UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Economiche "Marco Fanno"

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
University of Padova

January 2009

"MARCO FANNO" WORKING PAPER N.90

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¶

January, 2009

Abstract

Self-reported life satisfaction is highly heterogeneous across similar countries. We show that this phenomenon can be largely explained by the fact that individuals adopt different scales and benchmarks in evaluating themselves. Using a cross sectional dataset on individuals aged 50 and over in ten European countries, we compare estimates from an Ordered Probit in which life satisfaction scales are invariant across respondents with those from a Hopit model in which 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 dramatically depends on scale biases.

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

---

*We thank Arthur Van Soest, Rob Alessie, Karsten Hank, Guglielmo Weber, Nicolai Kristensen, Giovanni Porzio, Teresa Bago d'Uva, Luigino Bruni and Benedetto Gui for useful comments and suggestions. We are also grateful to participants at the seminar at the University of Groningen, the conference "Well being: Are We Happy with Our Standard of Living?" (Cassino) and the Third Nordic Workshop in Behavioral and Experimental Economics (Copenhagen) for helpful insights. Financial support from the European Commission through the STREP project COMPARE n. 028857 is gratefully acknowledged. This paper uses pre-release data from SHARE Wave 2. The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001-00360) and the 6th framework programme (projects SHARE-I3, RII-CT-2006-062193, and COMPARE, CIT5-CT-2005-028857). Additional funding came from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064, IAG BSR06-11, OGHA04-064).

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

Empirical studies on subjective well-being seek to identify the relation between life satisfaction (or other proxy of utility such as happiness) and economic, demographic, social and physical characteristics of individuals. In particular, in order to collect data on individuals' well-being, researchers use general surveys asking individuals to self-report their life satisfaction on a scale which goes from a minimum value to indicate very low satisfaction to a maximum value which expresses very high satisfaction.

There is no doubt about the importance of these studies. Starting from the famous Easterlin's paradox which 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 such as income, low inflation rate, being employed, there are non-economic variables which have a significant and positive effect on life satisfaction: high education, good health conditions and being physically autonomous, being female, being white rather than black, being local rather than immigrant, being involved in the community and volunteering, having frequent contacts with family, friends and neighbors, being married and having children, praying and attending religious communities.

A crucial methodological issue which limits cross-country comparability is that, when asked to self-report their life satisfaction, individuals, who are similar for both economic and non-economic conditions, can 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). 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"* .

In psychometrics this phenomenon is called *differential item functioning* (henceforth 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 biases 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 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).

On the other hand, scale biases 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).

When panel 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, these approaches present at least two main problems when used to analyze self-reported life satisfaction. First, they do not allow to control for DIF bias with cross sectional data in which respondents are not followed over time. Second, 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 individual and cross-country differences in self-reported life satisfaction are genuine or they just reflect individual-specific scale biases. Following this approach, individuals are

presented with two categories of questions on life satisfaction. First, as in traditional questionnaires, individuals are asked to self-report their level of life satisfaction. Second, each individual is asked to evaluate the life satisfaction of one or more hypothetical persons described in particular situations or conditions and kept constant across respondents (*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 filter out the level of self-reported life satisfaction from the DIF bias measured by vignettes and enhance the comparability of subjective assessments across individuals.

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 there is no 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, once controlled 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 biases, the ranking across countries dramatically 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 percentage of satisfied individuals in a country change when the re-

sponse scale of other countries is imposed to its inhabitants. When the Danish scale is used, more than 95 percent of respondents in all countries would rate themselves as satisfied with their own life. The picture considerably changes when using the Italian response scale. Indeed, the proportion of satisfied individuals according to the Italian scale drops significantly in all countries. Thus, variations in response scales explain a large part of the differences found in raw data. Concerning the determinants of individuals' well-being, once controlled 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. We also confirm the existence of a non linear relation between age and life satisfaction (see also Yang, 2008).

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) showing its properties and stating the statistical assumptions it is based on. 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, 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 percentage of satisfied individuals in a country change 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 6 we discuss our findings and we conclude.

2 Data and Descriptive Statistics

Our sample is drawn from the 2006 wave of the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE is a unique and multidisciplinary dataset that contains a large amount

of information on both the economic and non economic conditions of individuals aged 50 and over. As shown by Table 1, in our analysis we use a large number of individuals' characteristics that can be ideally classified into ten different categories: demographic conditions, education, employment, civil status, family backgrounds, health status, social activities, politics, religion and financial status.

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[1] 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?

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".

[1] The value is PPP-adjusted.

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 the proportion of respondents who rate themselves as either satisfied or very satisfied with their life. As documented in previous studies (Inglehart and Rabier, 1986; Kanhneman 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 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 Sweden, The Netherlands 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 (DIF) biases in evaluating themselves.

Analysing the distribution of vignettes evaluations at a pure descriptive level may help understanding how they can be used to purge self assessments from individual heterogeneity in reporting styles. Figure 2 presents the proportion of respondents that rate the persons described in the two vignettes (John and Carry) as satisfied with their life for each country. 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 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 satisfied with her life, the French are more reluctant to use the most extreme labelling of vignette evaluation. Remarkably, while nearly 80 percent of the Danes find Carry satisfied with her life, this probability falls to less than

40 percent for the French and lies between 50 and 60 percent for the remaining countries. 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.[2]

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 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 has been introduced by King et al. (2004) and it is usually referred to as the Hopit model. It mainly consists of two components

[2] See Bago d'Uva et al. (2008) for a discussion of a similar approach to assess the vignette equivalence hypothesis.

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^*, SHARE questionnaire picks 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.

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

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, SHARE questionnaire gathers respondent evaluations of vignette questions 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). Following King et al. (2004), the joint estimation is carried out via conditional maximum likelihood and implemented by the STATA module *gllamm*.[3]

4 Results

In our estimates we consider a large number of factors associated with life satisfaction: demographic characteristics (age and gender), socio-economic variables (employment, income and wealth and education), health (number of chronic diseases, arthritis, limitations with mobility, symptoms, ADL, IADL, obesity and having been diagnosed with depression) and social relationships (marital status, family bonds and extra-familiar activities), as well as country dummies.[4]

In Table 2 we present the results for the self-assessment equation of life satisfaction with heterogenous response scales (second column), comparing it with a baseline model not allowing for any threshold variation across respondents (first column).[5] 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 biases across countries or socio-economic groups, the parameter estimates of this model will reflect both true well-being 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

[3] For further details, see Rabe-Hesketh et al. (2004) and Rabe-Hesketh and Skrondal (2008).

[4] 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.

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

scale variation against the more general model that does allow for correction of the DIF bias.[6]

We first focus on the impact of different response scales on cross country variations in self-reported well-being. 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. Moreover, for both the Ordered Probit model and the Hopit model (Table 3) we report 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.

In the model without scale biases correction, 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 changes dramatically. 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.

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 their lives than unmarried ones, whose level of self-reported well-being is not significantly different from that of

[6] $\chi^2_{156} = 1103.177$, 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).

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

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 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 associated with lower levels of self-reported well-being: in particular, reporting at least two symptoms of diseases, limitations with mobility, ADL, IADL and having being diagnosed with affective or emotional disorders, all are significantly and negatively related to life satisfaction (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, substantially increases life satisfaction.

The overall 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 well-being of the elderly is the saving they have accumulated throughout their whole life and not current income. Even controlling for income and wealth, the less educated are the most dissatisfied with their life.

In Table 4 we estimate the same specifications as before but introducing in the model two additional sets of control variables related to the political and religious background of the respondent.[8]

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 (2004), who also show that people become happier the more often

[8] France is not included in the sample because in this country the questions about religion and politics were not asked for privacy reasons.

they attend church and the more often they pray. However, there seems to be no difference in self-reported 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.

Counterfactuals

We now present results of counterfactual simulations to assess the relevance of the DIF bias across countries. In particular, we simulate the number of persons in a country who would report to be satisfied with their life if they were asked to evaluate themselves according to the scale of response estimated for another 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 the proportions of individuals satisfied with their life if all the respondents used the Danish and Italian response scales respectively. The simulations show that, using the Danish thresholds, more than 95 percent of respondents in all countries would rate themselves as 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: in Poland the proportion of individuals who are satisfied with their life drops from about 97 percent to 60 percent and similar differences can be seen in all countries. A similar conclusion can be reached by looking at Figure 6. 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 they are more robust to the presence of outliers.[9] 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

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

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 to institutional differences or to differences in the composition of the sample with respect to the variables we control for in the estimation. In the second counterfactual exercise, we simulate the proportion of individuals that would rate themselves as satisfied with their life 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 do so by imposing the same country dummy to each individual in our sample. The results are presented in Figure 5 and, as in the previous exercise, the reference countries are Denmark and Italy. Since in both counterfactuals respondents use the same thresholds, the comparison between Figure 4 and Figure 5 highlights the institutional effects of living in Denmark and Italy rather than in the actual country of residence. 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, *ceteris paribus*. This evidence implies that the high proportion of satisfied individuals 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 for all countries but Czech Republic. While Figure 4 reports that 97% of the Swedes and 92% of the French would rate themselves as satisfied with their life if they adopted the Italian response scale, Figure 5 shows that these proportions would fall to 88% and 81% 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% of the Czechs are satisfied with their life, this proportion increases by more than 40% in the second counterfactual simulation. The remaining cross-country variability that we observe in Figure 5 is due only to differences in the composition of the sample across countries (e.g. if the proportion of females in country A is larger than that in

country B, the overall level of life satisfaction in country A will be higher, other things being equal).

5 Conclusion

In analyzing data of a Eurobarometer survey, Kahneman et al. (2004) notice that there is 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 authors, the puzzling differences in the level of self-reported life satisfaction across seemingly similar countries can be caused by the fact that the scales adopted by respondents to rate their satisfaction can be affected by individual-specific biases.

Using the 2006-2007 wave of SHARE, a cross-sectional dataset containing demographic, physical, social and economic information of individuals aged over 50 living in 10 European countries, we estimate and measure cross-country differences in individuals' scales of life satisfaction. In particular, we compare estimates from a model in which scales are assumed to be constant across country (Ordered Probit) with those from a model in which vignettes are used to correct for country specific scale biases (Hopit model).

Our results can be summarized as follows. According to the Ordered Probit that does not account for differences in reporting styles, once controlled 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 individual-specific biases, the ranking across countries dramatically change. 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. Interestingly, according to the Ordered Probit specification Danes are significantly more satisfied with their life than the French. On the contrary, an opposite and highly significant result is found when we estimate the Hopit model. This means that the assumption of invariant response scales across countries is far from being innocuous.

A formal likelihood ratio test strongly rejects the ordered probit not accounting for threshold variation across respondents against the more general Hopit model. We find evidence that the thresholds significantly depend not only on country dummies but also on age, education, employment status, health, social activities and wealth. We also present results from counterfactual simulations to study how the percentage of satisfied individuals in a country change when the response scale of other countries is imposed to its inhabitants. When the Danish scale is used, more than 95 percent of respondents in all countries would rate themselves as satisfied with their own life. The picture considerably changes when using the Italian response scale. Indeed, the proportion of satisfied individuals according to the Italian scale drops significantly in all countries. Thus, variations in response scales explain a large part of the differences found in raw data.

To our knowledge, this is the first study that applies the vignettes methodology to estimate and measure cross-country differences in the scale of self-reported life satisfaction. Our results inform the empirical literature studying the determinants of life satisfaction (or happiness) about the importance of correcting for individual-specific scale biases to produce more reliable cross-country comparisons.

As a proposal for further research, future waves of SHARE data will allow us to combine the vignettes approach with panel data techniques to control for time variant scales and to investigate to what extent individual reporting styles are stable over time.

References

Bago d'Uva, T., van Doorslaer, E., Lindeboom, M., O'Donnell, O. (2008). Does Reporting Heterogeneity Bias the Measurement of Health Disparities? Health Economics, 17, 351-375.

Bruni, L. and Porta, P.L. (2005). Economics and Happiness. Framing the Analysis. Oxford: Oxford University Press.

Charles, K.K. (2004). Is Retirement Depressing? Labor Force Inactivity and Psychological Well-Being in Later Life. In Polachek, S.W. (Ed-) "Accounting for Worker Well-Being" (pp.269-299),

Research in Labor Economics, vol. 23. Amsterdam, San Diego and Oxford: Elsevier, JAI.

Clark, A. (2003). Unemployment as a Social Norm: Psychological Evidence from Panel Data. Journal of Labor Economics, 21, 289-322.

Clark, A. and Lelkes, O. (2005) Deliver us from Evil: Religion as Insurance. PSE, Paris, November

Clark, A, Etilè, F., Postel-Vinay, F., Senik, C. and Van der Straeten, K. (2005). Heterogeneity in Reported Well-Being: Evidence from Twelve European Countries. Economic Journal, 115, C118-C132.

Clark, A.. and Oswald, A (1996). Satisfaction and Comparison Income. Journal of Public Economics, 61, 359-381.

Clark, A. and Oswald, A. (2002) Well-being in panels. DELTA, mimeo.

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

Diener, E. and Suh, E.M. (2000). Culture and Subjective Well-Being. Cambridge MA: MIT Press.

Easterlin, R. (2001) Income and happiness: towards a unified theory. Economic Journal, 111, 465-84.

Easterlin, R. (2003). Happiness of Women and Men in Later Life: Nature, Determinants, and Prospects. In Sirgy, M.J., Rahtz, D. and Samli, A.C. (Eds.) "Advances in Quality-of-Life Theory and Research" (pp.13-26), Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ferrer-i-Carbonell, A. and Frijters, P. (2004) How important is methodology for the estimates of the determinants of happiness. Economic Journal, 114, 641-659.

Frey, B.S. and Luechinger, S. (2007). Concepts of Happiness and their Measurement. In Meyerhoff, J. and Schwarze, R. (Eds.) "Jahrbuch Ökologische Ökonomik" (pp. 129-237), Vol. 5: Soziale Nachhaltigkeit. Marburg, DE: Metropolis Verlag.

Frey, B.S. and Stutzer, A. (2002a). Happiness and Economics; How the Economy and Institu-

tions Affect Well-Being. Princeton, NJ.

Frey, B.S. and Stutzer, A. (2002b). What Can Economists Learn from Happiness Research? Journal of Economic Literature, 40, 402-435.

Holland, P. W. and Wainer, H. (1993). Differential item functioning. Hillsdale, NJ: Erlbaum.

Inglehart, R., and Klingemann, H.D. (2000). Genes, culture, democracy and happiness. In Diener, E. and Suh, E.M. (Eds.) "Subjective well-being across cultures" (pp. 165–183). Cambridge, MA: MIT Press.

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

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

Kahneman, D., Krueger, A.B., Schkade, D., Schwarz, N. and Stone, A. (2004a). Towards National Well-Being Accounts. American Economic Review, Papers and Proceedings, 94, 429 - 434.

Kahneman, D., Krueger, A.B., Schkade, D., Schwarz, N. and Stone, A. (2004b). A Survey Method for Characterizing Daily Life Experience: the Day Reconstruction Method. Science, 306, 1776-1780.

Kapteyn, A., Smith, J.P. and Van Soest, A. (2007). Vignettes and Self-Reports of Work Disability in the U.S. and the Netherlands. American Economic Review, 97, 461-473.

King, G., Murray, C., Salomon, J. and Tandon, A. (2004). Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. American Political Science Review, 98, 567-583.

Kristensen, N. and Johansson, E. (2008). New Evidence on Cross-Country Differences in Job Satisfaction Using Anchoring Vignettes. Labour Economics, 15, 96-117.

Salomon, J.A., Tandon, A., Murray, C.J.L. (2004). Comparability of Self-Rated Health: Cross

Sectional Multi-Country Survey Using Anchoring Vignettes. British Medical Journal, 328, 258-261.

Senik, C. (2004). When Information Dominates Comparison: Learning from Russian Subjective Panel Data. Journal of Public Economics, 88, 2099-2123.

Uchida, Y., Norasakkunit, V. and Shinobu, K. (2004). Cultural Constructions of Happiness: Theory and Empirical Evidence. Journal of Happiness Studies, 5, 223-239.

van Beek, K.W.H., Koopmans, C.C. and van Praag, B.M.S. (1997). Shopping at the Labour Market: a Real Tale of Fiction. European Economic Review, 41, 295-317.

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

van Praag, B.M.S. and Ferrer-i-Carbonell, A. (2004). Happines Quantified: a Satisfaction Calculus Approach. Oxford: Oxford University Press.

Van Soest, A., Kapteyn, A., Andreyeva, T., Smith, J. (2006). Self-Reported Disability and Reference Groups. Wr-409, RAND Labour and Population, Working Paper Series.

Winkelmann, L. and R. Winkelmann (1998). Why Are the Unemployed so Unhappy? Evidence from Panel Data. Economica, 65, 1–15.

Yang, Y. (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

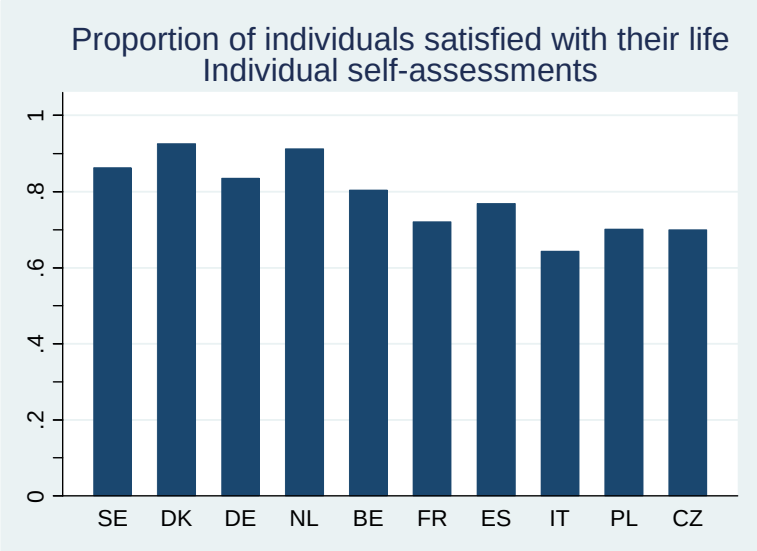


Figure 1: Life-satisfaction self-assessments.

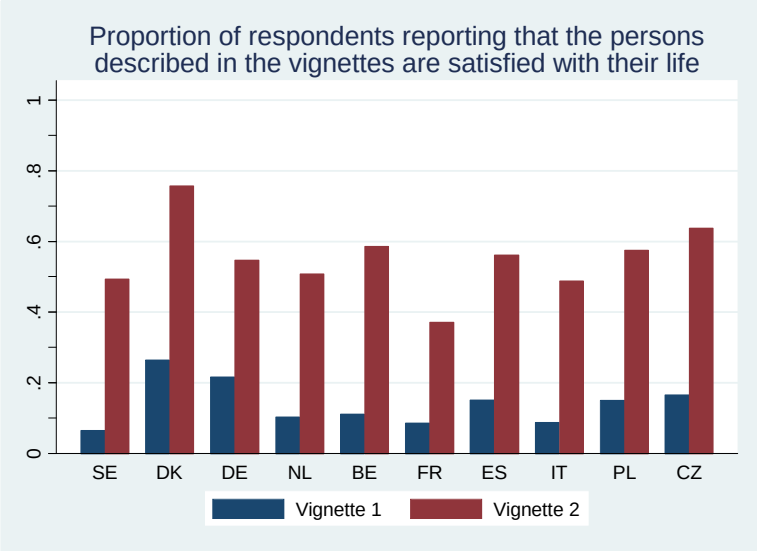


Figure 2: Vignette evaluations by country.

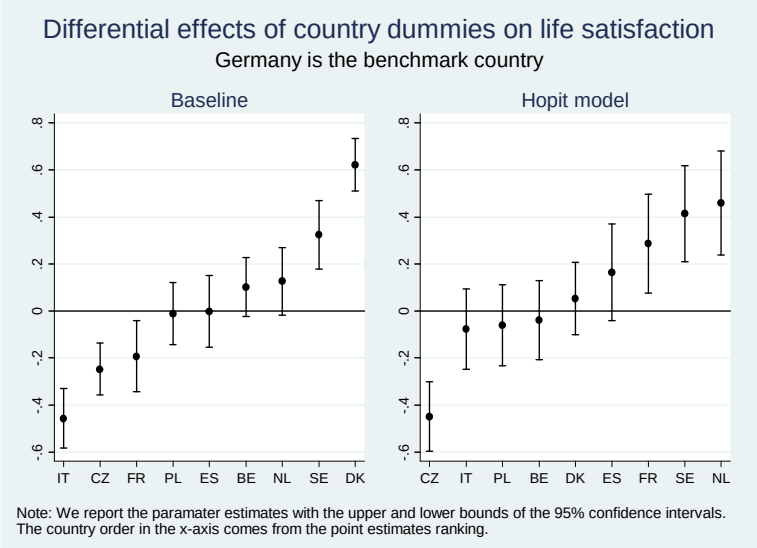


Figure 3: Country differences in life-satisfaction.

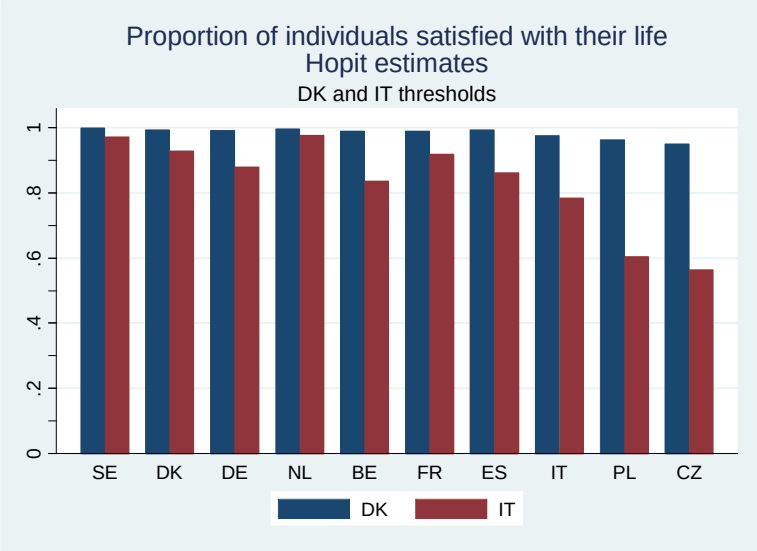


Figure 4: Counterfactual simulation: Danish and Italian thresholds.

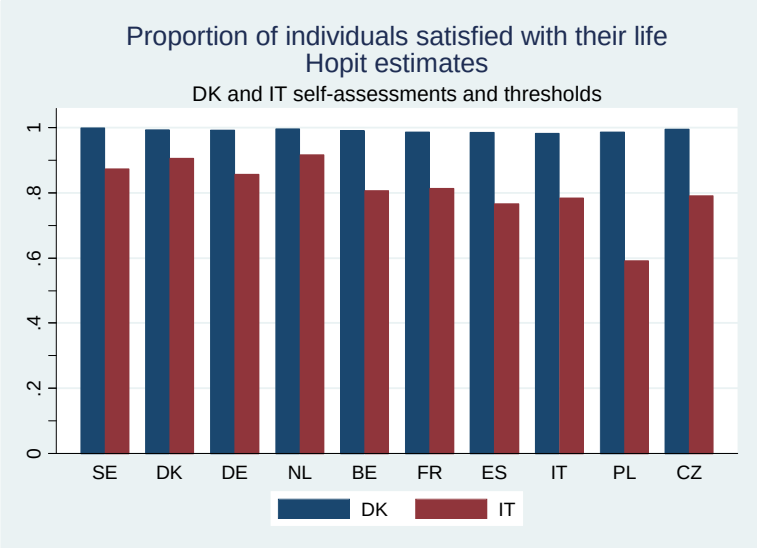
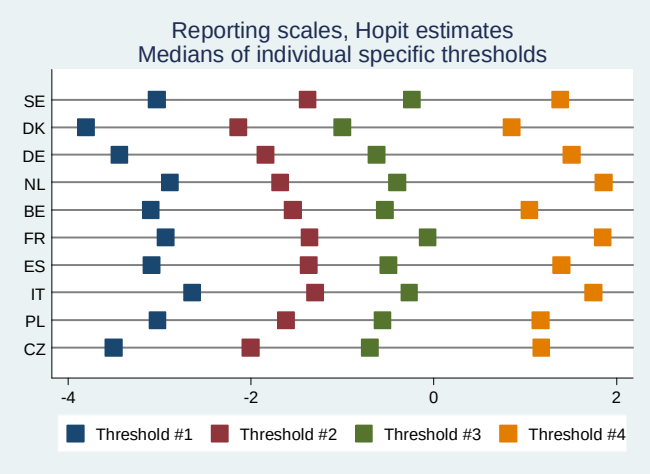


Figure 5: Counterfactual simulation: Danish and Italian thresholds and self-assessments.



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 6: Medians of individual-specific thresholds by country.

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)

* Henceforth, contacts include any kind of contact, either in person, by phone or mail.

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

** $\text{arcsinh}(x) = \ln\left(x + \sqrt{x^2+1}\right)$

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)
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)

(See the next page)

	Baseline	Self assessments	Threshold equations γ_1	γ_2	γ_3	γ_4
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		

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

Table 3: 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 of the dummy referring to the country on the vertical axis and that referring to 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 4: 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)
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)

(See the next page)

	Baseline	Self assessments	Threshold equations γ_1	γ_2	γ_3	γ_4
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)
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