J Econ Inequal (2015) 13:425–447
DOI 10.1007/s10888-015-9301-7

Alternative weighting structures for multidimensional poverty assessment

Danilo Cavapozzi · Wei Han · Raffaele Miniaci

Received: 15 January 2014 / Accepted: 18 February 2015 / Published online: 7 April 2015

Abstract A multidimensional poverty assessment requires a weighting scheme to aggregate the well-being dimensions considered. We use Alkire and Foster's J. Public Econ. **95**, 476–487 (2011a) framework to discuss the channels through which a change of the weighting structure affects the outcomes of the analysis in terms of overall poverty assessment, its dimensional and subgroup decomposability and policy evaluation. We exploit the Survey on Health, Ageing and Retirement in Europe to evaluate how alternative weighting structures affect the measurement of poverty for the population of over-50s in ten European countries. Further, we show that in our empirical exercise the results based on hedonic weights estimated on the basis of life satisfaction self-assessments are robust to the presence of heterogeneous response styles across respondents.

Keywords Anchoring vignettes · Life satisfaction · Multidimensional poverty measurement · Weighting schemes

D. Cavapozzi (✉)
Department of Economics, Ca' Foscari University of Venice, Fondamenta San Giobbe, Cannaregio 873, 30121 Venezia, Italy
e-mail: danilo.cavapozzi@unive.it

D. Cavapozzi
Netspar, Tilburg, The Netherlands

W. Han
Blavatnik School of Government, University of Oxford, Albion House, Littlegate Street, OX1 1AN, Oxford, UK
e-mail: wei.han@bsg.ox.ac.uk

R. Miniaci
Department of Economics and Management, University of Brescia, Via S.Faustino 74/b, 25122 Brescia, Italy
e-mail: raffaele.miniaci@unibs.it

R. Miniaci
IEFE, Bocconi University, Milano, Italy

1 Introduction

During the past decades it has been gradually recognized that the concept of well-being cannot be comprehensively captured by any conventional unidimensional indicator based on income, consumption or expenditure (Nussbaum 2001; Sen 1985). Focusing on a unique dimension keeps blind of the information about the overall life quality, of which it might be worthwhile for policy-makers to keep track given that pursuing well-being rather than wealth itself appears to be the ultimate goal of human society (Ruger 2010).

Although the multidimensional perspective on well-being measurement moves beyond the focus on a single indicator, it is still far from reaching an agreement on how to translate this perspective into practice. One of the complex and highly debatable issues emerging in a multidimensional context of well-being research lies in how to set the relative weights across the dimensions. Summarizing the achievements with respect to different well-being dimensions in a single indicator is needed to measure the diffusion of poverty, defined as pronounced deprivation in well-being, within a population.

This paper is aimed at showing how the adoption of different weighting schemes affects the outcomes of a multidimensional poverty study. We choose to conduct our analysis based on the Alkire and Foster's multidimensional poverty framework (Alkire and Foster 2011a). One of its advantages lies in that it allows exploiting the information coming from achievements measured on ordinal and categorical scales, which is of significant importance in the policy analysis. Despite the fact that this approach is not the only one that can deal with ordinal or categorical variables,[1] it is the one most used in informing policy. The United Nations Development Programme has been including the Multidimensional Poverty Index (MPI) inspired by the Alkire and Foster's method in the Human Development Report since 2010. Besides, a Multidimensional Poverty Peer Network including Ministers and senior officials from 22 countries and 5 institutions has been established recently to promote the application of MPI in policy making.

According to the Alkire and Foster's approach, well-being dimensions are described by a set of one or more achievement indicators. The results with respect to the whole battery of achievement indicators can be aggregated into a single well-being score according to a weighting structure specified a priori. Poor households are those whose well-being scores fail to reach a minimum threshold. Alkire and Foster (2011a) propose a poverty measure, the adjusted headcount ratio, which reflects prevalence of poverty in the population and the intensity of the poverty among the poor. This measure can be decomposed in order to assess both the contribution of each dimension to overall poverty and how poverty varies across subgroups. The subgroup decomposability of the adjusted headcount ratio is also useful to investigate the determinants of the variations in poverty measurement originated by changes in the weighting scheme or in the distribution of the achievements in the population. In both cases it is possible to recognize a first part of the variation due to the change in the pool of families identified as poor and a second part due to the change in well-being of those families who are identified as poor regardless of the weighting scheme adopted or the distribution of achievements considered. The decomposition allows the researcher to investigate if the poverty assessment changes mainly because the set of households in poverty varies or because of variations in the well-being of the poor households.

[1] See Aaberge and Peluso (2012), Bosmans et al. (2013), Bossert et al. (2013), and Decancq et al. (2014) and Rippin (2013).

Various approaches to the choice of the weighting schemes have been proposed in the literature. Decancq and Lugo (2013) surveyed three main classes of weights: normative, data-driven and hybrid. Normative weights are based on an explicit value judgment of analysts about the trade-offs between the well-being dimensions. Data-driven weights are based on the actual distribution of the achievements in the society with respect to the indicators of interest. Hybrid weights combine value judgements and statistical facts. They lie in the middle between defining weights by arbitrary decisions of analysts and letting data distribution be the only criterion used.

How poverty assessments are affected by the weighting scheme remains an empirical issue. In our exercise, we focus on the elderly population in Europe, follow the classification by Decancq and Lugo (2013) and choose one example for each of the classes discussed above. As for normative weights, we use equal weighting, which is the weighting scheme most widely used in measuring multidimensional well-being due to its simplicity.[2] We follow the Human Development Index and MPI to assign equal weights to each dimension and equal weights to each achievement indicator in each dimension (UNDP 2011).[3] Within the class of data-driven weights, we adopt the frequency weights, which are motivated by the idea that, when assessing well-being, individuals put a high value on the shortfalls where the majorities do not fall short. We follow Desai and Shah (1988) to set the weight of a given achievement indicator as the corresponding proportion of the non-deprived in the society. Finally, within the hybrid class, we choose the hedonic weights, that is weights derived from life satisfaction self-assessments. Doing so we circumvent one of the weaknesses of the equal weighting, that is, that the value judgement about dimension trade-offs is set a priori by researchers. In fact, as noticed by Kingdon and Knight (2006, p. 1204) "[t]he value judgement implicit in this weighting need not correspond at all well to the valuations of these capabilities made by individuals in society. Subjective well-being may be a narrow metric but at least it corresponds to individual valuations and it is a metric that can be measured." A possible drawback of hedonic and frequency weights is that they are time- and population-specific: as the distribution of achievements and preferences can vary across households and over time, the use of the same set of these weights for various subgroups or time periods needs extra caution with respect to the adoption of the normative weights.

Many social science surveys ask respondents to rate their satisfaction with life according to a predetermined scale usually spanning from "very dissatisfied" to "very satisfied". Life satisfaction self-assessments have been widely used in the applied research focusing on well-being determinants (see for instance Frey and Stutzer 2002, and Dolan et al. 2008). When dealing with self-reported life satisfaction data it is important to recognize that, as a subjective measure, its variability across socioeconomic groups can be ascribed to genuine differentials in well-being (Schokkaert 2007) as well as heterogeneity in the way in which individuals with different characteristics interpret the scale used to provide self-assessments. As an example, two individuals might have different expectations about the conditions that should realize to self-define as satisfied with their lives. Then, even if they experience the same level of well-being, they might produce different self-assessments due to their different reporting styles. Neglecting such heterogeneity when studying life satisfaction determinants might end up with assigning to an explanatory variable a biased role coming from the combination between its relationship with the reporting style used

[2] It has been employed in approximately 50 % of the published studies (see Decancq and Lugo 2013).

[3] It can be set equally either at the dimension level or at the indicator level. Inherently, it is an arbitrary approach regardless how we make it equal.

in life-satisfaction self-assessments and its actual role in explaining genuine differences in well-being (see King et al. 2004, and Angelini et al. 2012, 2014).

Fleurbaey et al. (2009) and Decancq et al. (2014) suggest a framework for poverty measurement that explicitly takes into account individual preferences. They elicit the relative concerns of the individuals about the various dimensions of well-being based on life satisfaction self-assessment regressions. They exploit the longitudinal nature of their data to control for the possible role played by heterogeneity in response styles. In a similar spirit, we use the estimated coefficients of a statistical model that relates the stated life satisfaction levels with the achievement indicators to derive the weights of the latters in the well-being score function of Alkire and Foster (2011a). As in Fleurbaey et al. (2009) and Decancq et al. (2014), we also take into consideration the presence of response style heterogeneity, but we tackle this problem in a different way.

To this end, we exploit the information made available by the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE is a cross-country study that administers a multi-disciplinary questionnaire to a representative sample of individuals aged 50 or more living in Europe. Moreover, the second wave of SHARE provides us with a survey-instrument designed to take into account heterogeneity in the reporting styles used in collecting subjective data on life satisfaction. This approach is based on anchoring vignettes. A representative subsample of respondents is asked to report their own life satisfaction self-assessments along with their assessments about the life satisfaction of hypothetical individuals described in vignettes kept constant across respondents. Differences in vignette-evaluations provided by respondents can then be of use to identify heterogeneity in reporting styles and disentangle such variability from actual differentials in well-being. We therefore estimate two sets of hedonic weights, the first based on the estimates of the relationship between life satisfaction self-assessments, achievement indicators and individuals' characteristics, the second that formally controls also for the heterogeneity in reporting styles.

We therefore investigate the multidimensional poverty of the elderly in Europe using four sets of weights to ascertain to what extent the weighting scheme adopted affects poverty assessment and to shed light on the robustness of the results based on hedonic weights with respect to the relaxation of the assumption of invariance of reporting styles in the population.

The paper is organised as follows. Section 2 describes Alkire and Foster's multidimensional poverty framework that our analysis is built on. Section 3 is devoted to describe the approach we follow to derive the hedonic weights used in our analysis. Section 4 describes the data used in our empirical exercise and all the ingredients, but weights, involved in our application of the Alkire and Foster's multidimensional framework. Results are reported in Section 5. Section 6 presents our conclusions and policy implications.

2 A framework for the multidimensional poverty assessment

Sen (1976) concisely summarized two problems that must be faced in the poverty measurement: (1) the identification problem, i.e. how to choose the criterion of poverty and then distinguish those who fall into that criterion and those who do not; and (2) the aggregation problem, i.e. how to construct a poverty index using the available information on the poor. Dealing with these two issues is particularly challenging in a multidimensional framework. Alkire and Foster (2011a) tackle the identification problem by defining indicator-specific thresholds – which refer to specific achievements – and an overall threshold, which refers to a comprehensive well-being score based on the achievements. Moreover, they adopt the

Foster et al. (1984) framework to handle the aggregation problem and deliver a methodology satisfying desirable properties for poverty measurement.

In our brief introduction to Alkire and Foster's method, for sake of simplicity and without loss of generality, we assume that each dimension is represented by one indicator. All the results can be easily generalized to the case in which more indicators are considered for each dimension.

The units of our analysis are n households. Each dimension, $k = 1, 2 \ldots D$ is described by the achievement of household h on the k-th dimension, y_k^h. Every dimension has its own threshold, z_k, to indicate the minimum standard to attain to be not deprived. Let w_k be the weight of the k-th dimension, the weights sum up to 1, i.e. $\sum_{k=1}^{D} w_k = 1$. The first step of Alkire and Foster's identification procedure defines the achievement status of household h in terms of the dimension k as $a_k^h = \mu\left(y_k^h > z_k\right)$, where $\mu(\cdot) = 1$ if the expression in parentheses is true, zero otherwise. The problem of aggregation across dimensions is solved by defining a scalar well-being function as weighted average of dimension achievement statuses, that is by computing the overall well-being score of household h:

$$s^h = \sum_{k=1}^{D} a_k^h \, w_k \tag{1}$$

Finally, Alkire and Foster's procedure identifies as poor those households whose well-being score is lower than an arbitrarily chosen well-being threshold φ, with $\varphi \in (0,1)$. More formally, the poverty status of household h is defined as $P^h = \mu\left(s^h < \varphi\right)$.

The standard method to overcome the aggregation problem is to refer to the poverty incidence measured by the headcount ratio:

$$H = \frac{1}{n} \sum_{h=1}^{n} P^h \tag{2}$$

However, this index violates the dimensional monotonicity property, that is, other things being constant, if the shortfall of those identified as 'poor' varies, the headcount index remains unchanged.[4] As early as in Sen's well-known paper on the ordinal approach to poverty measurement (Sen 1976), monotonicity has been listed as one of the most important axioms that a valid poverty index should satisfy. The adjusted headcount ratio M, proposed by Alkire and Foster (2011a), satisfies the monotonicity axiom by combining information on the incidence of the poor in the population with the degree of poverty among the poor. The former is measured by the headcount ratio H; the latter is measured by the average shortfall among the poor:

$$A = \frac{\sum_{h=1}^{n} P^h \left(1 - s^h\right)}{\sum_{h=1}^{n} P^h} \tag{3}$$

Formally, the adjusted headcount ratio is

$$M = HA = \frac{1}{n} \sum_{h=1}^{n} P^h \left(1 - s^h\right) \tag{4}$$

[4] In this paper, shortfall refers to the gap between the actual well-being score s_h and the full well-being score (i.e. when all the minimum standards are met and $s^h = 1$) rather than to the gap with respect to the well-being threshold φ.

which also represents the total shortfall experienced by the poor $\left(\sum_{h=1}^{n} P^h \left(1 - s^h\right)\right)$ divided by the maximum shortfall that could be experienced by the entire population. When none of the households meet any minimum standard with respect to any indicator, $s^h = \sum_{k=1}^{D} a_k^h w_k = 0$ and thus $\sum_{h=1}^{n} P^h \left(1 - s^h\right) = n$.

The adjusted headcount ratio satisfies a range of desirable properties for poverty indexes[5] (see Alkire and Foster 2011a). In particular, the decomposability by subgroup and dimension is of significant importance to policy makers.

Subgroup decomposability The adjusted headcount ratio M in the population is the weighted average of the same measure calculated for mutually exclusive subgroups, where weights are subgroup population shares. Formally, suppose population can be divided into q groups. Let θ_g be the population share of subgroup g. Denote the adjusted headcount ratio of subgroup g as M_g, so we have $M = \sum_{g=1}^{q} \theta_g M_g$. The relative contribution of subgroup g to overall poverty M depends on both θ_g and M_g and can be written as $RG_g = \frac{\theta_g M_g}{M}$.

Dimensional decomposability The overall poverty measure M is the weighted average of the *censored* deprivation indexes of each dimension, where the weights are the dimension-specific weights. The censored deprivation index I_k for the dimension k is the fraction of households who are poor and deprived with respect to the k-th dimension:

$$I_k = \frac{1}{n} \sum_{h=1}^{n} P^h (1 - a_k^h) \tag{5}$$

The term “censored” is used to emphasize that I_k considers only the deprivation status of the poor. The overall adjusted headcount ratio can be written as

$$M = \sum_{k=1}^{D} I_k w_k, \tag{6}$$

and the relative contribution of the dimension k to the overall poverty is $RD_k = \frac{I_k w_k}{M}$. If more than one indicator is used to describe a dimension, the relative contribution of the dimensions is the summation of the relative contributions of the indicators in that dimension.

The definition of the well-being score, $s^h = \sum_{k=1}^{D} a_k^h w_k$, makes apparent that changing the weights potentially has influence on who is identified as poor, on the headcount and the adjusted headcount ratios and on the dimensional as well as subgroup decompositions. The overall effects on the poverty assessment are difficult to predict. This is not a peculiarity of the Alkire and Foster’s approach. It is true for any index that aggregates individual positions and/or various dimensions, because of the role played by the joint distribution of the dimensions in the population (see e.g. Paruolo et al. 2013). Comparing alternative poverty assessments produced by alternative weighting schemes is standard practice in the literature (e.g. Bourguignon and Chakravarty 2003). We also run a similar exercise in the empirical section of the paper, but we enrich this practice by suggesting a decomposition

[5] It satisfies the following properties: replication invariance, symmetry, poverty and deprivation focus, weak and dimensional monotonicity, non-triviality, normalization, weak rearrangement and decomposability.

of the variation of the adjusted headcount index that provides insights on the mechanisms originating it.

More specifically and without loss of generality, consider the simple case of two weighting schemes w and w', that differ for the weights of the indicators p and q, with $w'_p - w_p = w_q - w'_q$.[6] The two sets of weights generate potentially different well-being scores for every household and consequently different pools of families identified as poor. We define n_{out} as the number of households identified as poor with the weighting scheme w but out of poverty with the new score vector w', n_{in} is the number of households falling in poverty with the weighting scheme w' (but not with w), and finally n_{stay} are the households who are identified as poor under both weighting schemes. Using the group decomposition property of the adjusted headcount ratio, the overall variation $\Delta M = M' - M$ can be written as the weighted average of the variations of the ratios of the three groups of households mentioned above:

$$\Delta M = \frac{n_{out}}{n} \Delta M_{out} + \frac{n_{in}}{n} \Delta M_{in} + \frac{n_{stay}}{n} \Delta M_{stay} \tag{7}$$

where the *in*, *out* and *stay* subscripts identify the pools of households the ratios $M$and M' refer to. The adjusted headcount ratio equals zero whenever it is computed over a set of non-poor households, then $\Delta M_{out} = \Delta M'_{out} - M_{out} = -M_{out}$, $\Delta M_{in} = M'_{in} - M_{in} = M'_{in}$, and the expression above can be rewritten as

$$\Delta M = \left(\frac{n_{in}}{n} M'_{in} - \frac{n_{out}}{n} M_{out} \right) + \frac{n_{stay}}{n} \Delta M_{stay} = \Delta M_1 + \Delta M_2 \tag{8}$$

The ΔM_1 term reflects a composition effect driven by the households who change their poverty status from one weighting scheme to the other. The term ΔM_2 is instead driven by the effect of the change in the weighting scheme on the well-being score of the households who are poor regardless of the weighting scheme used. Consequently, it is impossible to predict the sign of ΔM without knowing the underlying distribution of all the indicators within the sets of households changing their poverty status or remaining poor under both weighting schemes.

The group decomposability of the adjusted headcount ratio is also useful to analyse situations in which the weights remain constant, but the distribution of the achievements in the population changes. Consider for instance the case in which the adjusted headcount ratio is used to evaluate the effectiveness of an anti-poverty policy that targets the households deprived in the dimension p. Only the households who did not meet the minimum standard of indicator p and were identified as poor before the intervention contribute to reduce the original level of the adjusted headcount ratio M. Define M' the poverty measure after the intervention, the variation of interest $\Delta M = M' - M$ can be decomposed as:

$$\Delta M = \frac{n_{out}}{n} \Delta M_{out} + \frac{n_{stay}}{n} \Delta M_{stay} \tag{9}$$

where the *out* subscript identifies the pools of households beneficiary of the policy that exit the poverty status thanks to the intervention, while *stay* identifies the beneficiaries who do not change their poverty status. It can be shown that for the first group of households

[6] A more detailed discussion of these results can be found in Cavapozzi et al. (2013).

$\Delta M_{out} = -M_{out}$, while for the latter is $\Delta M_{stay} = -w_p$. The decomposition can therefore be written as

$$\Delta M = -\left(\frac{n_{out}}{n} M_{out} + \frac{n_{stay}}{n} w_p\right) = \Delta M_1 + \Delta M_2. \tag{10}$$

The expression above makes apparent that the effect of the policy is larger the larger the number of households escaping poverty (n_{out}), the more deprived are these households (M_{out}) and the higher is the weight associated with the targeted dimension (w_p). Changing the weighting scheme changes not only the overall effect ΔM but also the relative importance of ΔM_1 and ΔM_2. The latter effect impacts on the assessment of the relevance of the mechanisms that make the policy intervention effective.

3 Deriving weights with the hedonic approach

The hedonic approach of deriving a weighting scheme is hybrid since it combines value judgements about trade-offs among dimensions, as it is typical in the normative weighting, with statistical facts. We use life satisfaction self-assessments of respondents to elicit value judgements about trade-offs between well-being dimensions. A widely used approach to measure well-being in applied research is to ask individuals to evaluate their life satisfaction according to a predetermined scale, e.g. by answering the question “How satisfied are you with your life in general?”. Self-assessments are measured according to an ordinal scale, such as “Very dissatisfied”, “Dissatisfied”, “Neither satisfied nor dissatisfied”, “Satisfied”, “Very satisfied”. Frey and Stutzer (2002) and Dolan et al. (2008) survey the main findings of the empirical research on the determinants of life satisfaction self-assessments.

On the one hand, life satisfaction self-assessments have the advantage of summarizing in a single index all the factors that individuals consider relevant determinants of their well-being. On the other hand, a recent research vein (Angelini et al. 2012 and 2014, Kapteyn et al. 2009) has shown that the benchmarks used to self-evaluate life satisfaction are not invariant across individuals but depend on their own characteristics. Even if individuals are asked to self-evaluate their own life satisfaction according to the same survey question, they might provide different evaluations due to inter-personal and inter-cultural heterogeneity in the interpretation of the response scale. Furthermore, a phenomenon of adaptation might be at work. In fact, individuals may adjust their aspiration levels to their realistic opportunities (Schokkaert 2007). In psychometrics such heterogeneity has been called differential item functioning (DIF). If DIF is an issue, life satisfaction self-assessments fail to be comparable across individuals or socioeconomic groups since their differences might not reflect actual differences in well-being but only differences in the reporting styles adopted by respondents. Individuals with the same actual level of well-being might provide different life satisfaction self-evaluations because they have in mind different concepts about what being satisfied with their life means. As a consequence, the presence of DIF implies that a well-being analysis based on the comparison of life satisfaction self-evaluations should take into account heterogeneity in reporting styles in order to provide meaningful results.

This paper takes advantage of the SHARE data to control for DIF by a vignette methodology. After having provided life satisfaction self-assessments, a subsample of SHARE respondents are asked to evaluate the life satisfaction of two hypothetical individuals described in particular situations (anchoring vignettes), which are reported below.

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 per month.[7] 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?

Respondents' evaluations of vignettes are recorded according to the same response scale used for their self-assessments ("Very dissatisfied", "Dissatisfied", "Neither satisfied nor dissatisfied", "Satisfied", "Very satisfied").

The situations described in the vignettes do not vary across respondents, who are also explicitly asked to evaluate the vignettes according to their own preferences. Differences in the evaluations of the anchoring vignettes can be ascribed to the heterogeneity in the reporting styles of respondents and be of use to filter the life satisfaction self-assessments of respondents from DIF as long as respondents use the same reporting style when assessing the life satisfaction of themselves and of the hypothetical individuals described in the vignettes (response consistency) and the life satisfaction of the hypothetical individuals in the vignettes is on average perceived by respondents in the same way (vignette equivalence).

More specifically, we analyze the determinants of life satisfaction and control for the presence of DIF by the hierarchical ordered probit (Hopit) model introduced by King et al. (2004). This econometric specification consists of two components modeling self-assessments and vignette evaluations as ordered variables.

Self-assessment component Let Y_i^* be the life satisfaction perceived by individual $i = 1, \dots, I$ and assume that it comes from a linear combination of individual characteristics stored in the row vector X_i and an error term $\varepsilon_i \sim N(0, 1)$ independent of X_i,

$$Y_i^* = X_i \beta + \varepsilon_i \tag{11}$$

where β is a vector of unknown parameters. The vector X_i includes the achievement indicators as well as the individual characteristics related to their frame of reference (see Fleurbaey et al. 2009). Controlling for the individual characteristics is also "necessary to 'clean' the happiness measure to separate the 'ethically' relevant information from the irrelevant noise" (Schokkaert 2007, p. 428). Although Y_i^* cannot be observed, we know individual's life satisfaction self-evaluation Y_i, which is coded as an ordered discrete variable spanning from 1 ("Very dissatisfied") to 5 ("Very satisfied"). The transformation connecting the unobserved continuous variable Y_i^* with the observed discrete outcome Y_i can be written as follows

$$Y_i = j \text{ if } \tau_i^{j-1} \leq Y_i^* \leq \tau_j^i \quad j = 1, \dots, 5 \tag{12}$$

[7] This value is PPP-adjusted to account for cross-country differentials in price levels.

The thresholds τ_i^j are individual-specific and depend on X_i

$$\tau_i^0 = -\infty, \qquad \tau_i^5 = \infty \tag{13}$$

$$\tau_i^1 = X_i\gamma^1 \tag{14}$$

$$\tau_i^j = \tau_i^{j-1} + \exp\left(X_i\gamma^j\right), \quad j = 2, 3, 4 \tag{15}$$

where γ^j are vectors of unknown parameters. Notice that $X_i\gamma^j$, $j = 2, 3, 4$ is the argument of the exponential function to ensure the ascending order of the thresholds. The set of thresholds τ_i^j and the parameters γ^j formally allow individuals with different characteristics and achievements to provide different self-evaluations Y despite the same perceived level of life satisfaction Y^*. The Hopit model can then be seen as a generalization of the standard ordered probit specification, which restricts the thresholds to be invariant across individuals and implicitly assumes that reporting styles adopted by individuals do not depend on their own characteristics.

The self-assessment component of the Hopit model formally shows that the effect of X_i on the observed outcome Y_i is twofold since X_i affects the life satisfaction perceived by individuals Y_i^* as well as the way in which such perception is reported on a discrete scale, which is summarized by the thresholds τ_i^j. The information conveyed by life satisfaction self-evaluations is not sufficient to disentangle the effect of the individual characteristics X_i on Y_i^* and their effect on the thresholds τ_i^j. We make use of vignette evaluations to achieve this goal and identify the parameter vectors β and γ^j.

Vignette evaluation component Let Z_{il}^* be the life satisfaction of the hypothetical person in vignette $l = 1, 2$ perceived by individual i. We assume that

$$Z_{il}^* = \theta_l + v_{il} \tag{16}$$

where $v_{il} \sim N\left(0, \sigma_l^2\right)$ and v_{il} is independent of ε_i and X_i. The parameter θ_l is assumed to be vignette-specific and invariant across individuals. This restriction follows from the vignette equivalence assumption, according to which respondents have the same perception of the life satisfaction of the hypothetical person in the vignette up to an individual idiosyncratic error term. Again, we cannot observe the perception Z_{il}^* but we know the evaluation Z_{il}, defined as

$$Z_{il} = j \text{ if } \tau_i^{j-1} \leq Z_{il}^* \leq \tau_i^j, \quad j = 1, \ldots, 5 \tag{17}$$

The response consistency assumption implies that thresholds τ_i^j are those used to derive the life satisfaction self-assessments and therefore we can combine the information relevant for the two components of the model to identify all the parameters of interest in the Hopit model. Along the lines of King et al. (2004), the joint estimation can be carried out by maximum likelihood techniques.

We estimate two sets of hedonic weights. The first one derived from a standard ordered probit regression. The standardized weight for the achievement indicator k, w_k, is retrieved from the corresponding estimated coefficients $\hat{\beta}_k$ as

$$w_k = \frac{\hat{\beta}_k}{\sum_{l=1}^{D} \hat{\beta}_l} \quad k = 1, \ldots, D \tag{18}$$

The second set of standardized hedonic weights is derived from the estimated coefficients of the achievement indicators in the self-assessment component of the Hopit model.

This vector of hedonic weights is expected to reflect the relationship between achievement indicators and well-being once their effect on reporting styles in life satisfaction self-assessments has been filtered out.

4 Data, dimensions, indicators and thresholds

In this paper we use data from the 2006 wave of the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE is an interdisciplinary survey on ageing that is run every two years and collects extensive information on health, socioeconomic status and family interactions of individuals aged 50 and over in a host of European countries.[8] The choice of using SHARE rather than other well established surveys (e.g. EU-SILC for the European countries) is dictated by the fact that SHARE collects self-assessments and anchoring vignette evaluations on life satisfaction and makes it possible to investigate if the hedonic weights based on the life satisfaction self-assessments of respondents are robust to the presence of heterogeneity in response styles.

Data are collected by face-to-face, computer-aided personal interviews (CAPI), supplemented by a self-completion paper and pencil questionnaire, which collects self-assessments and vignette evaluations on life satisfaction. We select only those respondents who provide both the self-evaluation and at least one vignette evaluation. Our final estimation sample for the hedonic weights is composed by 3,804 households, corresponding to 5,545 individuals living in Sweden, Denmark, Germany, The Netherlands, Belgium, France, Greece, Italy, Spain and Czech Republic.[9]

Different multidimensional poverty indexes consider alternative sets of dimensions due to differences in theoretical perspectives, reference population and data limitation. According to Sen (2004), the choice of the dimensions should focus: (1) on those that are of special importance to the society or people in question; (2) on those that are an appropriate focus for public policy, rather than a private good or a capability that cannot be influenced from outside. Material deprivation, health conditions, educational attainments, empowerment, labour market participation, environmental quality, safety from violence, and social relationships are all relevant domains and their relevance has been assessed for the European Union population (Eurostat 2012).

In our illustrative exercise we focus on a representative sample of elderly respondents living in ten European countries and we consider three dimensions to represent the main drivers of their well-being: economic situation, housing and health conditions. The economic dimension is meant to describe the monetary resources available to the household. It includes two achievement indicators: per-capita net income and per-capita net wealth. The thresholds for income and wealth indicators are set equal to 60 % of the country specific median values. By doing so, we follow Stiglitz's Commission suggestions to consider both income and wealth. The housing dimension has one achievement indicator, a measure of accessibility of the dwelling given by the number of steps people have to climb up and/or down to the entrance of their home. The architectural barriers of the accommodation are potentially relevant for the population we consider, as ageing is often

[8] See Börsch-Supan et al. (2008) for further details.

[9] We restrict our sample to the countries in which vignette data have been collected with the exception of Poland, for which the data used for the analysis show some inconsistency with respect to the rest of the sample.

accompanied by limitation to the mobility. In our sample, 44 % of respondents report limitations with mobility, arm functions and fine motor function. This percentage is greater than 50 % for Belgium, Czech Republic, Greece and Italy. We considered also some overcrowding indicators, but, once controlled for other dimensions, none of them proved to have any significant effect on the self-assessed life satisfaction of the households. Unfortunately, we do not have information on the quality of the neighbourhood for most of the estimation sample.

Finally, we use three achievement indicators for the health domain: the presence of chronic diseases (in a list of 17 diseases) and the number of limitations with the activities of daily living (ADL, that is dressing, walking across a room, bathing or showering, eating, getting in and out of bed, using the toilet) to take into consideration physical health, and the presence of depression symptoms (EURO-D caseness, see Prince et al. 1999) to summarize mental wellbeing. The inclusion of an indicator of depression symptoms may be controversial because it could make more difficult the elicitation of the preferences and the estimation of the hedonic weights from questions on life satisfaction. In fact, the respondents may not clearly distinguish between the cognitive valuation of their life and their depressed feelings (see Fleurbaey et al. 2009). We decided to consider EURO-D caseness because mental wellbeing is a crucial dimension of the overall individual well-being and public intervention in this field has been often advocated (e.g. Nussbaum 2008). Moreover, the Hopit model takes explicitly into consideration that the depressed feelings may alter individuals' response styles, and by doing so it improves the possibility to separately identify the cognitive component of the life satisfaction evaluation.

Table 1 summarizes the details about the dimensions, the achievement indicators and the corresponding thresholds used to define the presence of deprivation.[10] Rephrasing Alkire and Foster's (2011b) words, the aim of our empirical exercise is not to suggest that this set of indicators, dimensions and cut-offs is appropriate in every application. Rather, the aim of our illustrative exercise is twofold. On the one hand, we aim at describing the effects of changes in weighting schemes on the outcomes of the multidimensional poverty analysis run according to the Alkire and Foster's methodology on a sample of elderly individuals living in ten European countries. On the other hand, we want to assess the robustness of the results obtained with hedonic weights based on respondents' life satisfaction self-assessments with respect to the presence of heterogeneity in reporting styles.

The proportion of households who meet the minimum standards with respect to single indicators ranges between 44.5 % for the presence of chronic diseases to 86.3 % for the presence of impediments with the activities of daily living. It is important to notice that the indicators are only weakly correlated between them (Table 2). This suggests that the information conveyed by the indicators considered is not statistically redundant.

Unlike the thresholds of the indicators (z_k) that can be mostly determined by convention, the choice of the overall well-being threshold φ seems more arbitrary and less grounded since it works across the dimensions where general understanding is hard to be applied. We take 0.6 as the well-being threshold in all the analysis and conduct a sensitivity analysis of the robustness of the results with respect to the choice of this parameter.

[10] All the descriptive statistics and findings are based on the sample of households, unless otherwise stated.

Table 1 Dimensions, indicators and thresholds

Dimensions	Achievement Indicators	Thresholds (meet the minimum standard if)	Percentage meeting the minimum standards
Economic	Per-capita net income	equal or above 60 % of median (country specific)	78.82
	Per-capita net wealth	equal or above 60 % of median (country specific)	66.70
Housing	Dwelling accessibility	less than 16 steps to climb up/down to entrance	82.62
	Chronic disease	none of household members have more than two chronic diseases	44.53
Health	ADL	none of household members have ADL problem	86.26
	EURO-D	none of household members have EURO-D caseness	66.83

Note: The percentage meeting the minimum standard of a given indicator is a weighted average for the entire sample of households

5 Results

5.1 Life satisfaction and hedonic regression

To set the hedonic weights, we exploit the individual question about life satisfaction in general. The top panel of Table 3 shows that about 77 % of the interviewed individuals declared to be satisfied or very satisfied, while 5.4 % declared to be dissatisfied or very dissatisfied. The lower panel provides a first insight on the relation between achievements and life satisfaction. For each achievement indicator, we compute the risk ratios for every level of life satisfaction, that is, the ratio of the probability that individuals deprived in that indicator declare a given level of satisfaction, over the same probability for those not deprived. Thus, we see that the percentage of income-poor individuals who declared themselves to be very satisfied with their life is only 3/4 of those whose income is above the income threshold. Viceversa, the percentage of dissatisfied among the income-poor is the double of those with higher income. The differences between being below and above the thresholds of indicators

Table 2 Tetrachoric correlation coefficients among indicators $\left(a_k^h\right)$

	Per-capita net income	Per-capita net wealth	Dwelling accessibility	Chronic disease	ADL	EURO-D
Per-capita net income	1.0000					
Per-capita net wealth	0.3565	1.0000				
Dwelling accessibility	−0.0164	0.2384	1.0000			
Chronic disease	0.0253	0.2000	0.0476	1.0000		
ADL	0.1155	0.2067	−0.0500	0.5238	1.0000	
EURO-D	0.0588	0.1458	0.0618	0.3755	0.4414	1.0000

Table 3 Distribution of the answers to the life-satisfaction question and relative risk ratios for each life satisfaction level by indicator

	Very dissatisfied	Dissatisfied	Neither satisfied nor dissatisfied	Satisfied	Very satisfied
Percentage of respondents	0.63	4.80	17.48	56.77	20.32
Relative risk ratios of deprived vs non deprived					
Per-capita net income	2.74	2.02	1.36	0.90	0.74
Per-capita net wealth	1.82	1.79	1.47	0.97	0.61
Dwelling accessibility	1.63	1.54	1.60	0.95	0.57
Chronic disease	5.84	2.24	1.29	0.97	0.72
ADL	6.62	2.97	1.72	0.85	0.42
EURO-D	4.40	5.85	1.89	0.85	0.45

Note: Sample of 5545 individuals used to estimate the hedonic weights. The relative risk ratio is the ratio of the fraction of deprived individuals declaring a specific level of satisfaction, over the same fraction computed among the non-deprived population.

are even more striking when focusing on the health indicators, suggesting a prominent role played by the health dimension on the overall life-satisfaction of the individuals.

As explained in the Section 4, the Hopit model can be seen as a generalization of the standard ordered probit model which exploits the information provided by the vignette evaluations to account for heterogeneity in response style. Table 4 shows the distribution of respondents' evaluations of the life satisfaction of the hypothetical individuals described in the vignettes. While about 44 % of respondents rate John (the person in vignette 1) as very dissatisfied or dissatisfied with his life, only 15 % of them think that John is at least satisfied. Also, while 13 % of respondents rate Carry (the person in the second vignette) as dissatisfied or very dissatisfied, 55 % of the sample think she is at least satisfied. Although the same vignettes about John and Carry have been administered to all the respondents, their evaluations show considerable variability and suggest the presence of heterogeneity in the way they report life satisfaction. If this is an issue, comparisons of life satisfaction self-assessments neglecting this source of heterogeneity might bring about misleading results.

We estimate the hedonic weights considering a full set of observable household and individual characteristics: country of residence, gender, age, presence of a cohabiting partner,

Table 4 Distribution of the answers to the vignette evaluation

	Very dissatisfied	Dissatisfied	Neither satisfied nor dissatisfied	Satisfied	Very satisfied
Vignette 1 (John)					
Percentage of respondents	5.78	38.59	40.35	14.32	0.96
Vignette 2 (Carry)					
Percentage of respondents	1.30	11.83	30.58	48.65	7.63

Note: Sample of 5545 individuals used to estimate the hedonic weights

Table 5 Weights derived from different approaches

	Equal weights	Frequency weights	Hedonic weights	
			Ordered probit	Hopit model
Per-capita net income	0.1667	0.1851	0.0569	0.1092
Per-capita net wealth	0.1667	0.1567	0.1167	0.1629
Dwelling accessibility	0.3333	0.1940	0.0986	0.0751
Chronic disease	0.1111	0.1046	0.1314	0.0788
ADL	0.1111	0.2026	0.2626	0.2359
EURO-D	0.1111	0.1570	0.3338	0.3380

children, and grandchildren, employment status, involvement in social activities, education, home ownership, type of the area the accommodation is located in, season at the time of the interview (see Appendix for the descriptive statistics).

The regression results (in Appendix) for the ordered probit and Hopit model show that the achievement indicators are strongly correlated with the self-reported life satisfaction and that demographic characteristics play a significant role.[11] Life satisfaction exhibits remarkable cross-country heterogeneity, it is higher for women, it is at its minimum among individuals aged between 50 and 55, which is consistent with Blanchflower and Oswald (2008) and de Ree and Alessie (2011). Further, life satisfaction increases with the presence of a cohabiting partner, the involvement in social activities and with being at work or retiree instead of out of work due to reasons other than retirement. As long as the achievement indicators are correlated with this set of individual and household characteristics, omitting the latter from the right-hand side of the regression will lead to a biased estimate of the actual relationship between life satisfaction and the achievement indicators (Schokkaert 2007).

When we use the Hopit model to control for the possible effect of the heterogeneity in response styles due to differential item functioning (DIF), we can see that this is correlated with country and seasonal dummies, age, the presence of a cohabiting partner, employment status, home-ownership, and the type of area in which the accommodation is located. Some of the achievement indicators, in particular those related to the presence of chronic diseases or depression symptoms, affect the thresholds τ_i^j. Overall, our results confirm the evidence provided by Angelini et al. (2012 and 2014) that there is heterogeneity in the response styles. Therefore, the estimation of the hedonic weights can be biased if such heterogeneity is neglected.

5.2 Comparing alternative weighting schemes

The four sets of weights are presented in Table 5. We set the equal weights mimicking the Human Development Index and Multidimensional Poverty Index, that is, the three dimensions have the same relevance and the achievement indicators share the same weight within each dimension (UNDP 2011). For the frequency weights, we follow Desai and Shah (1988)

[11] We refrain from interpreting these estimates in terms of causal effects because of the possible endogeneity of some of the explanatory variables. The estimated coefficients gauge the partial correlation of the variables with the reported life satisfaction.

Table 6 Dimensional decomposition

	Equal weights	Frequency weights	Hedonic weights	
			Ordered probit	Hopit model
H	0.2380	0.2759	0.3112	0.3165
M	0.1364	0.1496	0.1881	0.1848
Relative contribution of indicators and dimensions to the overall adjusted headcount index *M* (%)				
Per-capita net income	13.10	15.21	2.68	5.70
Per-capita net wealth	20.03	20.26	9.66	13.72
Economic dimension	*33.13*	*35.46*	*12.34*	*19.42*
Dwelling accessibility	36.03	12.67	4.14	3.11
Housing dimension	*36.03*	*12.67*	*4.14*	*3.11*
Chronic disease	15.32	17.38	18.75	11.35
ADL	5.49	14.20	14.97	13.38
EURO-D	10.02	20.28	49.80	52.75
Health dimension	*30.84*	*51.86*	*83.52*	*77.48*

Note: The relative contribution of each dimension is the sum of the relative contributions of the corresponding indicators

to set the weight of every achievement indicator as the proportion of the non-deprived households in the sample.[12]

As compared with the equal weighting scheme, the frequency one reduces the weight of the housing domain from 33.33 % to 19.40 % in favor of the health conditions, whose weight goes from 33.33 % to 46.42 %. The weights attached to the economic conditions remain almost unchanged. As compared with frequency weights, the hedonic approach doubles the weight of the EURO-D indicator to about 33 %, and in general it increases the prominence of the health domain. The weights of net income and net wealth are lower in the hedonic weights than in the other two schemes, and they are particularly low for the weights computed with the ordered probit model. With the hedonic weights, the accessibility of the accommodation loses most of its relevance.

Each weighting scheme gives origin to a different well-being score. Setting the poverty threshold φ equal to 0.6 produces four headcount ratios that vary between 23.8 % of the equal weights and 31.7 % of the hedonic weights with the vignettes, whereas the corresponding adjusted headcount ratios M range between 13.6 % and 18.8 % (see Table 6).

Although the level of the adjusted headcount indices is similar across alternative weighting structures, the relative contribution of the dimensions is remarkably different. As pointed out in Section 2, the relative contribution of each dimension comes from the summation of the relative contributions of all indicators in the dimension. Consider the health dimension: its contribution to the overall level of the adjusted headcount index is 51.9 % for the frequency weights and 77.5 % for the hedonic weights taking the heterogeneity in response styles into account (Table 6). This variation is driven by the sharp increase in the relative contribution of the EURO-D indicator, which changes from 20.28 % to 52.75 %. As for

[12] Frequency weights are standardized in order to sum up to one. Equal weights are standardized by construction.

Table 7 Adjusted headcount decomposition by age class, by weighting scheme

Age class	Population share (%) (θ_g)	Subgroup adjusted headcount ratio (M_g)			
		Equal weights	Frequency weights	Hedonic weights	
				Ordered probit	Hopit model
55 or less	17.93	0.1040	0.0888	0.1207	0.1266
56–60	18.64	0.1353	0.1345	0.1709	0.1696
61–65	15.88	0.1269	0.1287	0.1635	0.1589
66–75	28.22	0.1388	0.1412	0.1770	0.1730
76+	19.34	0.1718	0.2500	0.3034	0.2917

the economic condition, its relative contribution is 35.5 % with frequency weighting and reduces to 12.3 % with hedonic weighting based on the ordered probit model. Most of this reduction is explained by the change in the contribution of the per-capita income indicator that shrinks by about 80 %. This variability magnifies for the housing dimension: it explains more than one third of poverty under equal weighting but only 3 % under the hedonic weighting based on the Hopit model.[13]

Besides, the differences between subgroups can be affected by the weighting structure too. For instance, if we look at Table 7, we find that the households where the oldest member is aged 55 or less experience the lowest level of poverty, whereas those in which the oldest member is aged 76 or more the highest. However, while the adjusted headcount ratio of the youngest households is 40 % lower than the one of oldest households under equal weighting, this reduction is around 60 % for the remaining weighting schemes.

Alternative weighting schemes may deliver remarkably different pictures of poverty dynamics over time (e.g. Bourguignon and Chakravarty 2003) as well as different evaluations of anti-poverty policies. Consider for instance a hypothetical intervention that solves income deprivation. Table 8 shows that if the intervention were evaluated using equal weights, the policy maker would register 6.16 % of the population exiting the poverty status and 4.57 % enjoying the benefits of the policy without exiting poverty. In terms of adjusted headcount ratio, starting from an initial level of $M = 0.1364$ (see Table 6), the policy would produce a reduction of $\Delta M = -0.0383$ (−28.1 %), almost completely due to the well-being improvement of those exiting the poverty status ($\Delta M_1/\Delta M = 80.2$ %). The same policy evaluated with the hedonic weights computed with the ordered probit would account for only 1.01 % of the population exiting poverty, a very limited reduction of M ($\Delta M = -0.0086$, a mere $-$ 4.6 %), only half of it originated by households escaping poverty ($\Delta M_1/\Delta M = 49.1$ %). That such a policy would be judged more effective using equal weights rather than hedonic weights may seem obvious, given that in the first case the weight for income achievement is 0.1667 while in the latter is 0.0569. But this is not always the case. Assume the policy maker were able to eradicate ADL

[13] We carried out a sensitivity analysis by setting an alternative poverty threshold $\varphi = 0.47$, that is the 60 % of the median of the well-being score computed with the equal weights. Our results are confirmed. In order to check whether the number of indicators per dimension drives the decomposition analysis, we also estimated two alternative sets of hedonic weights by replacing the indicators in the ordered probit and Hopit models with their averages by dimension. By doing this, we estimate one parameter for each dimension, the indicators within a given dimension equally split the dimensional weight and are finally standardized. Our results are again confirmed.

Table 8 Effects of interventions under different weighting schemes

Indicator targeted by the policy	Weighting schemes	Relative group size (%)		Group-specific adjusted headcount index changes		Overall adjusted headcount index change	
		n_{out}/n	n_{stay}/n	ΔM_{out}	ΔM_{stay}	ΔM	ΔM_1
Per capita net income	1	6.16	4.57	−0.4988	−0.1667	−0.0383	−0.0307
	2	5.61	6.68	−0.4732	−0.1851	−0.0389	−0.0265
	3	1.01	7.87	−0.4131	−0.0569	−0.0086	−0.0042
	4	1.86	7.79	−0.4390	−0.1092	−0.0167	−0.0082
ADL	1	2.87	3.87	−0.4960	−0.1111	−0.0186	−0.0142
	2	4.20	6.29	−0.4774	−0.2026	−0.0328	−0.0200
	3	2.54	8.18	−0.5506	−0.2626	−0.0355	−0.0140
	4	2.08	8.40	−0.5321	−0.2359	−0.0309	−0.0111

Note: 1 - equal weights, 2 - frequency weights; 3 - hedonic weights based on ordered probit; 4 - hedonic weights based on the Hopit model

deprivation, whose weight is 0.2026 with frequency weights and 0.2626 for hedonic weights derived from the ordered probit model. Such intervention turns out to be more effective if assessed under frequency weighting rather than the hedonic weighting scheme considered since the relative variation in the M amounts to $\Delta M/M = -0.328/0.1496 = -21.9$ % and $\Delta M/M = -0.0355/0.1881 = -18.9$ % respectively.

If we look at the effects of the interventions under the hedonic weights based on the Hopit model, we find that they are extremely close to their analogues under the alternative set of hedonic weights. In particular, the percentages of households exiting the poverty status or enjoying the benefits of the policy without exiting poverty and the variation in the adjusted headcount ratio explained by well-being improvement of the households exiting the poverty status are virtually unaffected. We interpret these results as evidence in favor of the stability of the poverty assessment with hedonic weights to the relaxation of the assumption of invariance of reporting styles in the population.

Finally, Table 9 helps to understand the origin of the differences between poverty measurements by decomposing the variation of the adjusted headcount ratios observed in Table 6. Abandoning the equal weights for the frequency weights, 4.98 % of the population exits the poverty status, while 8.77 % enters it. Despite the fact that the correlation between the two well-being scores is 94.8 %, only 79.1 % of the poor under equal weights are classified as such also with the frequency weights. The overall change in the headcount ratio is $\Delta M = 0.0132$ (+9.7 % with respect to the starting level of $M = 0.1362$), the variation due to the change in the pool of households in poverty is $\Delta M_1 = 0.0162$, and, consequently, the variation in the poverty assessment of the households identified as poor under both weighting schemes is $\Delta M_2 = 0.0132 - 0.0162 = -0.003$. The ΔM_1 component predominates over ΔM_2. This decomposition shows that the information provided by the achievements of the set of households originally classified as poor is insufficient to predict the variations in the poverty assessments due to the changes in the weighting schemes. Most of the variation in the adjusted headcount ratio is explained by the outcomes of the households changing their poverty status from one weighting scheme to the other.

Although going from equal weights to the hedonic weights computed with the ordered probit, implies a wider reshuffle of the set of poor households, the contribution of this

Table 9 Effects of changes of weighting scheme on the adjusted headcount ratio M

Weighting schemes		Relative group size (%)			Group-specific adjusted headcount ratio changes			Overall adjusted headcount ratio change	
From w	To w'	n_{out}/n	n_{in}/n	n_{stay}/n	ΔM_{out}	ΔM_{in}	ΔM_{stay}	ΔM	ΔM_1
1	2	4.98	8.77	18.82	−0.4636	0.4480	−0.0162	0.0132	0.0162
	3	9.19	16.51	14.61	−0.5126	0.5456	0.0592	0.0516	0.0430
3	4	0.25	0.77	30.88	−0.4926	0.4472	−0.0180	−0.0033	0.0022

Note: 1 - equal weights, 2 - frequency weights; 3 - hedonic weights based on ordered probit; 4 - hedonic weights based on the Hopit model

component to the overall change of the adjusted headcount ratio is lower ($\Delta M_1/\Delta M = 0.043/0.0516 = 83\ \%$). The same table shows also that the two hedonic weight vectors deliver two poverty measurements that are barely distinguishable from each other: the variation in the adjusted headcount ratio is a mere 0.1 %. Furthermore, 99.2 % of the households classified as poor with one set of weights are poor also according to the other weighting scheme. This finding, together with the 98.5 % correlation between the two scores, clearly shows that, in our sample, the poverty measurement based on the hedonic weights is not significantly affected by the presence of reporting styles heterogeneity in life satisfaction self-assessments.

6 Conclusions

Using multidimensional poverty measures instead of simple monetary poverty indicators is now a standard practice. The increase in the number and heterogeneity of the dimensions makes the weighting scheme a key ingredient of the poverty assessment. In this paper we carry out a multidimensional poverty assessment framed in the approach proposed by Alkire and Foster (2011a) to assess empirically to what extent the outcomes of a multidimensional poverty analysis depends on the weighting scheme adopted and whether the use of hedonic weights derived from self-assessed life satisfaction questions is robust to the presence of heterogeneity in response styles.

We draw data from the second wave of SHARE, a multi-country survey of the Europeans aged 50 or over, consider three dimensions (economic, housing and health) and compute the headcount and the adjusted headcount ratios using equal, frequency and hedonic weighting schemes. Two sets of hedonic weights are estimated, one by means of an ordered probit regression having respondents' life satisfaction self-assessments as dependent variable, the other using a Hopit model, a model which takes into account the variability of response styles across individuals by means of an anchoring vignette methodology (King et al. 2004).

Our results show that changes in the weighting scheme produces substantial differences in poverty assessment, both in terms of headcount and adjusted headcount ratios. Decomposing the variation of the adjusted headcount ratio, we see that the pools of households entering or exiting poverty when varying the weighting scheme can explain most of the differences in the measurement of poverty. Moreover, the contribution of each dimension to the overall poverty level changes widely across alternative weighting schemes, the same holds true for the contribution of different subgroups of households and for the evaluation of hypothetical policy interventions.

Although our empirical exercise confirms that the heterogeneity in response styles is an important issue in modeling life satisfaction self-assessments, it does not highlight significantly differences in neither the level nor the decomposition of the poverty index based on the hedonic weights. Overall, our results based on hedonic weights proved to be robust to the relaxation of the assumption that the reporting styles adopted by individuals in assessing life satisfaction are invariant in the population.

Our empirical results relate to a sample of Europeans aged 50 or over, and they are influenced by the framework adopted and by the achievements considered. Nevertheless, they clearly warn us that the choice of the weighting schemes is not innocuous for the outcomes of a multidimensional poverty analysis. Comparisons of poverty across groups and policy evaluations should then take into account this issue in order to provide meaningful and reliable conclusions.

Acknowledgments This paper uses data from SHARE wave 4 release 1.1.1, as of March 28th 2013 or SHARE wave 1 and 2 release 2.5.0, as of May 24th 2011 or SHARELIFE release 1, as of November 24th 2010. The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5-CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHARE-PREP, N° 211909, SHARE-LEAP, N° 227822 and SHARE M4, N° 261982). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions). We are grateful to the Editor and two anonymous referees for helpful comments and suggestions. We also thank the participants at the Italian SHARE Users' Conference (2012), 12th Journées Louis-André Gérard-Varet Conference in Public Economics (2013), SIS 2013 Statistical Conference, ASSET Annual Meeting (2013), International SHARE User Conference (2013), NETSPAR International Pension Workshop (2014), ESPE Annual Conference (2014) and at seminars at the University of Groningen and CEMFI.

Appendix

Table 10 ML estimates of ordered probit and Hopit models and sample mean of covariates

	Ordered probit model	Hopit model					
		main equation	/cutoff1	/cutoff2	/cutoff3	/cutoff4	Sample mean
per-capita net income	0.0876*	0.1662***	0.0358	0.0295	−0.0043	−0.0111	0.8399
per-capita net wealth	0.1797***	0.2480***	0.0963	−0.0505	0.0604	−0.0171	0.7078
dwelling accessibility	0.1518***	0.1143*	−0.1555*	0.0426	−0.0033	0.0775**	0.8465
chronic disease	0.2022***	0.1200***	−0.1327**	0.0467	−0.0123	−0.0125	0.4298
ADL	0.4043***	0.3591***	−0.1222	0.0262	0.0384	0.0111	0.8871
EURO-D	0.5138***	0.5145***	−0.1899***	0.0403	0.0702**	0.0806***	0.6967
SE	0.3422***	0.3828***	0.5189***	−0.0508	−0.1528**	−0.2432***	0.0772
DK	0.5840***	0.0268	−0.2729	−0.0260	−0.1010	−0.1474***	0.1717
NL	0.2203***	0.5862***	0.6714***	−0.2989***	0.0175	0.0311	0.0833

Table 10 (continued)

	Ordered probit model	Hopit model					
		main equation	/cutoff1	/cutoff2	/cutoff3	/cutoff4	Sample mean
BE	0.0184	−0.0292	0.5334***	−0.0797	−0.2115***	−0.2839***	0.0931
FR	−0.1265	0.2492*	0.7092***	−0.1762	0.0335	−0.0560	0.0637
GR	−0.5224***	−0.5964***	0.5848***	−0.1854	−0.0299	−0.6128***	0.0895
IT	−0.3165***	−0.0312	0.7414***	−0.1196	−0.2477***	0.0136	0.1068
ES	0.0061	0.0636	0.5812***	−0.0416	−0.4044***	−0.0927	0.0828
CZ	−0.2729***	−0.5580***	−0.0003	−0.1265	−0.0038	−0.0682	0.1473
male	−0.1078***	−0.0807*	0.0152	0.0034	0.0050	0.0107	0.4449
aged 55 or less	−0.2000***	−0.4066***	−0.1798	0.0188	−0.0290	−0.0366	0.2341
aged 56 − 60	−0.0541	−0.1977**	−0.0669	−0.0300	−0.0042	−0.0422	0.2076
aged 61 − 65	−0.0098	−0.0918	−0.2059*	0.0928	−0.0147	−0.0177	0.1729
aged 66 − 75	0.0011	0.0419	0.1722*	−0.0904	0.0254	−0.0433	0.2530
living with cohabiting	0.4119***	0.4279***	−0.0182	0.0111	−0.0242	0.0700**	0.7803
have children	0.0051	−0.0167	−0.0778	0.0934	−0.0384	−0.0491	0.9039
have grand children	0.0386	0.0737	−0.0023	−0.0018	0.0256	0.0178	0.6388
retired from work	0.2053***	0.2827***	−0.0393	0.0450	0.0257	0.0527	0.4956
employed or self-employed	0.2458***	0.4288***	−0.0913	0.0889	0.0581	0.1136***	0.3203
not involved in social activity	−0.2266***	−0.2536***	0.0989	−0.0808*	0.0443	−0.0643***	0.4923
low education	−0.0233	−0.0491	−0.1170	0.0525	−0.0016	0.0183	0.4855
middle education	−0.0059	−0.0110	−0.0245	−0.0232	0.0095	0.0473	0.2819
house owner	−0.0388	−0.0558	−0.0772	0.0194	−0.0261	0.0668**	0.7482
residing in city	0.0643	0.0739	0.0640	−0.0205	−0.0650*	0.0476	0.3234
residing in town	0.1142***	0.0600	−0.0401	0.0171	−0.0733**	0.0516*	0.4227
interviewed in winter	−0.0045	−0.0991	0.1904*	−0.1983	0.0284	0.0189	0.4074
interviewed in spring	0.0624	0.028	0.2265*	−0.1574**	0.0145	−0.0075	0.3627
interviewed in summer	−0.0202	−0.1626	0.1554	−0.2489*	0.0787	0.0346	0.0294
/cutoff1	−1.4208	–	–	–	–	–	
/cutoff2	−0.3970	–	–	–	–	–	
/cutoff3	0.6308	–	–	–	–	–	
/cutoff4	2.4940	–	–	–	–	–	
vignettes question 1	–	−0.4821***	–	–	–	–	
vignettes question 2	–	0.6794***	–	–	–	–	
Constant	–	–	−2.0703***	0.4395**	0.1647	0.4631***	
McFadden Pseudo R^2	0.1194				0.0711		

Note: Sample of 5545 individuals used to estimate the hedonic weights

References

Aaberge, R., Peluso, E.: A counting approach for measuring multidimensional deprivation. Discussion paper, Statistics Norway Research Department, 700 (2012)

Alkire, S., Foster, J.: Counting and multidimensional poverty measurement. J. Public Econ. **95**, 476–487 (2011a)

Alkire, S., Foster, J.: Understandings and misunderstandings of multidimensional poverty measurement. J. Econ. Inequal. **9**, 289–314 (2011b)

Angelini, V., Cavapozzi, D., Corazzini, L., Paccagnella, O.: Age, health and life satisfaction among older Europeans. Soc. Indic. Res. **105**, 293–308 (2012)

Angelini, V., Cavapozzi, D., Corazzini, L., Paccagnella, O.: Do Danes and Italians rate life satisfaction in the same way? Using vignettes to correct for individual-specific scale biases. Oxf. Bull. Econ. Stat. **76**(5), 643–666 (2014)

Blanchflower, D.G., Oswald, A.J.: Is well-being U-shaped over the life cycle? Soc. Sci. Medicine **66**, 1733–1749 (2008)

Börsch-Supan, A., Brugiavini, A., Jürges, H., Kapteyn, A., Mackenbach, J., Siegrist J., Weber, G.: First results from the Survey of Health, Ageing and Retirement in Europe (2004-2007). Starting the longitudinal dimension. Mannheim Research Institute for the Economics of Aging (MEA), Mannheim (2008)

Bourguignon, F., Chakravarty, S.R.: The measurement of multidimensional poverty. J. Econ. Inequal. **1**, 25–49 (2003)

Bosmans, K., Lauwers, L., Ooghe, E.: Prioritarian poverty comparisons with cardinal and ordinal attributes. Discussion paper, Center for Economic Studies. Katholieke Universiteit Leuven 13.10 (2013)

Bossert, W., Chakravarty, S.R., D'Ambrosio, C.: Multidimensional poverty and material deprivation with discrete data. Rev. Income Wealth **59**(1), 29–43 (2013)

Cavapozzi, D., Han, W., Miniaci, R.: Alternative weighting structures for multidimensional poverty assessment. University of Groningen, Research Institute SOM (Systems, Organisations and Management) Research Report 13018-EEF (2013)

Decancq, K., Fleurbaey, M., Maniquet, F.: Multidimensional poverty measurement with individual preferences. Discussion paper, pp. 058–2014. Princeton University William S. Dietrich II Economic Theory Center Working Paper (2014)

Decancq, K., Lugo, M.A.: Weights in multidimensional indices of well-being: An overview. Econom. Rev. **32**, 7–34 (2013)

Desai, M., Shah, A.: An econometric approach to the measurement of poverty. Oxf. Econ. Pap. **40**, 505–522 (1988)

Dolan, P., Peasgood, T., White, M.: Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. J. Econ. Psychol. **29**, 94–122 (2008)

Eurostat: Measuring material deprivation in the EU — Indicators for the whole population and child-specific indicators. Publications Office of the European Union, Luxembourg (2012)

Fleurbaey, M., Schokkaert, E., Decancq, K.: What good is happiness?. Université Catholique de Louvain, Center for Operations Research and Econometrics (CORE) Discussion Paper 2009/17 (2009)

Foster, J.E., Greer, J., Thorbecke, E.: A class of decomposable poverty measures. Econom. **52**, 761–766 (1984)

Frey, B.S., Stutzer, A.: What can economists learn from happiness research? J. Econ. Lit. **40**, 402–435 (2002)

Kapteyn, A., Smith, J.P., Van Soest, A.H.O.: Comparing life satisfaction. RAND Working Paper WR-623-1 (2009)

King, G., Murray, C., Salomon, J., Tandon, A.: Enhancing the validity and cross-cultural comparability of measurement in survey research. Am. Polit. Sci. Rev. **98**, 567–583 (2004)

Kingdon, G.G., Knight, J.: Subjective well-being poverty vs. income poverty and capabilities poverty? J. Devel. Stud. **42**(7), 1199–1224 (2006)

Nussbaum, M.C.: Women and Human Development: The Capabilities Approach. Cambridge University Press, Cambridge (2001)

Nussbaum, M.C.: Who is the happy warrior? Philosophy poses questions to psychology. J. Leg. Stud. **37**(S2), S81–S113 (2008)

Paruolo, P., Saltelli, A., Saisana, M.: Ratings and rankings: voodoo or science? J. Roy. Statistical Society A **176**(2), 1–26 (2013)

Prince, M.J., Reischies, F., Beekman, A.T., Fuhrer, R., Jonker, C., Kivela, S.L., Lawlor, B.A., Lobo, A., Magnusson, H., Fichter, M., van Oyen, H., Roelands, M., Skoog, I., Turrina, C., Copeland, J.R.: Development of the EURO-D scale - a European Union initiative to compare symptoms of depression in 14 European centres. Br. J. Psychiatry **174**, 330–338 (1999)
de Ree, J., Alessie, R.: Life satisfaction and age: Dealing with underidentification in age-period-cohort models. Soc. Sci. Medicine **73**, 177–182 (2011)
Rippin, N.: Considerations of efficiency and distributive justice in multidimensional poverty measurement. Doctoral Thesis. University of Göttingen (2013)
Ruger, J.P.: Health and Social Justice. Oxford University Press, New York (2010)
Schokkaert, E.: Capabilities and satisfaction with life. J. Hum. Dev. **8**, 415–430 (2007)
Sen, A.: Poverty: An ordinal approach to measurement. Econom. **44**, 219–31 (1976)
Sen, A.: Commodities and Capabilities. Amsterdam, North-Holland (1985)
Sen, A.: Capabilities, lists, and public reason: Continuing the conversation. Fem. Econ. **10**, 77–80 (2004)
United Nation Development Programme: Human Development Report 2011, Sustainability and Equity: A Better Future for All. United Nations Development Programme, New York (2011)