuals. Under the assumption that the situation described in the vignettes is perceived by respondents in the same way (*vignette equivalence*), variability in vignette evaluations is only due to the different reporting styles adopted. Hence, if the same response style is used for both self-ratings and vignette evaluations (*response consistency*), the additional information provided by vignettes acts as an *anchor* to adjust the self-assessments of different individuals

according to a homogenous classification allowing for inter-personal comparisons. Our econometric specification is usually referred to as the Hopit model (King et al., 2004). It mainly consists of two components modelling self-assessments and vignette evaluations as standard ordered variables.

Let us denote with Y_i^* the life satisfaction perceived by individual $i = 1, ..., n$ and assume that it is the result of the linear specification

$$Y_i^* = X_i\beta + \varepsilon_i; \tag{1}$$
$$\varepsilon_i | X_i \sim N(0, 1),$$

where X_i includes observed covariates, β is a set of parameters to estimate and ε_i is a stochastic component normally distributed and encompassing unobserved factors relevant for the determination of life satisfaction levels.

Although we do not observe Y_i^*, survey questions pick up its discrete counterpart Y_i, which is the answer to the life satisfaction self-assessment and it is recorded as an ordered variable taking on values 1 ("very dissatisfied"), ..., 5 ("very satisfied"). In particular,

$$Y_i = j \quad \text{if} \quad \tau_i^{j-1} < Y_i^* \leq \tau_i^j, \; j = 1, ..., 5. \tag{2}$$

The thresholds τ_i^j are individual-specific and are given by

$$\tau_i^0 = -\infty\,; \quad \tau_i^5 = \infty;$$
$$\tau_i^1 = X_i\gamma^1; \tag{3}$$
$$\tau_i^j = \tau_i^{j-1} + \exp(X_i\gamma^j), \quad j = 2, 3, 4. \tag{4}$$

Allowing the thresholds to vary across individuals entails that respondents with the same perceived life satisfaction Y^* may report different self-assessments Y because of different sets of cut-off points τ^j used. This is the main difference with respect to standard ordered probit specifications, which formally state that the same thresholds are utilized by the whole population of interest. The exponential function in (4) has been chosen to guarantee monotonicity of the thresholds.

Using self-reports on life satisfaction alone is not enough to separately identify the para-

meters in β and γ. In fact, if a covariate in X influences both the self-assessment Y^* and the thresholds τ^j, we are not able to distinguish these two effects on the basis of the information conveyed by a single self-evaluation. To achieve this goal we need at least two distinct evaluations in which the same response scale is adopted.

In our sample each respondent is asked to answer two vignette questions. We denote with Z^*_{il}, $l = 1, 2$ the variable indicating how the actual level of the domain of interest described in the vignette l is perceived by respondent i. We assume that

$$\begin{aligned} Z^*_{il} &= \theta_l + \nu_{il}; \\ \nu_{il} &\sim N(0, \sigma^2_v), \end{aligned} \tag{5}$$

where θ_l is the actual level of the domain of interest described in the vignette l and ν_{il} is a stochastic component assumed to be independent of ε_i. The requirement of vignette equivalence assumes that the situation described in the vignettes is perceived by respondents in the same way and formally restricts θ_l to not vary over i.

Although the actual perceived value Z^*_{il} is unobserved, survey questions gather respondent evaluations of vignettes according to the same 5-point scale used for self-assessments. As a result, we are provided with the ordered response Z_{il} such that

$$Z_{il} = j \quad \text{if} \quad \tau_i^{j-1} < Z^*_{il} \leq \tau_i^j, \quad j = 1, ..., 5. \tag{6}$$

It is worth noting that the same set of thresholds is also found in equation (2) and this follows from the response consistency hypothesis claiming that the same reporting styles are used for both self-assessments and vignette evaluations.

In this set-up the specifications modelling self-assessment and vignette ordered responses are connected via the utilization of the same set of thresholds. This implies that the information relevant to estimate equation (1) and equation (5) in the sample should be combined to estimate the common set of parameters showing up in the threshold equations (3) and (4).

4 Results

Starting from the famous Easterlin's paradox that states the existence of a reversed U-shaped relation between happiness and income in the US between 1946-1996 (Easterlin, 2001), social scientists have devoted a large quantity of research using self-reported data to identify the main determinants of life satisfaction. As an established result (see Frey and Stutzer, 2002a and 2002b; van Praag and Ferrer-i-Carbonell, 2004; Bruni and Porta, 2005; Dolan et al., 2008 for surveys), "*money is not enough to make people happy*" and in addition to economic factors, there are non-economic variables that exert a significant and positive effect on life satisfaction. In our estimates, we control for a large number of factors that are associated with life satisfaction: demographic characteristics (age and gender), socio-economic variables (employment, income and wealth and education), health (number of chronic diseases, arthritis, symptoms, limitations with mobility, limitations with activities of daily living - ADL - and instrumental activities of daily living - IADL-, obesity and having been diagnosed with depression) and social relationships (marital status, family bonds and extra-familiar activities), as well as country dummies.[3] The description of the variables used in the empirical analysis is reported in Table 1.

[Table 1 about here]

In Table 2 we present the results for the self-assessment equation of life satisfaction with heterogeneous response scales (second column), comparing it with a baseline model not allowing for any threshold variation across respondents (first column).[4]

[Table 2 about here]

The equation in the latter model is almost identical to an ordered probit model, not taking into account potential differences in reporting styles: in the presence of scale differences across

[3] Notice that, rather than being interested in establishing causal relations between variables, our aim is to identify the factors with which life satisfaction is associated and identify whether these relations are genuine or they just reflect differences in response scales.

[4] In order to control for within-household correlation, we also estimate the same Hopit specification allowing for unobserved household-specific effects. The results confirm those of the model discussed in this section and are available upon request.

countries or socio-economic groups, the parameter estimates of this model will reflect both true life satisfaction effects and the effects of reporting heterogeneity. The results for the threshold equations are presented in columns 3 to 6. The estimates show that the thresholds significantly depend on a number of variables, such as country dummies, age, education, employment and marital status, several health conditions (chronic diseases and symptoms of diseases, arthritis, mobility problems and having been diagnosed with depression), social activities and wealth. Indeed a formal likelihood-ratio test strongly rejects the model not allowing for response scale variation against the more general model that does allow for correction of the DIF bias.[5] Table 3 displays the parameters estimates of the vignette equation (5) for both the Ordered Probit and the Hopit model. The θ parameter associated to the first vignette (John) is lower than the one referring to the second vignette (Carry). On average, Carry is perceived by respondents in our sample as more satisfied with her life than John. This difference is statistically significant and in line with the evidence summarized in Figure 2.

[Table 3 about here]

We first focus on the impact of different response scales on cross country variations in self-reported life satisfaction. As shown in Table 2, our specifications assign to the coefficients on country dummies the corresponding deviations from Germany (baseline). Parameter estimates as well as their 95% confidence intervals are plotted in Figure 3.

[Figure 3 about here]

Moreover, for both the Ordered Probit model and the Hopit model, we report in Table 4 differences between estimates of the country dummies for each pairwise comparison as well as the significance levels from a Wald test for the null hypothesis of equal estimates.

[Table 4 about here]

[5] $\chi^2_{156} = 1103.177$, p-value = 0.000. We also test the joint significance of all the coefficients but the constant in the threshold equations separately for each threshold and the null hypothesis is always rejected ($\gamma_1 : \chi^2_{39} = 230.07$, p-value = 0.000, $\gamma_2 : \chi^2_{39} = 116.50$, p-value = 0.000, $\gamma_3 : \chi^2_{39} = 143.63$, p-value = 0.000, $\gamma_4 : \chi^2_{39} = 170.47$, p-value = 0.000).

In the model without corrections for differences in scales, Denmark is the country with the highest reported level of life satisfaction, while Italy is the one with the lowest. As regards the other countries, Sweden and the Netherlands rank higher than Germany, while France and Czech Republic, together with Italy, are the only countries with a level of life satisfaction that is lower than in the baseline country. When we correct for the DIF bias, the ranking of countries significantly changes. The most striking result is that now the life satisfaction of the Danes is not significantly different from that of the Italians, while the Netherlands, Sweden and France all rank higher than Denmark. Czech Republic, on the other hand, reports the lowest level of life satisfaction.

Interestingly, the ranking obtained by correcting for differences in reporting styles is very similar to that implied by other (and more objective) indicators, such as the *Human Development Index* (HDI).[6] In particular, while we find a positive and highly significant correlation between the country fixed effects from the Hopit model and the correspondent 2006 HDI (Spearman's rank correlation coefficient: 0.900; p-value$<$0.01), the relation is much weaker when we use the fixed effects from a model with no correction for differences in reporting styles (Spearman's rank correlation coefficient: 0.584; p-value$<$0.1).

The parameter estimates for the other variables are sensible and consistent with the literature. The relation between age and life satisfaction is non-linear: life satisfaction increases up to age 85 and then gradually decreases (Clark and Oswald, 1996; Yang, 2008). Note that the age-profile correcting for the DIF bias is similar to that found with the model not accounting for threshold variation across respondents but it is steeper, pointing to more marked age differences.[7] Women seem to be happier than men and married individuals are more satisfied with life than unmarried ones, whose level of life satisfaction is not significantly different from that of divorced and widowers. As regards family bonds, parents who have a child with whom they have less than weekly contacts are the most dissatisfied with their life. In other words, it is better not having any child than having a child and not being in contact with her. Having a

[6] The HDI is an index used to rank countries by level of "human development", which is obtained by combining three dimensions: a. Life expectancy at birth, as an index of population health and longevity; b. Knowledge and education, as measured by the adult literacy rate (with two-thirds weighting) and the combined primary, secondary, and tertiary gross enrolment ratio (with one-third weighting); c. Standard of living, as measured by the natural logarithm of gross domestic product per capita at purchasing power parity. The ranking of the countries considered in our analysis according to the 2006 HDI (for references, see http://hdr.undp.org/en/statistics/indices/) is as follows: the Netherlands (0.958), Sweden (0.958), France (0.955), Denmark (0.952), Spain (0.949), Belgium (0.948), Italy (0.945), Germany (0.940), Czech Republic (0.897), Poland (0.875).

[7] However, with cross-sectional data we cannot distinguish between age, cohort and time effects.

living parent or a grandchild is significantly and positively correlated with life satisfaction only if the contacts with her are frequent (at least once a week).

Not surprisingly, health problems are negatively associated with life satisfaction. In particular, reporting at least two symptoms of diseases, limitations with mobility and with activities of daily living (ADL and IADL) and having being diagnosed with affective or emotional disorders, all are significantly and negatively related to satisfaction with one's own life (Easterlin, 2003).

We also classify individuals by their employment status as employed (either employee or self-employed), retired or out of the labour force (that includes unemployed, homemakers and the disabled). Our results show that the retired are worse off than the employed (Charles, 2004) but are more satisfied with their life than those out of work for other reasons.

For a given employment status, taking part in social activities, such as voluntary or charity work, caring for a sick person, educational courses and social clubs, is associated to higher levels of life satisfaction.

The overall life satisfaction is positively and significantly correlated with wealth, while income does not seem to play any role. Our explanation for this result is that, given the nature of our dataset, what really matters for the elderly is the saving they have accumulated throughout their life and not current income. Even controlling for income and wealth, education is positively correlated with life satisfaction.

We also estimate a richer specification that adds controls for the political and religious background of the respondent.[8] to the covariates we used before The results show that life satisfaction is positively correlated with the frequency of praying, the most satisfied being those who pray more than once a day. This evidence is consistent with the findings of Clark and Lelkes (2005), who show that people become happier the more often they attend church and the more often they pray. However, there seems to be no difference in life satisfaction between those who never pray and those who pray less than once a week. As for politics, even if we control for income, wealth and occupational status, life satisfaction increases when moving from left to right. The estimates on the other variables remain qualitatively unchanged.

[8] The estimates are available upon request from the authors. France is not included in the sample because in this country the questions about religion and politics were not asked for privacy reasons.

5 Discussion and testing of the identifying assumptions

The validity of the vignette approach to identify reporting heterogeneity relies on two assumptions, vignette equivalence and response consistency.

As stated in equation (5), vignette equivalence claims that, on average, the life satisfaction level of the hypothetical persons described in the vignettes is perceived similarly by different respondents. One might argue that respondents in different countries with different institutional settings might not perceive the situation described in the vignettes in the same way. In a country with a more developed welfare state, the same situation might be considered less problematic than in a country with a less generous social safety net. However, the descriptive evidence reported in Figure 2 does not seem to support this interpretation. For example, the percentage of respondents who consider John and Carry as either satisfied or very satisfied with their life is higher in Poland and Czech Republic than in countries like Sweden and Denmark (for John it is double in Czech Republic than in Sweden). Moreover, in countries with very similar institutional settings, such as Sweden and Denmark or Italy and Spain, there is substantial variation in the way in which respondents evaluate the same vignettes. Note also that there is consistency in the responses to the vignette questions, since in each country Carry is always rated as more satisfied with her life than John. This result is confirmed in all countries even if we control for the individual and household characteristics included in the Hopit estimation.

We also perform several formal tests to check the validity of the vignette equivalence assumption.

First, we use a test based on the global ordering of vignettes (Murray et al., 2003; Kristensen and Johansson, 2008; Rice et al., 2010). A minimal condition for the assumption of vignette equivalence to hold is that individual responses are consistent with the global ordering of vignettes. The global ordering can be obtained by pooling all the responses across countries and considering the average categorical response for each vignette (Murray et al., 2003). We then compute the average percentage of respondents in each country that gave an ordering of vignettes consistent with the global ordering. The results are shown in Table 5. In each country the percentage of respondents who are consistent with the global ordering of vignettes is equal or greater to 89%, supporting the vignette equivalence assumption. As a robustness check, we also calculate the average percentage of respondents who gave an ordering of vignettes

consistent with the global ordering when countries are stratified in groups that share similar social norms and values according to the Inglehart-Welzel cultural map of the world (as in Rice et al., 2010, and Kristensen and Johansson, 2008): Protestant Europe (Sweden, Denmark, The Netherlands and Germany), Catholic Europe (Belgium, France, Spain and Italy) and Ex-Communist countries (Czech Republic and Poland). The average percentages are between 92 and 93, with a very small variation across groups, providing further support to the vignette equivalence hypothesis. Similar percentages are found when the sample is stratified according to each explanatory variable included in our empirical specification.

[Table 5 about here]

Second, we re-estimate the Hopit model stratifying countries according to the Inglehart-Welzel map. If the vignette equivalence assumption holds and the model is correctly specified, then the coefficients estimated separately on the three groups of countries should be similar to those estimated on the pooled sample. The idea behind this test, first proposed by Kristensen and Johansson (2008), is that, if the countries in our sample were too different for the vignette equivalence to hold, then we would expect very different coefficients when estimating the model separately by groups of countries with similar cultures, where the vignette equivalence assumption is more likely to be supported by the data. We focus our attention on the country dummy estimates and on the θ parameters, which measure the average level of life satisfaction of the persons described in the vignettes as perceived by respondents. Results are displayed in Table 6. Our sensitivity analysis shows that the estimates of the country dummy parameters are left virtually unaffected by stratifying the sample and the conclusions in terms of cross-country comparisons are confirmed. Most importantly, also the θ parameters preserve the same ranking found in the pooled estimation. In fact, in all groups of countries Carry is perceived as more satisfied with her life than John and this difference is always statistically significant. Overall, the results of all our tests support the vignette equivalence assumption.

[Table 6 about here]

As regards response consistency, there is a growing empirical literature that tries to test the validity of this assumption (Van Soest et al., 2007; Bago d'Uva et al., 2009; Datta Gupta et

al., 2009). The tests proposed in the literature require the availability of an objective measure of the construct of interest. For example, Van Soest et al. (2007) apply the vignette approach to investigate drinking behaviour. In their study they take advantage of the information on the number of drinks consumed by the respondents and find that vignette based corrections appear quite effectively in bringing objective and subjective measures closer together. However, given the multidimensional nature of life satisfaction, defining an objective measure at the individual level is arguable.[9] Therefore, we propose a new way to validate the response consistency assumption. This hypothesis states that individuals adopt the same response style when evaluating themselves and when evaluating the vignettes. Hence, if this assumption holds, respondents who closely match the hypothetical individual described in one of the two vignettes should give a similar evaluation of their own level of life satisfaction and that of the vignette person. Our results show that, as we select respondents with characteristics increasingly similar to those of the person described in the vignette, the correlation between the self-report and the vignette evaluation becomes closer to 1 (of course the sample size becomes also very small). For example, in the case of John we start by selecting respondents who are males and aged between 65 and 75 and then we gradually match more and more closely the description of John up to the point of selecting only males aged between 55 and 65, who are neither in the lowest nor in the highest income and wealth quartiles, who have children and grandchildren, have no serious limitation with activities of daily living, do not work and are widowed[10]: as we "get closer" to John, the correlation between the self-reports and the vignette evaluation increases from 0.09 to 0.80. In the case of Carry we find similar results. Therefore, our results support the response consistency assumption.

6 Counterfactuals

We now present results of counterfactual simulations to assess the relevance of the DIF bias across countries. In particular, we simulate the distribution of self-reported levels of life satisfaction that would have been provided by respondents if they were asked to evaluate themselves according to the (estimated) scale of another country. First, for all respondents in our sample we

[9] Although the vignette methodology has been traditionally applied to unidimensional concepts, Kapteyn et al. (2010) show that the multidimensional nature of life satisfaction does not invalidate the anchoring vignette approach.

[10] Note that we cannot go as far as selecting only respondents identical to John because of the reduced sample size.

calculate the average perceived value of life satisfaction given their own characteristics[11]. Then, we predict their thresholds setting the dummy of the reference country equal to one irrespective of respondents' actual country of residence but still conditioning on their other characteristics. The comparison between the perceived life satisfaction and the so-defined thresholds gives the prediction of the self-reported life satisfaction provided by respondents according to the reporting style of the reference country. In what follows we classify individuals as satisfied with their life if they are either satisfied or very satisfied. In Figure 4 we compare how respondents would rate their own level of life satisfaction if they used the Danish and Italian response scales respectively.

[Figure 4 about here]

Using the Danish thresholds, more than 95 percent of respondents in all countries would rate themselves as satisfied or very satisfied with their own life. This means that, given the same true level of life satisfaction, the Danes are more likely to rank themselves high in the 5-point response scale for life satisfaction. The picture considerably changes when using the Italian response scales. Now cross country heterogeneity is much more evident: the proportion of individuals satisfied or very satisfied with their life decreases in all countries. In Poland it drops from about 96 percent to 60 percent and similar differences can be seen in all countries. To sum up, given the same true level of life satisfaction, the Danes are more likely to rank themselves high in the 5-point response scale for life satisfaction.

Similar conclusions are reached by looking at Figure 5.

[Figure 5 about here]

On the basis of the estimates of the parameters in the threshold equations, we are able to associate to each individual the cut-off points used in equation (3). Figure 6 shows the medians of the individual-specific thresholds by country. We report medians rather than means because

[11] More formally, we calculate the conditional expectation $E[y^*|X] = X'\widehat{\beta}$.

they are more robust to the presence of outliers.[12] The fact that the third threshold, which is the one that determines whether someone is satisfied with her own life, is lower for Denmark than for the other countries implies that the Danes will be more likely to report that they are satisfied, *ceteris paribus*. Thus, variations in response scales explain a large part of the differences found in raw data.

When we control for response scale heterogeneity, cross-country differences decrease but some variability still remains. This variability might be due either to institutional differences or to differences in the composition of the sample with respect to the covariates used in the estimation. In the second counterfactual exercise, we simulate the distribution of life satisfaction if all respondents used the same response scales and lived in the same country, thus facing the same set of institutional constraints and general socioeconomic conditions. We predict the perceived value of life satisfaction and the thresholds of each respondent by setting to one the dummy corresponding to the reference country and conditioning on the other respondents' characteristics.

The results are presented in Figure 6 and, as in the previous exercise, the reference countries are Denmark and Italy.

[Figure 6 about here]

Since in both counterfactuals respondents use the same thresholds, the comparison between Figure 4 and Figure 6 highlights the institutional effects of living in Denmark and Italy rather than in the actual country of residence once conditioning on the thresholds. As an example, when the Danish thresholds are used, simulating that all respondents live in Denmark seems not to produce sizeable effects on their probability of being satisfied or very satisfied with their life. This evidence implies that the low proportion individuals not satisfied with their life found in raw data for the Danes is mainly driven by their propensity towards using the modalities at the top of the self-assessment scale and not to cross-country institutional differences. On the contrary, simulating that all respondents live in Italy produces reductions in the probability of being satisfied or very satisfied for all countries but Czech Republic.

[12] Anyway, we take variability into account in the estimates of the Hopit model.

While Figure 4 reports that 97 percent of the Swedes and 92 percent of the French would rate themselves as at least satisfied with their life if they adopted the Italian response scale, Figure 6 shows that these proportions would fall to 88 percent and 81 percent respectively if they were not only using the Italian thresholds but they were also living in Italy. The reverse pattern is found for Czech Republic. While in Figure 4 less than 60 percent of the Czechs are at least satisfied with their life, this proportion increases by more than 40 percent in the second counterfactual simulation.

The remaining cross-country variability that we observe in Figure 6 is due only to differences in the composition of the sample across countries, which, nevertheless, are taken into account in the estimates.

7 Conclusion

Is the true level of life satisfaction in Denmark really higher than in Italy? Does cross-country comparability depend on how respondents interpret subjective survey questions?

In this paper, we address these questions by applying the vignette methodology to data from the 2006-2007 wave of the Survey of Health, Ageing and Retirement in Europe (SHARE), a cross-sectional dataset containing demographic, physical, social and economic information of individuals aged 50 and over living in several European countries.

Although vignettes have already been successfully used in several domains, such as political efficacy (King et al., 2004), health (Salomon et al., 2004; Bago d'Uva et al., 2008), employer preferences (van Beek et al., 1997), work disability (Kapteyn et al., 2007; van Soest et al., 2006) and job satisfaction (Kristensen and Johansson, 2008) to correct for differences in the scales used by respondents across countries and socio-economic groups in cross-sectional data, to our knowledge this is the first empirical study applying this methodology to self-reported life satisfaction.

The main results of our paper can be summarized as follows. We compare estimates from a model in which scales are assumed to be constant across individuals (Ordered Probit) with those from a model in which vignettes are used to correct for the DIF bias (Hopit model). According to the Ordered Probit, after controlling for economic, demographic, health and social conditions, Danes and Italians result to be the most and the least satisfied with life respectively.

However, by correcting for scale differences, the ranking across countries significantly changes. The difference in self-reported life satisfaction between Danes and Italians disappears and the Netherlands and Czech Republic respectively replace Denmark and Italy in the ranking of life satisfaction. A formal likelihood ratio test strongly rejects the ordered probit not allowing for response scale variation against the more general Hopit model. We find evidence that the thresholds significantly depend on the explanatory variables used in the regressions. We also present results from counterfactual simulations to study how the distribution of self-assessed life satisfaction in a country changes when the response scales of other countries are imposed on its inhabitants. When the Danish scale is used, more than 95 percent of respondents in all countries would rate themselves as satisfied or very satisfied with their own life. The picture considerably changes when using the Italian response scale. Indeed, self-reported life satisfaction drops significantly in all countries when the Italian scale is imposed on respondents. Thus, variations in response scales explain a large part of the differences found in raw data. Concerning the determinants of life satisfaction, after controlling for heterogeneity in reporting scales, we find that life satisfaction is positively and significantly correlated with being married, being female, being not retired and having not experienced unemployment, having frequent (almost weekly) contacts with sons, parents and grandchildren, enjoying good health conditions, being active in the community and volunteering, having high wealth.

Our results inform the empirical literature studying the determinants of life satisfaction about the importance of correcting for individual-specific scale differences, such as cultural connotations of respondents, to produce more reliable cross-country comparisons.

References

Bago d'Uva, Teresa, Eddy van Doorslaer, Marteen Lindeboom, and Owen O'Donnell. 2008. "Does Reporting Heterogeneity Bias the Measurement of Health Disparities? Health Economics, 17: 351-375.

Bago d'Uva Teresa, Martin Lindeboom, Owen O'Donnell and Eddy van Doorslaer. 2009. "Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity." HEDG Working Paper 09/30, Health and Econometrics Data Group, The University of York.

Bruni, Luigino, and Pier Luigi Porta. 2005. Economics and Happiness. Framing the Analysis. Oxford: Oxford University Press.

Charles, Kerwin Kofi. 2004. "Is Retirement Depressing? Labor Force Inactivity and Psychological Well-Being in Later Life." In Accounting for Worker Well-Being, Research in Labor Economics, ed. Solomon W. Polachek, vol. 23, 269-299. Amsterdam, San Diego and Oxford: Elsevier, JAI.

Clark, Andrew E. 2003. "Unemployment as a Social Norm: "Psychological Evidence from Panel Data." Journal of Labor Economics, 21, 289-322.

Clark, Andrew E., and Orsolya Lelkes. 2005. "Deliver us from Evil: "Religion as Insurance." PSE, Paris, November.

Clark, Andrew E., Fabrice Etilè, Fabien Postel-Vinay, Claudia Senek, and Karine Van der Straeten. 2005. "Heterogeneity in Reported Well-Being: Evidence from Twelve European Countries." Economic Journal, 115: C118-C132.

Clark, Andrew E., and Andrew J. Oswald. 1996. "Satisfaction and Comparison Income." Journal of Public Economics, 61: 359-381.

Clark, Andrew E., and Andrew J. Oswald. 2002. "Well-Being in Panels." DELTA, mimeo.

Datta Gupta, Nabanita, Nicolai Kristensen and Dario Pozzoli. 2009. "External Validation of the Use of Vignettes in Cross-Country Health Studies." IZA Discussion Paper 3989, Institute for the Study of Labor, Bonn.

De Neve, K.M., and Cooper, H. 1999. "The Happy Personality: a Meta-Analysis of 137 Personality Traits of Subjective Well-Being." Psychological Bulletin, 125: 197-229.

Diener, Ed, and Eunkook M. Suh. 2000. "Culture and Subjective Well-Being." Cambridge

MA: MIT Press.

Dolan, Paul, Tessa Peasgood, and Matthew White. 2008. "Do We Really Know What Makes Us Happy? A Review of the Economic Literature on the Factors Associated with Subjective Well-Being." Journal of Economic Psychology, 29: 94 - 122.

Easterlin, Richard A. 2001. "Income and Happiness: Towards a Unified Theory." Economic Journal, 111: 465-84.

Easterlin, Richard A. 2003. "Happiness of Women and Men in Later Life: Nature, Determinants, and Prospects". In Advances in Quality-of-Life Theory and Research, ed. M. Joseph Sirgy, Don R. Rahtz and A. Coskun Samli, 13-26. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ferrer-i-Carbonell, Ada, and Frijters, Paul. 2004. "How Important Is Methodology for the Estimates of the Determinants of Happiness". Economic Journal, 114: 641-659.

Frey, Bruno S., and Simon Luechinger. 2007. "Concepts of Happiness and their Measurement." In Jahrbuch Ökologische Ökonomik, ed. Jürgen Meyerhoff and Reimund Schwarze, vol. 5, 129-237. Soziale Nachhaltigkeit. Marburg, DE: Metropolis Verlag.

Frey, Bruno S., and Alois Stutzer. 2002a. Happiness and Economics; How the Economy and Institutions Affect Well-Being. Princeton, NJ.

Frey, Bruno S., and Alois Stutzer. 2002b. "What Can Economists Learn from Happiness Research?" Journal of Economic Literature, 40: 402-435.

Holland, Paul W., and Howard Wainer. 1993. Differential Item Functioning. Hillsdale, NJ: Erlbaum.

Inglehart, Ronald, and Hans-Dieter Klingemann. 2000. "Genes, Culture, Democracy and Happiness." In Subjective Well-Being across Cultures, ed. Ed Diener and Eunkook M. Suh, 165–183. Cambridge, MA: MIT Press.

Inglehart, Ronald, and Jacques-René Rabier. 1986. "Aspirations Adapt to Situations – But Why are the Belgians So Much Happier than the French?" In Research on the Quality of Life, ed. F.M. Andrews, 1-56. Ann Arbor, MI: Inter-University Consortium for Political and Social Research.

Jürges, Hendrik. 2007. "True Health vs Response Styles: Exploring Cross-Country Differences in Self-Reported Health." Health Economics, 16: 163-178.

Kahneman, Daniel, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A.

Stone. 2004a. “Towards National Well-Being Accounts.” American Economic Review, Papers and Proceedings, 94: 429 - 434.

Kahneman, Daniel, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone. 2004b. “A Survey Method for Characterizing Daily Life Experience: the Day Reconstruction Method.” Science, 306: 1776-1780.

Kapteyn, Arie, James P. Smith, and Arthur van Soest. 2007. “Vignettes and Self-Reports of Work Disability in the U.S. and the Netherlands.” American Economic Review, 97: 461-473.

Kapteyn, Arie, James P. Smith, and Arthur van Soest. 2010. "Life Satisfaction". IZA Discussion Paper 4015, Institute for the Study of Labor, Bonn.

King, Gary, Christopher J.L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. “Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research.” American Political Science Review, 98: 567-583.

Kristensen, Nicolai, and Edvard Johansson. 2008. “New Evidence on Cross-Country Differences in Job Satisfaction Using Anchoring Vignettes". Labour Economics, 15: 96-117.

Murray, Christopher J.L., Emre Özaltin, Ajay Tandon, and Joshua A. Salomon. 2003. "Empirical Evaluation of the Anchoring Vignettes Approach in Health Surveys". In Health Systems Performance Assessment: Debates, Methods and Empiricism, ed. Christopher J.L. Murray and David B. Evans, 369-399. Geneva: World Health Organization.

Rice, Nigel, Silvana Robone and Peter Smith. 2010. "Analysis of the Validity of the Vignette Approach to Correct for Heterogeneity in Reporting Health System Responsiveness". European Journal of Health Economics, forthcoming.

Salomon, Joshua A., Ajay Tandon, and Christopher J.L. Murray. 2004. “Comparability of Self-Rated Health: Cross Sectional Multi-Country Survey Using Anchoring Vignettes.” British Medical Journal, 328: 258-261.

Senik, Claudia. 2004. “When Information Dominates Comparison: Learning from Russian Subjective Panel Data.” Journal of Public Economics, 88: 2099-2123.

Uchida, Yukiko, Vinai Norasakkunit, and Shinobu Kitayama. 2004. “Cultural Constructions of Happiness: Theory and Empirical Evidence.” Journal of Happiness Studies, 5: 223-239.

van Beek, Krijn W.H., Carl C. Koopmans, and Bernard M.S. van Praag. 1997. “Shopping at the Labour Market: a Real Tale of Fiction.” European Economic Review, 41: 295-317.

van Praag, Bernard M.S. 1971. "The Welfare Function of Income in Belgium: an Empirical Investigation." European Economic Review, 2: 337-369.

van Praag, Bernard M.S., and Ada Ferrer-i-Carbonell. 2004. Happines Quantified: a Satisfaction Calculus Approach. Oxford: Oxford University Press.

Van Soest, Arthur, Arie Kapteyn, Tatiana Andreyeva, and James P. Smith. 2006. "Self-Reported Disability and Reference Groups." Wr-409, RAND Labour and Population, Working Paper Series.

Van Soest Arthur, Delaney Liam, Colm Harmon, Arie Kapteyn and James P. Smith. 2007. "Validating the Use of Vignettes for Subjective Threshold Scales." IZA Discussion Paper 2860, Institute for the Study of Labor, Bonn.

Winkelmann, Liliana, and Rainer Winkelmann. 1998. "Why Are the Unemployed so Unhappy? Evidence from Panel Data." Economica, 65: 1–15.

Yang, Yang. 2008. "Social Inequalities in Happiness in the United States, 1972 to 2004: An Age-Period-Cohort Analysis." American Sociological Review, 73: 204-226.

A Appendix

Figure 1: Life satisfaction self-assessments.

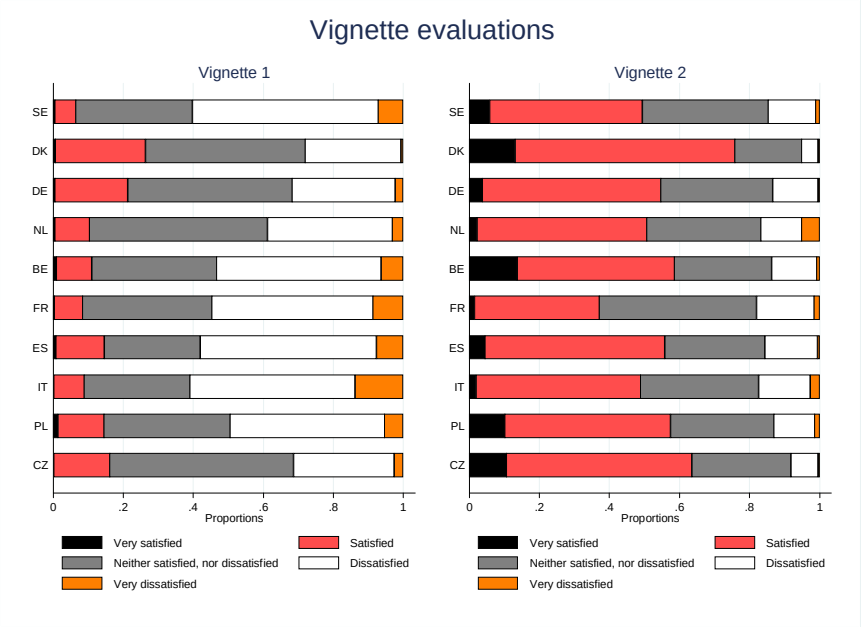


Figure 2: Vignette evaluations by country.

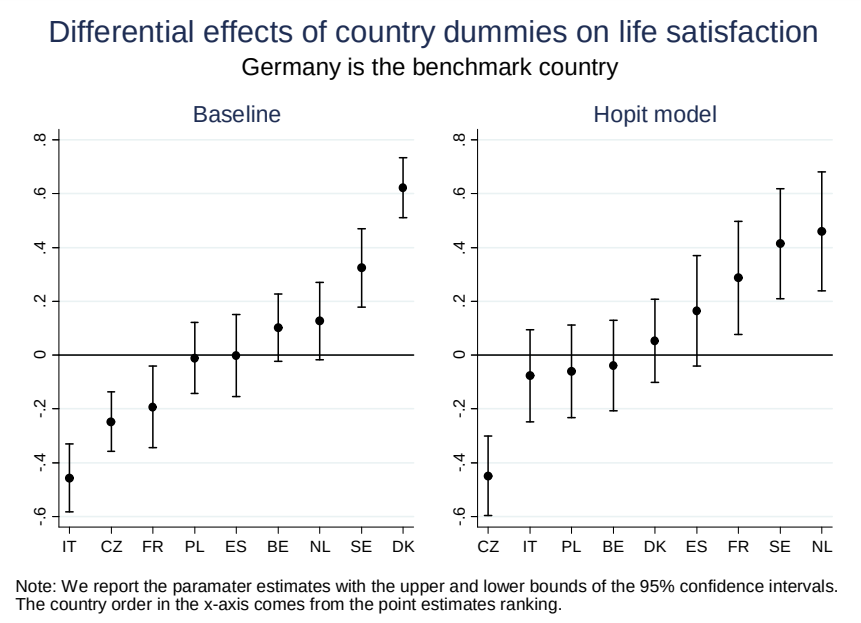


Figure 3: Country differences in life satisfaction.

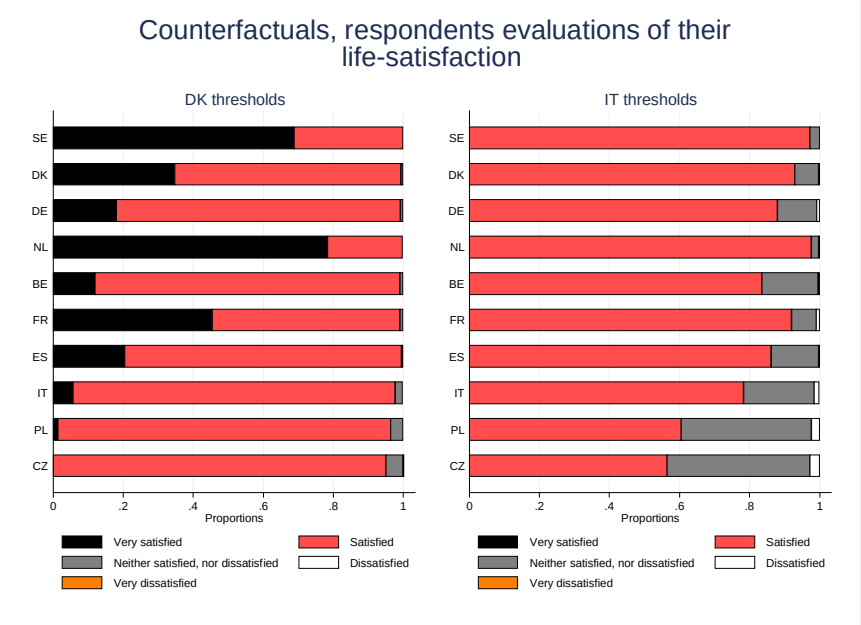
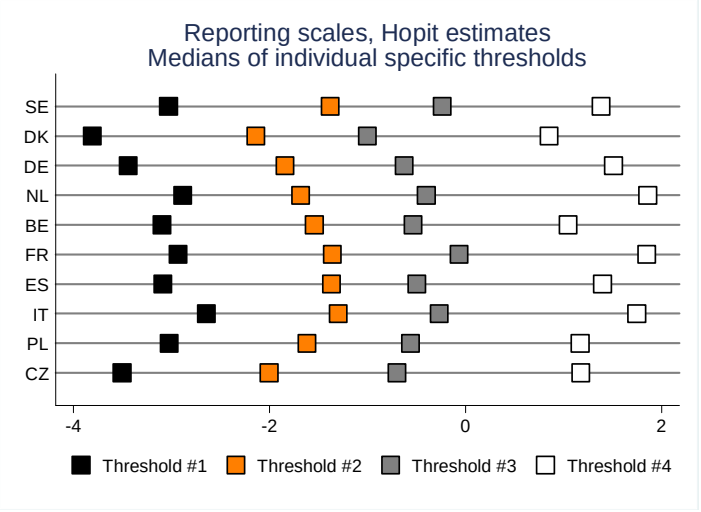


Figure 4: Counterfactual simulation: Danish and Italian thresholds.



Country	Medians of the thresholds			
	τ_1	τ_2	τ_3	τ_4
SE	-3.03	-1.38	-0.24	1.38
DK	-3.81	-2.14	-1	0.85
DE	-3.44	-1.84	-0.63	1.51
NL	-2.89	-1.69	-0.4	1.86
BE	-3.1	-1.54	-0.54	1.04
FR	-2.93	-1.36	-0.07	1.85
ES	-3.09	-1.37	-0.5	1.4
IT	-2.65	-1.3	-0.27	1.75
PL	-3.02	-1.62	-0.56	1.17
CZ	-3.51	-2.01	-0.7	1.17

Figure 5: Medians of individual-specific thresholds by country.

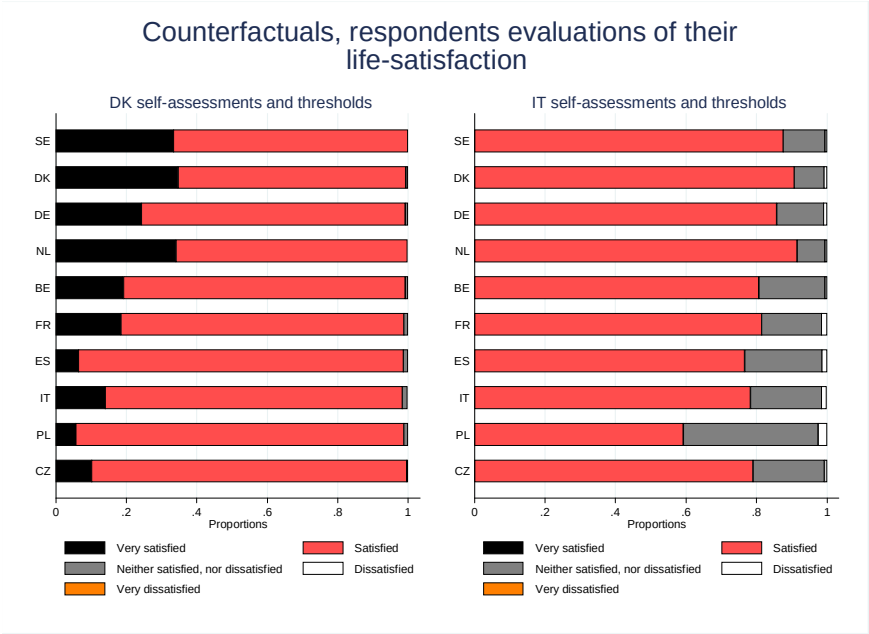


Figure 6: Counterfactual simulation: Danish and Italian thresholds and self-assessments.

Table 1: Description of the variables included in the regressions.

Variable	Description	Mean	Std. Dev.
Country of residence			
SE	dummy=1 if the person lives in Sweden	0.064	0.245
DK	dummy=1 if the person lives in Denmark	0.149	0.356
NL	dummy=1 if the person lives in The Netherlands	0.062	0.241
BE	dummy=1 if the person lives in Belgium	0.092	0.290
FR	dummy=1 if the person lives in France	0.056	0.229
IT	dummy=1 if the person lives in Italy	0.113	0.317
ES	dummy=1 if the person lives in Spain	0.064	0.244
PL	dummy=1 if the person lives in Poland	0.087	0.281
CZ	dummy=1 if the person lives in Czech Republic	0.148	0.355
DE	dummy=1 if the person lives in Germany (baseline)		
Demographics			
male	dummy=1 if the person is male	0.447	0.497
age	$(\text{age} - 65)/100$	-0.007	0.097
age2	$[(\text{age} - 65)/100]^2$	0.010	0.012
Education			
low_edu	dummy=1 if the ISCED code is at most 1	0.161	0.368
med_edu	dummy=1 if the ISCED code is either 2 or 3	0.602	0.490
high_edu	dummy=1 if the ISCED code is at least 4 (baseline)		
Employment			
retired	dummy=1 if retired from work	0.528	0.499
not_at_work	dummy=1 if homemaker, unemployed or disabled	0.173	0.378
employed	dummy=1 if employee or self-employed (baseline)		
Civil status			
spouse	dummy=1 if the person has a cohabiting partner	0.777	0.416
widowed	dummy=1 if widow or widower	0.125	0.330
divorced	dummy=1 if the person is divorced	0.065	0.247
single	dummy=1 if the person has never been married (baseline)		
Family background			
child	dummy=1 if the person has at least one living child	0.911	0.285
child_weekly	dummy=1 if the person has almost weekly contacts with her child(ren)	0.881	0.324
parent	dummy=1 if the person has at least one parent alive	0.265	0.441
parent_weekly	dummy=1 if the person has almost weekly contacts with her parent(s)	0.229	0.420
gchild	dummy=1 if the person has at least one grandchild	0.640	0.480
gchild_weekly	dummy=1 the person has almost weekly contacts with her grandchild(ren)	0.161	0.367

(See the next page)

Variable	Description	Mean	Std. Dev.
Health status			
obese	dummy=1 if the person is obese	0.192	0.394
chronic	dummy=1 if the person has at least two chronic diseases	0.447	0.497
arthritis	dummy=1 if the person is affected by arthritis	0.222	0.415
symptoms	dummy=1 if the person has at least two symptoms of diseases	0.439	0.496
mobility	dummy=1 if the person has mobility limitations	0.312	0.463
adl	dummy=1 if at least one limitation with activities of daily living	0.094	0.292
iadl	dummy=1 if at least one limitation with instrumental adl	0.135	0.341
depression	dummy=1 if diagnosed with affective or emotional disorders	0.085	0.279
Social activities			
volunteer	dummy=1 if done voluntary or charity work	0.138	0.345
caregiving	dummy=1 if cared for a sick person or provided informal help	0.227	0.419
training	dummy=1 if attended an educational course or gone to a social club	0.263	0.441
community	dummy=1 if taken part in political or religious activities	0.123	0.329
no_activities	dummy=1 if the person has not done any social activity (baseline)		
Politics			
left_wing	dummy=1 if the person is left-wing in politics	0.181	0.385
right_wing	dummy=1 if the person is right-wing in politics	0.252	0.434
center	dummy=1 if the person is centrist (baseline)		
Religion			
rel_more_day	dummy=1 if the person prays more than once a day	0.086	0.281
rel_day	dummy=1 if the person prays once a day	0.186	0.389
rel_week	dummy=1 if the person prays at least once a week	0.152	0.359
rel_less_week	dummy=1 if the person prays less than once a week	0.163	0.369
rel_nev	dummy=1 if the person never prays (baseline)		
Financial status			
income	arcsinh(household income)*, where income is PPP-adjusted and in euros	10.092	1.892
wealth	arcsinh(household wealth)*, where wealth is PPP-adjusted and in euros	11.647	3.940

* $\text{arcsinh}(x) = \ln\left(x + \sqrt{x^2+1}\right)$. This transformation allows us to take into account observations with zeros.

Table 2: Hopit model, determinants of life satisfaction. The first column refers to a baseline Hopit specification not allowing for threshold variation across individuals

	Baseline	Self assessments	Threshold equations γ_1	γ_2	γ_3	γ_4
SE	0.324***	0.414***	0.474***	-0.018	-0.047	-0.291***
	(0.074)	(0.104)	(0.125)	(0.077)	(0.060)	(0.053)
DK	0.622***	0.053	-0.285**	-0.009	-0.086*	-0.177***
	(0.057)	(0.079)	(0.130)	(0.078)	(0.047)	(0.036)
NL	0.127*	0.460***	0.694***	-0.356***	0.066	0.011
	(0.074)	(0.112)	(0.120)	(0.085)	(0.057)	(0.050)
BE	0.102	-0.038	0.341***	-0.007	-0.207***	-0.283***
	(0.064)	(0.086)	(0.111)	(0.069)	(0.054)	(0.043)
FR	-0.192**	0.287***	0.544***	-0.024	0.065	-0.069
	(0.077)	(0.107)	(0.126)	(0.080)	(0.059)	(0.057)
IT	-0.456***	-0.076	0.783***	-0.122*	-0.196***	-0.001
	(0.065)	(0.087)	(0.104)	(0.069)	(0.052)	(0.046)
ES	-0.002	0.165	0.347***	0.122	-0.333***	-0.055
	(0.077)	(0.105)	(0.133)	(0.081)	(0.068)	(0.052)
PL	-0.011	-0.06	0.379***	-0.114	-0.126**	-0.163***
	(0.067)	(0.088)	(0.114)	(0.074)	(0.054)	(0.045)
CZ	-0.247***	-0.448***	-0.088	-0.051	0.081*	-0.119***
	(0.056)	(0.075)	(0.113)	(0.072)	(0.043)	(0.037)
male	-0.159***	-0.133***	-0.018	0.05	-0.025	0.001
	(0.033)	(0.044)	(0.056)	(0.036)	(0.026)	(0.022)
age	1.175***	1.787***	0.773*	-0.558*	0.723***	-0.009
	(0.275)	(0.363)	(0.469)	(0.302)	(0.224)	(0.186)
age2	-2.619*	-4.463**	-5.686**	2.385	-0.895	1.716*
	(1.516)	(1.996)	(2.686)	(1.659)	(1.206)	(1.012)
low_edu	-0.074	-0.177**	-0.039	-0.041	0.008	-0.035
	(0.060)	(0.081)	(0.094)	(0.062)	(0.049)	(0.042)
med_edu	-0.041	-0.02	-0.062	-0.003	0.060*	0.039
	(0.040)	(0.054)	(0.069)	(0.044)	(0.033)	(0.026)
retired	-0.056	-0.121*	0.124	-0.031	-0.094**	-0.046
	(0.051)	(0.069)	(0.088)	(0.055)	(0.041)	(0.034)
not_at_work	-0.258***	-0.399***	0.05	-0.028	-0.080*	-0.082**
	(0.052)	(0.069)	(0.089)	(0.057)	(0.042)	(0.035)
spouse	0.310***	0.369***	0.396**	-0.188*	-0.062	0.005
	(0.095)	(0.124)	(0.164)	(0.099)	(0.072)	(0.064)
widowed	0.119	0.128	0.345*	-0.132	-0.092	-0.067
	(0.104)	(0.135)	(0.178)	(0.109)	(0.079)	(0.071)
divorced	-0.028	0.001	0.307	-0.144	-0.066	0.011
	(0.109)	(0.141)	(0.188)	(0.114)	(0.083)	(0.074)

(See the next page)

	Baseline	Self assessments	Threshold equations			
			γ_1	γ_2	γ_3	γ_4
child	-0.214**	-0.409***	0.081	-0.112	-0.105	-0.04
	(0.107)	(0.138)	(0.160)	(0.121)	(0.083)	(0.072)
child_weekly	0.231***	0.399***	-0.257*	0.254**	0.059	0.029
	(0.089)	(0.113)	(0.136)	(0.107)	(0.070)	(0.061)
parent	-0.081	-0.173	0.106	-0.06	0.06	-0.224***
	(0.084)	(0.108)	(0.138)	(0.091)	(0.066)	(0.059)
parent_weekly	0.135	0.279**	-0.025	0.014	-0.016	0.211***
	(0.086)	(0.111)	(0.140)	(0.091)	(0.067)	(0.060)
gchild	0.027	0.066	-0.003	0.031	-0.019	0.021
	(0.042)	(0.056)	(0.068)	(0.043)	(0.033)	(0.027)
gchild_weekly	0.098**	0.103*	-0.093	0.008	0.065*	0.03
	(0.046)	(0.061)	(0.076)	(0.049)	(0.037)	(0.030)
obese	0.059	0.025	-0.053	-0.008	0.019	0.015
	(0.039)	(0.052)	(0.066)	(0.043)	(0.031)	(0.026)
chronic	-0.112***	-0.072	0.145**	-0.104***	0.038	0.017
	(0.037)	(0.049)	(0.060)	(0.039)	(0.029)	(0.024)
arthritis	-0.071*	-0.065	-0.122*	0.031	0.098***	-0.03
	(0.041)	(0.053)	(0.067)	(0.044)	(0.032)	(0.028)
symptoms	-0.349***	-0.348***	-0.056	0.063	0	-0.038
	(0.038)	(0.050)	(0.063)	(0.040)	(0.030)	(0.025)
mobility	-0.144***	-0.128**	0.098	0.001	-0.064*	-0.027
	(0.042)	(0.055)	(0.069)	(0.046)	(0.034)	(0.029)
adl	-0.394***	-0.304***	0.128	-0.013	-0.068	0.01
	(0.060)	(0.076)	(0.090)	(0.061)	(0.047)	(0.046)
iadl	-0.158***	-0.174**	-0.033	0.001	0.035	-0.052
	(0.054)	(0.068)	(0.082)	(0.056)	(0.041)	(0.039)
depression	-0.321***	-0.317***	0.275***	-0.158***	-0.043	-0.069*
	(0.056)	(0.070)	(0.078)	(0.056)	(0.043)	(0.040)
volunteer	0.151***	0.274***	-0.199**	0.114**	0.066*	0.044
	(0.049)	(0.070)	(0.089)	(0.052)	(0.039)	(0.032)
caregiving	0.104***	0.133**	-0.074	0.078*	-0.034	0.02
	(0.039)	(0.053)	(0.068)	(0.042)	(0.031)	(0.025)
training	0.175***	0.262***	-0.095	0.076*	-0.011	0.059**
	(0.038)	(0.053)	(0.069)	(0.042)	(0.031)	(0.025)
community	0.148***	0.079	-0.046	0.006	-0.052	0.019
	(0.049)	(0.067)	(0.085)	(0.053)	(0.040)	(0.032)
income	0.017**	0.016	0.001	0	-0.004	-0.002
	(0.009)	(0.011)	(0.012)	(0.008)	(0.007)	(0.006)
wealth	0.017***	0.022***	-0.004	-0.001	0.008**	0.006**
	(0.004)	(0.005)	(0.006)	(0.004)	(0.003)	(0.003)
constant			-3.610***	0.480***	0.271**	0.703***
			(0.286)	(0.150)	(0.111)	(0.097)
Log-likelihood	-18855.232			-18303.643		
Sample size	5606			5606		

Table 3: Hopit model, estimates of the vignette equation parameters.

	Baseline	Hopit
θ_1 (John)	-1.546***	-1.540***
	(0.144)	(0.186)
θ_2 (Carry)	-0.389***	-.0373**
	(0.143)	(0.186)
σ_v^2	1.167	1.102

Note: Standard errors in parentheses, *** p-value<0.01, ** p-value<0.05, * p-value<0.1.

Table 4: Pairwise comparisons of country effects in the two models

Baseline

DK	1.078***	0.869***	0.814***	0.633***	0.624***	0.622***	0.520***	0.495***	0.297***	-
SE	0.781***	0.572***	0.517***	0.335***	0.326***	0.324***	0.222***	0.198**	-	
NL	0.583***	0.374***	0.319***	0.137	0.129	0.127*	0.024	-		
BE	0.559***	0.350***	0.295***	0.113	0.104	0.102	-			
DE	0.456***	0.247***	0.192**	0.011	0.002	-				
ES	0.455***	0.245***	0.190**	0.009	-					
PL	0.446***	0.237***	0.181**	-						
FR	0.264***	0.055	-							
CZ	0.209***	-								
IT	-									
	IT	CZ	FR	PL	ES	DE	BE	NL	SE	DK

Hopit

NL	0.908***	0.536***	0.520***	0.498***	0.460***	0.407***	0.295**	0.173	0.046	-
SE	0.862***	0.490***	0.474***	0.452***	0.414***	0.361***	0.249**	0.127	-	
FR	0.736***	0.364***	0.347***	0.325***	0.287***	0.234**	0.122	-		
ES	0.613***	0.241**	0.225**	0.203*	0.165	0.112	-			
DK	0.501***	0.130	0.113	0.091	0.053	-				
DE	0.448***	0.076	0.06	0.038	-					
BE	0.410***	0.038	0.022	-						
PL	0.388***	0.016	-							
IT	0.372***	-								
CZ	-									
	CZ	IT	PL	BE	DE	DK	ES	FR	SE	NL

Note: Each cell reports the value of the difference between the estimated coefficient on the dummy of the country on the vertical axis and that of the country on the horizontal axis. The country order on both axes comes from the point estimates ranking. Significance levels: *** p-value<0.01, ** p-value<0.05, * p-value<0.1

Table 5: Percentage of respondents who ordered vignettes consistently with the global ordering.

	Percentage
By country	
Sweden	95.26%
Denmark	95.19%
Germany	88.93%
The Netherlands	88.70%
Belgium	92.23%
France	92.28%
Spain	89.17%
Italy	93.85%
Poland	92.93%
Czech Republic	93.15%
By group of countries	
Protestant Europe	91.94%
Catholic Europe	92.21%
Ex-communist countries	93.07%
All sample	92.29%

Table 6: Estimates of the Hopit model on groups of countries: Protestant Europe, Catholic Europe and Ex-communist countries.

	Protestant Europe		Catholic Europe		Ex-communist countries	
Parameters on the country dummies in the self-assessment equation						
	SE	0.331***	BE	-0.070	PL	-0.088
		(0.113)		(0.088)		(0.099)
	DK	0.024	FR	0.218**	CZ	-0.484***
		(0.081)		(0.111)		(0.082)
	NL	0.479***	IT	-0.108		
		(0.115)		(0.092)		
			ES	0.120		
				(0.110)		
θ parameters						
θ_1 (John)		-0.585*		-1.344***		-1.570***
		(0.334)		(0.245)		(0.323)
θ_2 (Carry)		0.439		-0.242		-0.452
		(0.334)		(0.244)		(0.322)
σ_v^2		0.978		1.084		1.213
Log-likelihood		-7699.709		-9063.494		-7202.962
Sample size		2472		2749		2243

Note: Standard errors in parentheses, *** p-value<0.01, ** p-value<0.05, * p-value<0.1. In all specifications Germany is the baseline country as in the pooled estimation reported in Table 2.

B Appendix not intended for publication

Hopit model, determinants of life satisfaction when accounting for political and religious background.

	Baseline	Self assessments	Threshold equations γ_1	γ_2	γ_3	γ_4
SE	0.342***	0.412***	0.502***	-0.022	-0.080	-0.290***
	(0.077)	(0.109)	(0.136)	(0.083)	(0.063)	(0.055)
DK	0.612***	0.050	-0.241*	-0.017	-0.102**	-0.174***
	(0.060)	(0.083)	(0.139)	(0.084)	(0.049)	(0.038)
NL	0.106	0.462***	0.745***	-0.378***	0.046	0.023
	(0.077)	(0.119)	(0.128)	(0.090)	(0.060)	(0.052)
BE	0.101	-0.073	0.342***	0.004	-0.238***	-0.288***
	(0.068)	(0.091)	(0.122)	(0.075)	(0.057)	(0.046)
IT	-0.502***	-0.193**	0.859***	-0.154*	-0.231***	-0.053
	(0.072)	(0.098)	(0.121)	(0.080)	(0.058)	(0.051)
ES	-0.021	0.089	0.410***	0.110	-0.415***	-0.081
	(0.086)	(0.117)	(0.155)	(0.094)	(0.077)	(0.058)
PL	-0.093	-0.194*	0.526***	-0.170*	-0.183***	-0.219***
	(0.075)	(0.099)	(0.131)	(0.088)	(0.061)	(0.051)
CZ	-0.244***	-0.463***	-0.096	-0.050	0.080*	-0.124***
	(0.059)	(0.079)	(0.123)	(0.078)	(0.045)	(0.039)
male	-0.137***	-0.093*	-0.065	0.077*	-0.024	0.021
	(0.036)	(0.049)	(0.065)	(0.042)	(0.029)	(0.024)
age	1.084***	1.548***	0.754	-0.457	0.591**	-0.211
	(0.302)	(0.399)	(0.527)	(0.339)	(0.242)	(0.202)
age2	-2.772*	-3.015	-4.681	1.762	0.431	1.653
	(1.676)	(2.228)	(2.983)	(1.865)	(1.317)	(1.101)
low_edu	-0.126*	-0.163*	-0.077	-0.003	0.040	-0.009
	(0.068)	(0.091)	(0.110)	(0.072)	(0.056)	(0.048)
med_edu	-0.063	-0.023	-0.054	0.000	0.043	0.059**
	(0.042)	(0.058)	(0.077)	(0.049)	(0.035)	(0.027)
retired	-0.057	-0.122*	0.146	-0.058	-0.069	-0.051
	(0.055)	(0.074)	(0.097)	(0.061)	(0.044)	(0.036)
not_at_work	-0.273***	-0.414***	0.110	-0.063	-0.062	-0.102***
	(0.057)	(0.075)	(0.099)	(0.064)	(0.046)	(0.038)
spouse	0.326***	0.406***	0.408**	-0.194*	-0.030	-0.008
	(0.105)	(0.136)	(0.196)	(0.117)	(0.081)	(0.069)

(See the next page)

	Baseline	Self assessments	Threshold equations γ_1	γ_2	γ_3	γ_4
widowed	0.174	0.214	0.360*	-0.156	-0.048	-0.053
	(0.116)	(0.150)	(0.214)	(0.131)	(0.090)	(0.077)
divorced	0.056	0.100	0.408*	-0.222*	-0.026	-0.009
	(0.120)	(0.155)	(0.218)	(0.133)	(0.093)	(0.079)
child	-0.237**	-0.437***	0.119	-0.121	-0.130	-0.043
	(0.117)	(0.151)	(0.182)	(0.137)	(0.092)	(0.077)
child_weekly	0.215**	0.392***	-0.275*	0.237*	0.081	0.065
	(0.098)	(0.123)	(0.156)	(0.122)	(0.077)	(0.065)
parent	-0.035	-0.125	0.074	-0.042	0.065	-0.203***
	(0.092)	(0.118)	(0.154)	(0.102)	(0.072)	(0.064)
parent_weekly	0.099	0.229*	-0.054	0.024	-0.005	0.181***
	(0.094)	(0.121)	(0.153)	(0.099)	(0.073)	(0.065)
gchild	0.032	0.074	-0.059	0.064	-0.011	0.015
	(0.045)	(0.061)	(0.075)	(0.048)	(0.036)	(0.029)
gchild_weekly	0.088*	0.152**	-0.074	0.001	0.079**	0.067**
	(0.049)	(0.067)	(0.087)	(0.056)	(0.039)	(0.032)
obese	0.052	0.014	-0.010	-0.040	0.026	0.006
	(0.043)	(0.057)	(0.075)	(0.050)	(0.034)	(0.029)
chronic	-0.122***	-0.074	0.107	-0.074*	0.036	0.019
	(0.040)	(0.053)	(0.067)	(0.043)	(0.031)	(0.026)
arthritis	-0.089*	-0.079	-0.138*	0.037	0.110***	-0.030
	(0.046)	(0.059)	(0.076)	(0.050)	(0.035)	(0.031)
symptoms	-0.346***	-0.372***	-0.009	0.036	-0.018	-0.053**
	(0.041)	(0.054)	(0.071)	(0.045)	(0.032)	(0.027)
mobility	-0.118**	-0.102*	0.079	0.001	-0.058	-0.011
	(0.046)	(0.060)	(0.076)	(0.050)	(0.036)	(0.031)
adl	-0.374***	-0.233***	0.178*	0.000	-0.100*	0.026
	(0.068)	(0.086)	(0.106)	(0.073)	(0.053)	(0.051)
iadl	-0.151**	-0.183**	-0.069	-0.009	0.075	-0.042
	(0.060)	(0.076)	(0.098)	(0.068)	(0.046)	(0.043)
depression	-0.274***	-0.294***	0.220**	-0.132*	-0.044	-0.077*
	(0.065)	(0.081)	(0.096)	(0.068)	(0.051)	(0.045)
volunteer	0.147***	0.245***	-0.215**	0.103*	0.073*	0.041
	(0.053)	(0.077)	(0.103)	(0.061)	(0.043)	(0.035)
caregiving	0.094**	0.122**	-0.085	0.086*	-0.047	0.025
	(0.042)	(0.057)	(0.078)	(0.048)	(0.034)	(0.027)

(See the next page)

	Baseline	Self assessments	Threshold equations γ_1	γ_2	γ_3	γ_4
training	0.184***	0.236***	-0.081	0.062	-0.025	0.048*
	(0.041)	(0.057)	(0.075)	(0.047)	(0.033)	(0.026)
community	0.104*	0.029	-0.002	0.006	-0.080*	-0.007
	(0.055)	(0.076)	(0.101)	(0.064)	(0.046)	(0.035)
income	0.013	0.014	0.012	-0.005	-0.006	0.002
	(0.009)	(0.012)	(0.014)	(0.009)	(0.007)	(0.006)
wealth	0.018***	0.023***	-0.004	0.000	0.007*	0.005*
	(0.004)	(0.006)	(0.007)	(0.005)	(0.004)	(0.003)
left_wing	-0.130***	-0.137**	0.000	0.026	-0.053	0.004
	(0.045)	(0.059)	(0.075)	(0.048)	(0.036)	(0.030)
right_wing	0.064	0.114**	0.026	0.007	-0.012	0.025
	(0.040)	(0.055)	(0.067)	(0.044)	(0.032)	(0.026)
rel_more_day	0.138**	0.190**	-0.071	-0.011	0.007	0.117**
	(0.070)	(0.094)	(0.115)	(0.080)	(0.057)	(0.047)
rel_day	0.121**	0.153**	-0.195**	0.047	0.091**	0.059*
	(0.052)	(0.070)	(0.097)	(0.062)	(0.041)	(0.034)
rel_week	0.072	0.132*	-0.098	-0.003	0.063	0.091***
	(0.053)	(0.072)	(0.095)	(0.062)	(0.043)	(0.035)
rel_less_week	0.015	0.009	-0.029	-0.049	0.065*	0.020
	(0.049)	(0.066)	(0.090)	(0.059)	(0.039)	(0.032)
constant			-3.646***	0.548***	0.254**	0.611***
			(0.325)	(0.166)	(0.123)	(0.105)
Log-likelihood	-15970.158	-15495.803				
Sample size	4801	4801				

Note: Standard errors in parentheses, *** p-value<0.01, ** p-value<0.05, * p-value<0.1