CrossMark

RESEARCH PAPER

Data Versus Survey-based Normalisation in a Multidimensional Analysis of Social Inclusion

Ludovico Carrino[1,2]

Abstract In the context of the multidimensional measurement of complex phenomena, the major focus of the recent literature has been on the choice of the dimensions' weights and the shape of the aggregation function, while few studies have concentrated on how normalisation influences the results. With the aim of building a measure of Social Inclusion for 63 European regions between 2004 and 2012, we adopt a standard linear aggregation framework and compare two alternative normalisation approaches: a *data-driven* min–max function, whose parameters depend solely on the available data, and an *expert-based* function, whose parameters are elicited through a survey at the University of Venice Ca' Foscari. Regardless of the adopted strategy, we show that normalisation plays a crucial part in defining variables' weighting. The *data-driven* strategy allocates a large relative weight to the longevity dimension, whereas the *survey-driven* results in a rather equal distribution of weights. The *data-driven* approach produces trade-offs that are hard to interpret in economic terms and debatable from a social desirability perspective, thus constituting a *positive* analysis of Social Inclusion. Moreover, it softens the aftermaths of the recent economic crisis on

✉ Ludovico Carrino
  ludovico@unive.it

[1] Economics Department, Ca' Foscari University of Venice, Venice, Italy

[2] Department of Economics, Ca' Foscari University of Venice, Cannaregio 873, Fondamenta S.Giobbe, 30121 Venice, Italy

Social Inclusion, by putting a consistent weight on the longevity variable. Conversely, the *expert-based* normalisation is heavily affected by elicitation techniques, and allows for a *normative* interpretation of the resulting index. Furthermore, it emphasizes the worsening trends in long-term unemployment and the relevance of early school leaving in the Social Inclusion measure. The two strategies lead to substantially different conclusions in terms of levels (both *between* and *within countries*) and distribution of Inclusion: numerous rank-reversals occur when switching the normalisation methods.

Keywords CES · Normalisation · Aggregation · Weighting · Experts · Multidimensionality · Social inclusion

JEL Classification C43 · D63 · I32

1 Introduction

Although there is quite a consensus on the need for broadening the scope of the analysis of Well-being beyond the monetary dimension (see, e.g., the influential report by Stiglitz et al. 2010), there is not equal agreement on how such an ambitious task should be operationalized. It is well known that subjectivity and arbitrariness exist with respect to the choice of the dimensions to be included in the composite index, the normalisation of the variables, and the characterisation of the aggregation function (see, e.g., Ravallion 2012a; Decancq and Lugo 2013; Martinetti and von Jacobi 2012).[1] The socio-economic literature highlighted that no unanimous method exists to perform such choices, pointing out numerous theoretical issues (Stiglitz et al. 2010; Ravallion 2011, 2012a; Klugman et al. 2011; Maggino and Nuvolati 2012; Decancq and Lugo 2013), testing empirical robustness (Kasparian and Rolland 2012; Lefebvre et al. 2010; Saisana et al. 2005; Ravallion 2012b). Yet, although there may be no "absolute cure" for multidimensional evaluations, a good practice could consist in enhancing methodological transparency (Sen and Anand 1997).

While the major focus of the recent literature has been devoted to the choice of the dimensions' weights, few studies have concentrated on the role played by normalisation in influencing the final results (Lefebvre et al. 2010; Saisana et al. 2005). Our contribution highlights that, in fact, normalisation is a crucial stage where an "early" implicit weighting takes place, which can strongly affect the overall results of the multidimensional analysis. We show that, since no golden rules exist on how a normalisation function should be selected and characterised, different strategies, all

---

[1] The act of synthesizing a composite latent phenomenon encompasses methodological issues that have economic, philosophical (as well as psychological) and political connotations. Indeed, these issues arise from a fundamental mismatch between the kind of multiplicity inherent in the latent concept and the multiplicity characterizing the forged measure (the result of the researcher's work). In a sense, the latent multidimensional concept (e.g., Well-being or Social Inclusion) is an un-synthesized multiplicity, in that it is composite by nature and perceived as a whole by the human sensibility. Since the phenomenon is unmeasurable per se, the researcher is forced to separate it, operationally, in numerous measurable components, in order to aggregate them back to provide a proxy of the latent phenomenon. In other words, building a synthetic index of Well-being requires that the indeterminate nature of multiplicity is made determinate through a specification of its contents, and of their relationship.

acceptable *a-priori*, can lead to very different weighting structures and, ultimately, opposite results for the composite indicator. Therefore, the unavoidable arbitrariness regarding the choice of the normalisation function, as well as its methodological justification, should be made as transparent as possible.

To illustrate these points, in this paper we will build a composite measure of Social Inclusion for 63 European administrative regions from 2004 to 2012, using data from EUROSTAT. The aggregation framework is a CES function (constant elasticity of substitution), and the selection of variables follows the relevant literature on this topic (stemming from Atkinson et al. 2002).[2]

In this analysis, we adopt a baseline linear aggregation model where the normalised components have equal weights and we look at what happens to the aggregate measure of Social Inclusion when the sole normalisation function changes. In particular, we apply the widely used *data-driven* min–max normalisation strategy, whose parameters depend on the available data. This data-driven function generates implicit trade-offs (between the index' components) and shadow prices with weak economic justification. We also propose a novel strategy, an *expert-based* min–max function, whose parameters are grounded on the responses to a survey conducted on a population of 150 professors of Economics or Management at the Ca' Foscari University of Venice.

Our results indicate that, even within a simple-average framework, changing the normalisation function substantially affects the relative relevance of each component of the aggregate measure. As a consequence, significant differences emerge in the levels and rankings of regional Social Inclusion in Europe, leading to very different policy implications. The data-driven strategy softens the heterogeneities *within* and *between* European countries by putting a substantial weight on the longevity variable rather than on educational and economic statuses. As a result, the European regional distribution of Inclusion appears to be uni-modal around the mean. Conversely, the expert-based normalisation emphasises the unemployment and the school-dropouts variables, and returns a bi-modal distribution of Social Inclusion. We, thus, discuss how the different premises of the two strategies characterise the interpretation of the results: the data-driven approach allows for a positive interpretation of the index, while the survey-driven approach allows for a normative one. In other words, if the index' intrinsic trade-offs are grounded on statistical terms, its results should be interpreted accordingly.

The remaining of the paper is organized as follows. Section 2 briefly describes the concept of Social Inclusion and the data. Section 3 sets a standard framework for multidimensional aggregation and details the baseline model. Section 4 introduces the normalisation strategies, while Sect. 5 discuss the implicit trade-offs resulting from applying the aforementioned normalisation functions on the baseline model. Section 6 details the results of the Social Inclusion indices, Sect. 6.1 concludes.

---

[2] Through this analysis, we do not aim at providing efficiency index for the Welfare States, which would require a much more structured set of information. We, rather, limit ourselves at evaluations of performances, as suggested by Pestieau (2009), Lefebvre et al. (2010).

2 Social Inclusion, Definition and Sample Selection

Social inclusion (as its corresponding opposite, social exclusion)[3] is one of the five priorities selected by the European Commission in the context of the Europe 2020 Strategy. A definition of Exclusion was already drawn in December 1992 by the Commission of the European Communities (European Communities Commission 1992): "Social exclusion is a multidimensional phenomenon stemming from inadequacies or weaknesses in the services offered and policies pursued in these various policy areas. Such insufficiencies and weaknesses often combine to affect both people and regions via cumulative and interdependent processes of such a nature that it would be futile to try to combat exclusion by tackling only one of its dimensions. More clearly than the concept of poverty, (…) it states out the multidimensional nature of the mechanisms whereby individuals and groups are excluded from taking part in the social exchanges, from the component practices and rights of social integration and of identity".

The Laeken European Council in 2001 has developed a set of unanimously agreed indicators that could capture the multifaceted aspects and outcomes of Social Inclusion, thus providing reliable and comparable data to monitor the social and economic conditions of European citizens (European Council 2001), through the Open Method of Coordination. In particular, four basic dimensions have been identified: the level and distribution of income, the performance in the labour-market, education and health areas. For each of them, a set of primary indicators were adopted: income (Poverty rate (after social transfers), Persistent risk-of-poverty rate, Relative median at risk-of-poverty gap, Inequality of income distribution); labour market (Long-term unemployment, Regional cohesion, Persons living in jobless households); education (early school leavers); health (life expectancy at birth, Self-defined health status by income level).[4]

The target of this paper is to build an aggregate index of social exclusion at administrative-regional level in Europe. We choose administrative regions as the main territorial unit of this analysis, with the aim of capturing higher variability than it can be inferred from aggregate national data. Data-availability is often mentioned as a serious constraint for analyses which focus on a wide set of countries for a long time-period (Lefebvre et al. 2010; Martinetti and von Jacobi 2012). In the context of social exclusion at administrative regional level, we are able to gather data for four out of the 10 aforementioned indicators, one per dimension: poverty-rate, long-term unemployment, early school-leavers and life expectancy at birth. Our data-source is the on-line

---

[3] The concept of social inclusion/exclusion should not be confused with the variable 'at risk of poverty or social exclusion' in the Eurostat database, which defines an individual as at risk of poverty or social exclusion when at least one of the following conditions hold: (a) equivalent household income below 60 % of national median; (b) households with at least 4 of the following 9 issues: (i) impossibility to bear unexpected expenses, (ii) cannot afford a week holiday, (iii) issues with the mortgage, rent, bills; (iv) cannot afford a proper meal every two days; (v) not able to adequately heat the house; (vi) not able to afford a washing machine, (vii) a colour TV, (viii) a phone, (ix) an automobile; (c) living in families whose members aged 18-59 work less than a fifth of their time.

[4] We refer to (Atkinson et al. 2002, 2004), as well as to European Commission (2009) and Eurostat (2010) for further details on the rationale of Social Inclusion indicators and on the issues related to their measurement.

Table 1 Variable definitions

Variable	Definition
Poverty rate	Share of individuals living in households with an income below 60 % national median equivalised disposable income
Long-term unemployment rate	Total long-term unemployed population (≥ 12 months; ILO definition) as proportion of total active population
Early school-leavers	Share of total population of 18-24-year olds having achieved ISCED level 2 or less and not attending education or training
Life expectancy at birth	Number of years a person may be expected to live, starting at age 0

Eurostat Regional Database 2015, and the longest data interval available for all the four variables spans from year 2004 to 2012, for 63 administrative regions in five countries (Belgium, Denmark, Germany, Italy and Spain), even though data for Denmark are not available for 2006 and earlier. For other countries (e.g., Greece, France, Czech Republic and Norway), data were either not available for all the indicators, or they are available for statistical-regions, but not for administrative regions.[5]

As argued in Lefebvre et al. (2010), "these indicators cover the most relevant concerns of a modern welfare state, also reflecting aspects that people who want to enlarge the concept of GDP to better measure social welfare generally take into account". The latter referenced paper discusses, as do Atkinson et al. (2004), the limitations of these data and the necessary simplifying assumptions that have to be done when translating a complex multidimensional phenomenon like social exclusion in empirical terms. Table 1 provides a brief definition for our four variables:

The following table and figure report descriptive statistics on the four indicators in our sample (Table 2).

Appendix A includes further descriptive statistics on correlations and data-distribution of the selected variables in the sample-data.

3 Aggregation Framework

Let us consider *m dimensions* (hereinafter also *variables*, *attributes, components*) of Social Inclusion, observed for n regions. For a generic region i we can therefore build the vector $\mathbf{x}^i = (x_1^i, ..., x_m^i)$, while $\mathbf{X} \in \mathbb{R}^{n \times m}$ is the distribution matrix of m attributes for n regions. To retrieve an aggregated measure for region i, we consider the function F defined as:

$$F^i\left(v\left(\mathbf{x}^i\right)\right) = \left[w_1 v_1\left(x_1^i\right)^\beta + \cdots + w_m v_m\left(x_m^i\right)^\beta\right]^{1/\beta} \tag{1}$$

---

[5] As a robustness check we enlarged the sample with data for statistical-regions for Czech Republic, Greece, Norway and the Netherlands, without any significant change to the results of the analysis. Besides, as stated in the introduction, the purpose of this paper is to offer a methodological discussion that can be applied to composite analyses in various fields and from various data-selections.

Table 2 Descriptive statistics for four indicators of social inclusion, years 2004–2012

	Belgium	Denmark	Germany	Italy	Spain
Longevity (in years)					
Min	77.5 *Wallonie '04*	77.6 *Sjælland '07*	77.8 *Sachsen-A. 04*	79.5 *Campania '05*	78.3 *Ceuta '04*
National mean	79.8	79.2	80.2	82	81.4
Max	81.6 *Vlaanderen '11*	80.7 *Midtjylland '11*	82.2 *Baden-W. '12*	83.8 *Marche '10*	84.2 *Madrid '12*
Early school leaving (in %)					
Min	8.6 *Vlaanderen '08*	8.5 *Midtjylland '12*	5.4 *Thüringen '09*	10.2 *Molise '12*	11.5 *P. Vasco '12*
National mean	12.1	11.1	12.3	18.4	29.3
Max	20 *Brussels '07*	16.7 *Sjælland '09*	21 *Saarland '06*	32.7 *Sardegna '05*	54.2 *Ceuta '05*
Long-term unemployment (in %)					
Min	1.4 *Vlaanderen '08*	0.3 *Midtjylland '08*	1.1 *Bayern '12*	0.5 *TAA '04*	0.9 *Aragón '08*
National mean	3.8 %	1.2	4.1 %	3.6 %	4.8 %
Max	9.9 *Brussels '06*	2.4 *Syddanmark '04*	13.8 *Sachsen-A. '04*	12.2 *Campania '12*	18.2 *Ceuta '12*
At-risk-of-poverty rate (in %)					
Min	9.8 *Vlaanderen '11*	8.7 *Sjælland '07*	10 *Baden-W. '07*	5.2 *VDA '06*	8.1 P. *Vasco '09*
National mean	14.8	12.6	14.5	16.9	20.8
Max	33.7 *Brussels '11*	16 *Hovedstaden '10*	24.3 *Mecklem. '07*	44.3 *Sicilia '11*	48.9 *Ceuta '08*

Eurostat Regional Database 2016. Country labels correspond the ISO 3166-1 alpha-2 standard
VDA Valle d'Aosta; *TAA* Trentino–Alto Adige

which is often referred to as a constant elasticity of substitution (CES) function, or a generalized mean of order β. Its arguments are the elements $v_1, \ldots, v_m$ which are transformations of the original variables $x_1, \ldots, x_m$ (defined hereafter). The function F is non-decreasing, separable, weakly scale-invariant and homogenous of degree-one in its arguments v; we refer to Blackorby and Donaldson (1982) and (Decancq and Lugo 2009, 2008) for an analytic characterization of these properties.

The parameters $w_1, \ldots, w_m$, the weights of the normalised dimensions v, are non-negative and sum to one.

Provided that a choice of the m dimension has been performed, the main methodological task is now the selection of the set of functions $v_1, \ldots, v_m$, of parameters $w_1, \ldots w_m$, as well as of β.

3.1 Baseline Linear Aggregation Model

The parameter β in (1) determines the elasticity of substitution $\varepsilon_{k,j}$ between any pair (v_k, v_j). In the CES function, the elasticity between any pair k, j is constant and equal to $1/1 - \beta$. The elasticity of substitution determines the percentage change in v_j/v_k, which would result from a percentage change in the slope along a level-set (the marginal rate of substitution, MRS, along an indifference curve). The parameter β must be lower than one to generate iso-inclusion contours convex to the origin in the two-dimensional region of the space of attributes (Bourguignon and Chakravarty 2003). The smaller is β, the higher is the increase in dimension v_j needed to keep constant the overall index after a one-unit decrease in dimension v_k.

Since the focus of this paper is on the normalisation choices, let us adopt a standard aggregation framework by setting $\beta = 1$ in (1), therefore obtaining a linear weighted average with linear indifference curves, constant MRS and infinite elasticity of substitution between pairs of normalised dimensions. We also let the weights w_j be equal, i.e., $w_1 = \cdots = w_m = 1/m = 1/4$ (since $m = 4$ in our case study). The resulting model will be, for a generic region i (time subscripts are omitted) an aggregation function L, as in *linear*, such as:

$$L^i \left(v \left(\mathbf{x}^i \right) \right) = \frac{1}{m} v_1 \left(x_1^i \right) + \cdots + \frac{1}{m} v_m \left(x_m^i \right) \tag{2}$$

The arbitrary choice of setting equal weights is a widely adopted strategy in the literature of multidimensional measurement. As Hoskins and Mascherini (2009) and Decancq and Lugo (2013) highlight, this approach is often justified with the argument that all the dimensions are equally important (Atkinson et al. 2002) or, conversely, that there is insufficient knowledge for setting a more detailed weighting scheme (sometimes referred to as an "agnostic view"). Although being frequently described as a *simple* and relatively *neutral* strategy, "equal weighting" does not mean "no weighting", because it involves an implicit judgment on the weights being equal, and because it often applies just on the normalised dimensions of the index.[6]

---

[6] In the words of Martinetti and von Jacobi (2012), the implicit assumption for equal weighting is that "in absence of any objective mechanism for determining the relative importance of the considered dimensions,

In the following Sections, we will investigate how original attributes contribute to the overall measure, and what characterizes the relationship between attributes within the linear framework. In general, such effects can be retrieved via the partial derivative of the aggregate measure L with respect to variable x_j (region-specific indices are dropped for convenience), as follows:

$$\frac{\partial L\left(v\left(\mathbf{x}\right)\right)}{\partial x_j} = w_j v'_j\left(x_j\right) \tag{3}$$

From (3) we can identify two main drivers that determine how the aggregate measure L reacts at small changes in the j-th real-valued dimension x_j. First, the higher is the weight of the normalised j-th dimension, the higher will be the marginal variation in the L. Second, the steeper is the normalisation function, the higher will be the effect of a change in the j-th dimension on the aggregate measure.

Within the linear aggregation function L, the marginal rate of substitution between a pair of observed-indicators x_j and x_k is:

$$MRS_{x_k,x_j} = -\frac{dx_j}{dx_k} = \frac{\frac{\partial L(v(\mathbf{x}))}{\partial x_k}}{\frac{\partial L(v(\mathbf{x}))}{\partial x_j}} = \frac{w_k}{w_j}\frac{v'_k\left(x_k\right)}{v'_j\left(x_j\right)} \tag{4}$$

Both the marginal contribution of the j-th attribute and its MRS depend on the shape of the normalisation function v_j. If, however, the transformation function is the identity function ($v_j(x_j) = x_j$), the effect of a change in x_j can be uniquely determined by its weight w_j, while the MRS between a pair of dimensions j and k is determined by the ratio between their weights.[7]

4 Normalisation Framework

Raw variables are usually observed and measured with different measurement units. The component $v_i(x_j)$ is a weakly monotonic and continuous normalisation function that maps the values of the j-th variable x_j on the closed interval [0, 100], i.e., $v_j(x_j) \in$ [0, 100]. Moreover, attributes might be, alternatively, positively or negatively related to the latent phenomenon, i.e., they may have a positive or negative *polarity*. Hence, in order to ensure comparability and monotonicity of any aggregation function, each variable must be normalised such that better performances in the *j*-th dimension correspond to non-lower values of $v_j(x_j)$ and therefore of the aggregated value L. In other words, each normalised variable should have a positive polarity. The normalisation function thus ensures that L is bounded between 0 and 100 when the weights w sum to one. In what follows we will briefly present two normalisation strategies, and we refer

---

Footnote 6 continued

the most neutral method is assigning an equal weight to each of them". Indeed, both Chowdhury and Squire (2006) and Nguefack-Tsague et al. (2011) provide evidence in favour of equal weighting after collecting expert preferences.

[7] The MRS between two observed dimensions will be equal to their "weights" also if the derivatives of their normalisation functions are equal, i.e., if $v'_k\left(x_k\right)/v'_j\left(x_j\right) = 1$.

the reader to Giovannini et al. (2008) for their detailed description. The first one is "data-driven", that is, a transformation whose characteristics are entirely determined by the data at hand. The second one, conversely, is defined through the elicitation of explicit value judgements.

4.1 Data-Driven Normalisation Function

The *min–max normalisation function* is widely used in the literature of multidimensional measures (see, e.g., Cherchye et al. 2007; Silva and Ferreira-Lopes 2013; Pinar et al. 2014; Mazziotta and Pareto 2015), as well as in the Human Development Index (Anand and Sen 1994) and in the OECD Better Life (Boarini and D'Ercole 2013).

For two given variables x_+ (with positive polarity) and x_- (with negative polarity) observed in region i at a time t, the corresponding generic *min–max* normalised values v_+ and v_- are defined as follows:

$$v^{i,t}_{MM+}\left(x^{i,t}_+\right) = 100 * \frac{x^{i,t} - b_+\min(x_+)}{b_+\max(x_+) - b_+\min(x_+)}$$

$$v^{i,t}_-\left(x^{i,t}\right) = 100 * \frac{b_-\max(x) - x^{i,t}}{b_-\max(x) - b_-\min(x^{i,t})}$$

$$v^{i,t}_{MM+}\left(x^{i,t}_+\right) = 0 \quad \text{if } x^{i,t}_+ \le b_+\min(x_+) \quad v^{i,t}_-\left(x^{i,t}_-\right) = 0 \quad \text{if } x^{i,t}_- \ge b_-\max(x_-)$$

$$v^{i,t}_{MM+}\left(x^{i,t}_+\right) = 100 \text{ if } x^{i,t}_+ \ge b_+\max(x_+) \quad v^{i,t}_-\left(x^{i,t}_-\right) = 100 \text{ if } x^{i,t}_- \le b_-\min(x_-)$$

(5)

In the generic *min–max* function, the coefficients $b_\pm min_i$ and $b_\pm max_i$ are the highest and lowest values to be used as benchmarks for the x_+ variable for region i. Regardless on how the benchmarks are defined, it is straightforward that, for x_+, b_+max corresponds to a more desirable performance than b_+min, while the opposite is true for x_-. The *min–max* strategy rescales indicators into an identical range [0, 100].[8] E.g., for x_+, 0 is given to values lower or equal to b_+min, while 100 is given to those higher or equal to b_+max. The values within these benchmarks are proportionally converted into the 0–100 scale. Hence, v is a stepwise continuous function.

The data-driven min–max normalisation (6) defines the benchmarks *min* and *max* as the best and worst observed performance among selected regions (Lefebvre et al. 2010; Silva and Ferreira-Lopes 2013; Murias et al. 2012), and across a time-series, in order to take into account the evolution of indicators and offer time-comparability (Giovannini et al. 2008). In our case study, this corresponds to assigning a value of 0 to the region which reports the worst-observed performance in the period from 2004 to 2012, while assigning a value of 100 to the "best-observed" one.

For each region i where an attribute x is observed at a time t, the corresponding normalised value $v^{i,t}_{dM}, (x^i, t)$, where the subscript dM stands for "data-driven min–max", is determined as:

---

[8] The choice of multiplying by 100 eases readability of the results in the remaining of the paper, and does not affect any result.

Table 3 Data-driven benchmarks

	Observed minimum	Observed maximum
Longevity (good)	77.5 (Région wallonne)	84.2 (Comunidad de Madrid 2012)
Early school leaving (bad)	5.4 % (Thüringen 2009)	42.8 % (Murcia 2004)
Long-term unemployment (bad)	0.3 % (Midtjylland 2008)	15.3 % (Canarias 2012)
At-risk-of-poverty rate (bad)	5.2 % (Valle d'Aosta 2006)	44.3 % (Sicilia 2011)

Data from the Eurostat Database 2016 (2004–2012). Data for autonomous cities of Ceuta and Medilla (Spain) are excluded (see Footnote 10)

$$v_{dM+}^{i,t}\left(x_{+}^{i,t}\right) = 100 * \frac{x_{+}^{i,t} - \min_{t \in T} \min_{i}\left(x_{+}^{t}\right)}{\max_{t \in T} \min_{i}\left(x_{+}^{t}\right) - \min_{t \in T} \min_{i}\left(x_{+}^{t}\right)}$$

$$\text{or } v_{dM-}^{i,t}\left(x_{-}^{i,t}\right) = 100 * \frac{\max_{t \in T} \max_{i}\left(x_{-}^{t}\right) - x_{-}^{i,t}}{\max_{t \in T} \max_{i}\left(x_{-}^{t}\right) - \min_{t \in T} \min_{i}\left(x_{-}^{t}\right)} \qquad (6)$$

where x$_+$ and x$_-$ have the usual meaning of a "good" and a "bad" attribute, respectively.

Table 3 displays the data-driven thresholds for our sample of regions.[9]

4.2 A Novel Expert-Based Normalisation Function

Consistently with what is often debated with respect to the aggregation function, the parameters of the normalisation function can either reflect a predetermined choice by the researcher herself (e.g., through a *data-driven* strategy), or be elicited from some stakeholders group, e.g., field-experts, members of institutions, citizens (Kim et al. 2015; Decancq and Lugo 2013 produce a recent review of elicitation strategies).

In a simple linear model with $\beta = 1$ and $w_j = 1/m$ for each *j*-th indicators, the crucial determinant of a dimension's relevance relies heavily on the normalisation function, as visible from (3). As we will discuss in the next section, if the function's parameters are data-driven, then their implications in terms of dimensions' weights and MRS are to be interpreted under a mathematical perspective, yet it is harder to determine what do they reflect in economic terms (Lefebvre et al. 2010). As an example, in the data-driven min–max, a variable with transformed-value equal to "0" just implies it being "the last one", or "the worst one" observed among the available

---

[9] The autonomous cities of Ceuta and Medilla, located on the Mediterranean coast of Morocco but belonging to Spain since 15 century, are substantially different from other Spanish regions. Given that their values for school-dropouts, long-term unemployment and poverty rate are sensibly higher than the rest of the sample, we prudently decided to treat them as outliers and exclude them from the computation of the thresholds. This decision has no significant consequences on the results of the paper, nor on its implications. Including them in the sample would raise the maximum values for early school-leaving rate to 54.2 % (Ceuta 2005), for long-term unemployment to 18.2 % (Ceuta 2012), and for poverty rate to 48.9 % (Ceuta 2008). A graphical distribution of the data used for the min–max normalisation is reported in Fig. 8 (Appendix A).

data, which does not necessarily corresponds to an undesirable condition of Well-being. A similar reasoning, with opposite meaning, can be done for normalised values of "100".

An alternative to the data-driven normalisation would require to incorporate some value judgments in the normalisation (e.g., goalposts, see Mazziotta and Pareto 2015). This translates to linking the extreme values "0" and the "100" with, e.g., a certain definition of desirability, thus making the normalisation independent from the data. When an indicator lies above or below such fixed bounds, further variations do not contribute to the latent variable under study (see e.g., the discussion in Anand and Sen 1994; Klugman et al. 2011; Ravallion 2012b; Lefebvre et al. 2010; Gidwitz et al. 2010). A major example of fixed threshold is the Human Development Index that, since 1994, adopted "goalposts" as minimum and maximum values in the normalisation function. The interpretation behind these fixed thresholds relies on the belief that objective upper and lower bounds can be identified and defined as "subsistence" minimum or "satiation" points, beyond which additional increments would not contribute to the expansion of capabilities.

Contrary to Human Development, social exclusion's concept has been developed with reference to advanced industrialized economies, as are those of the European Union members. Therefore, rather than on "subsistence", its focus is posed on the "unacceptability" and "undesirability" of living conditions, as in an enlarged definition of poverty. Accordingly, a positive threshold for each of our four social-inclusion attributes would refer to a "certainly desirable and favourable condition of Well-being", to which a normalised value of 100 would correspond. Conversely, a negative threshold would refer to a "certainly undesirable and harmful conditions of Well-being", corresponding to a normalised value of 0.

In order to select the actual thresholds, we chose to elicit expert preferences through a survey, rather than to pre-determine them in a top-down fashion. To the best of our knowledge, this is a strategy rarely applied to normalisation stage, and mostly adopted for the aggregation phase instead.

Following Chowdhury and Squire (2006) and Hoskins and Mascherini (2009) (who, indeed, both elicit weights on aggregation rather than on normalisation), we intended to involve informed opinions and therefore selected the population of professors and researchers in the Departments of Economic and Management of the Ca' Foscari University of Venice. Specifically, our population consisted of 149 professors (57 + 38 full or associate professors of Economics and Management, respectively; 29 + 25 assistant professors, *ricercatore universitario*, of Economics and Management, respectively).[10] As for any expert sample, issues could be raised on our group's capability of ensuring all values of efficiency, equity and democracy in the elicitation process. As Kim et al. (2015) pointed out, there is no elicitation method that can ensure all the aforementioned problems. Moreover, being concerned with democratic representativeness, one

---

[10] Although, in principle, it would be of interest to widen the Survey population to professors of other Departments (Asian and North African Studies, Environmental Sciences, Humanities, Linguistic, Molecular Sciences and Philosophy), we were led by time and resources constraints to focus on those Faculty more specifically connected to the issues of social inclusion and to the disciplines related to the four indicators over which a judgment was asked.

could argue that greater citizen participation were required; nevertheless, such strategy would likely cause loss of efficiency and quality of the elicitation, together with a lower degree of representativeness (given the resources' constraints). Conversely, we selected a narrow population with specific characteristics but with a working experience that is, at least, partially related with the issues involved in social inclusion. Moreover, thanks to an adequate response rate, we are able to statistically represent it.

The survey was worded in Italian and conducted in electronic-form with the QUALTRICS software, a web-based tool that enables users to build custom surveys and distribute them via email.[11] Participants were invited with an email including a link to take part to the on-line questionnaire on an anonymous basis. Appendix B includes further details on the survey pages and wording.

4.2.1 The Expert-Based Thresholds and the Min–Max Normalisation Function

We implement a min–max normalisation as in (5), where the benchmarks corresponds to the median values elicited through the Qualtrics survey (Table 4). In particular, the favourable threshold for life expectancy is chosen at 83 years old, while the negative threshold is 73 years old. Early school-leaving's range lies between 10 % (which corresponds to the EUROPE 2020's target for members of the European Union). A rate of 9 % (or higher) of long-term unemployment denotes a median certainly undesirable condition, while the positive threshold is determined at 3 %. As for poverty rate, a certainly harmful level has its median value at 20 %, while desirability corresponds to 5 % (or lower) share of population below the poverty line set by the Eurostat.

The interquartile ranges are always relatively small, except for the negative threshold for early-school-leaving (15–25 %). Nevertheless, we are aware that no "true values" exist, with respect to these thresholds. In the words of Mascherini and Hoskins (2008), "the judgment of one of the outline may be correct, and those who share a consensus view may be wrong".

A quick comparison of Tables 3 and 4 suggests that "certain desirability" and "certain undesirability" largely differ from observed minimum or maximum achievements. Indeed, the lowest observed level of longevity (77.4 years) is considered to be "certainly undesirable" just by a small fraction of respondents (Fig. 10 in Appendix A). Similarly, any rate of long-term unemployment beyond 9 %, or of school dropouts higher than 20 %, or of poverty-rate beyond 20 %, is regarded as certainly negative, while the actual observed maximums are quite higher. A capping on the positive threshold occurs for those regions which report long-term unemployment lower than 3 % or early school leaving rates lower than 10 %.[12]

For each region i where an attribute x is observed at a time t, the corresponding normalised value $v_{sM}^{i,t}(x^{i,t})$, where the subscript sM stands for "survey-driven min–max", is determined as:

---

[11] For further details, please refer to http://www.qualtrics.com/.

[12] No territories in our sample reach 5 % poverty-rate or 73 years in longevity-at-birth, so no "positive" capping occurs.

Table 4 Survey-elicited benchmarks

	Median elicited minimum	Median elicited maximum
Longevity	73 years	83 years
Early school leaving (%)	10	20
Long-term unemployment (%)	3	9
At-risk-of-poverty rate (%)	5	20

Details on the survey's results are available in Appendix A

$$v_{sM+}^{i,t}\left(x_+^{i,t}\right) = 100 * \frac{x_+^{i,t} \text{ lowest threshold}}{\text{thresholds'range}} \quad \text{or}$$

$$v_{sM-}^{i,t}\left(x_-^{i,t}\right) = 100 * \frac{\text{highest threshold} - x_-^{i,t}}{\text{thresholds'range}} \quad \text{with}$$

$$v_{sM+}^{i,t}\left(x_+^{i,t}\right) = 0 \quad \text{if } x_+^{i,t} < \text{lowest thres hold} \quad \text{and}$$

$$v_{sM+}^{i,t}\left(x_+^{i,t}\right) = 100 \quad \text{if } x_+^{i,t} > \text{highest threshold}$$

$$v_{sM-}^{i,t}\left(x_-^{i,t}\right) = 0 \quad \text{if } x_-^{i,t} > \text{highest thres hold} \quad \text{and}$$

$$v_{sM-}^{i,t}\left(x_-^{i,t}\right) = 100 \quad \text{if } x_-^{i,t} < \text{lowest threshold} \tag{7}$$

where x_+ and x_- have the usual meaning of a "good" and a "bad" attribute, respectively.

The expert-based normalisation is closer in nature to a "social value-functions", in that the rescale is performed according to how much a value fulfils a "desirability" requirement. Moreover, the normalisation function may become weakly monotonic (instead of being *strongly monotonic* as the data-driven min–max), when the elicited constraints are binding for some observed variable. Indeed, there are regions having attributes with observed performances outside the elicited boundaries, which will receive a normalised value of 100 or 0. Therefore, as we will discuss in the next section, when an attribute's value lies outside the thresholds, its marginal contribution to the aggregate measure is zero.[13]

5 Implicit Trade-Offs from Normalisation

When implementing the min–max normalisation, both in the data-driven (6) and in the survey-driven (7) setup in the baseline linear model with "equal weighting" (2) we obtain two aggregation functions: LD (linear, data-driven min–max), LS (linear, survey-driven min–max). Normalisation benchmarks for the LD and LS models are taken from Tables 3 and 4, respectively. For a generic region i, such aggregation functions take the followingform:

---

[13] The min–max normalisation function can be smoothed, in order to avoid the step-wise shape (see, e.g., the discussion in Ravallion 2012b; Lefebvre et al. 2010; Martinetti and von Jacobi 2012; Meyer and Ponthière 2011; Pinar et al. 2014).

$$LD^i\left(v_{dM}\left(\mathbf{x}^i\right)\right) = 100\left(0.25 * \frac{x_1 - 77.5}{84.2 - 77.5} + 0.25 * \frac{42.8 - x_2}{42.8 - 5.4}\right.$$

$$\left. +0.25 * \frac{15.3 - x_3}{15.3 - 0.3} + 0.25 * \frac{44.3 - x_4}{44.3 - 5.2}\right) \tag{8}$$

$$LS^i\left(v_{sM}\left(\mathbf{x}^i\right)\right) = 100\left(0.25 * \frac{x_1 - 73}{83 - 73} + 0.25 * \frac{20 - x_2}{20 - 10}\right.$$

$$\left. +0.25 * \frac{9 - x_3}{9 - 3} + 0.25 * \frac{20 - x_4}{20 - 5}\right) \quad \text{with}$$

$$v_{sM}(x_1) = 0 \quad \text{if } x_1 < 73 \quad \text{and} \quad v_{sM}(x_1) = 100 \quad \text{if } x_1 > 83$$

$$v_{sM}(x_2) = 0 \quad \text{if } x_2 > 20 \quad \text{and} \quad v_{sM}(x_2) = 100 \quad \text{if } x_2 < 10$$

$$v_{sM}(x_3) = 0 \quad \text{if } x_3 > 9 \quad \text{and} \quad v_{sM}(x_3) = 100 \quad \text{if } x_3 < 3$$

$$v_{sM}(x_4) = 0 \quad \text{if } x_4 > 20 \quad \text{and} \quad v_{sM}(x_4) = 100 \quad \text{if } x_4 < 5. \tag{9}$$

Before implementing such models on the sample data, it is useful to highlight the implicit economic and statistical mechanisms acting beyond these aggregation functions, through the normalisation stage. The most direct way to do it is to investigate the "relative importance" (marginal contribution) that each dimension is given in each of the three aforementioned models of social inclusion. Indeed, the aggregation function is kept constant and it is characterised by *equal* weighting to the normalised dimensions. Nevertheless, since no such things as normalised-longevity or normalised-unemployment rates exist in reality, it is particularly useful to focus on how observed-attributes contribute to the overall measure of social inclusion, and what characterize the relationship between raw-variables within the aggregation framework.

The marginal contribution of each j-th raw-indicator to the overall synthetic measure can be determined by computing the partial derivative as in (3). Given that the selected aggregation model has $w_j = 0.25$, the magnitude of the marginal contribution is entirely determined by the steepness of the adopted normalisation function. Indeed, for a generic linear aggregation model L, and for any normalisation function v, it holds that:

$$\frac{\partial L\left(v\left(\mathbf{x}\right)\right)}{\partial x_j} = 0.25 \cdot v'_j\left(x_j\right). \tag{10}$$

The derivative v' represents the link implicitly imposed, when normalising data, between the original variable x and its counterpart $v(x)$. The partial derivatives in (11) and (12) illustrate such link for the data-driven min–max (MM), and the survey-driven min–max (sM), respectively:

$$\frac{\partial v_{MM\pm}^i\left(x^i\right)}{\partial x^i} = \pm\frac{25}{\max\limits_{t\in T}\max\limits_{i}\left(x_+^t\right) - \min\limits_{t\in T}\min\limits_{i}\left(x_+^t\right)} \tag{11}$$

$$\frac{\partial v_{sM\pm}^i\left(x^i\right)}{\partial x^i} = \pm\frac{25}{[\text{thresholds' range}]} \quad \text{if } x^i \in [\text{thresholds'range}]$$

$$= 0 \quad \text{if } x^i \leq \text{lowest threshold} \vee x^i \geq \text{highest threshold} \tag{12}$$

We can immediately notice that, in all of the cases, the effect of a one-unit increment in x on the transformed variable $\nu(x)$ is constant. This is due to the transformation functions being linear, and the benchmarks bmax, bmin, being fixed (they are either extracted from the data, or elicited from experts).[14] In particular, the higher the range or the standard deviation of a raw variable, the lower its unitary marginal contribution to the normalised one. Hence, unless all the attributes have very similar distributions and comparable units of measurement (which would make the normalisation itself of secondary importance), all the normalisations are performing a preliminary, and unequal, weighting of the original variables, regardless of the choice of the aggregation function.

A partial exception must be highlighted for the survey-driven min–max (12), for which the usual *non-satiation* hypothesis (more of a "good" is always preferred to less) is maintained in a weaker form. Indeed, more of a "good" is *non-ill favoured* with respect to less of it, after a certain performance is reached (and, conversely, more of a "bad" is *non-preferred* to less of it). As a rough realisation of the diminishing sensitivity hypothesis, the effects of a change in variables' score on the social utility is zero after the thresholds are crossed. From a policy-implication point of view, this suggests to focus on those dimensions whose performances lie farther away from the "desirability" level.

To help further clarifying the aforementioned observations, Fig. 1 reports a graphical visualisation of the two versions of the min–max transformation implemented on the selected data. The heterogeneity of the functions' steepness both *within* and *between* normalisation frameworks reflects the differences in each variable's thresholds' range. In particular, the steep of the survey-based functions for unemployment and school-dropouts is higher with respect to the data-driven version because of the shorter min–max range imposed by the experts. The opposite is true for life expectancy, which has a steeper normalisation under the data-driven strategy. To make some examples, a long-term unemployment rate of 3 % is normalised to 100 under the expert thresholds, to around 80 under the data-driven threshold. A life expectancy of 80 years old results in a transformed value around 70 under the expert-function, whereas it is around 40 in the data-driven normalisation.

We can now compute, by the means of partial derivations, the marginal contribution of each indicator with respect to the three synthetic measures LD, LS, that is, the impact that a unitary change in the original attribute has on the overall measure of social inclusion. Table 5 illustrates the results.

Since these marginal contribution coefficients cannot be easily compared across normalisation-methods, we normalised them row-wise, so that their sum is always 100. This allows us to interpret the results in terms of "relative weights", i.e., how much weight (in percentage) is given to a specific raw variable. Such normalised weights are shown in Fig. 2.

Although the weights were set as equal for each normalised attribute, those related to the actual indicators are highly un-balanced, regardless of the transformation adopted.

---

[14] The linearity hypothesis of the min–max can be relaxed by imposing a non-linear shape (convex, concave or s-shaped, see, e.g., Martinetti and von Jacobi 2012; Meyer and Ponthière 2011). Such alternatives were tested and do not in any way alter the implications of this paper.

Fig. 1   Min–max normalisation: survey-driven vs data-driven benchmarks
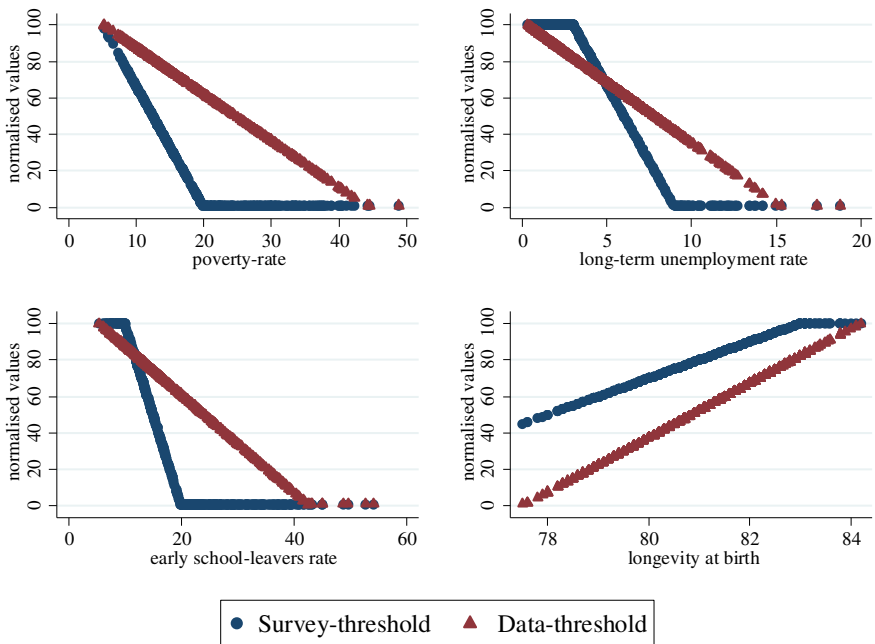
Data: Eurostat Regional Database 2016, Belgium, Denmark, Germany, Italy, Spain, years 2004-2012

Table 5 Dimensions' relative weights in linear data-driven model under different normalisations

	Longevity	Early school leavers	Long-term unemployment	Poverty-rate
Data-driven min–max	3.7	−0.7	−1.6	−0.6
Survey-based min–max*	2.5	−2.5	−4	−1.65

Data from the Eurostat Database 2016 (2004–2012)

* Derivatives refer to unitary increments of the original variable from a starting value within the interval $[\max(x_j) - \min(x_j)]$ as detailed in Table 4. For all the values outside the boundaries, a unitary variation would produce no zero change in the index

In particular, the longevity dimension is assigned a predominant role in the aggregation (a relative weight higher than 55 %) under the data-driven min–max. Indeed, this is the variable for which the data-driven min–max exhibits the highest slope (Fig. 1). Much lower effects derive from a decrease of one unit in long-term unemployment, and an even lower one from reductions in school-dropouts and poverty-rate.

Such trade-offs change significantly when the expert-based min–max is adopted. The dimensions' weights appear slightly more homogeneous: longevity and school-dropouts have equal relative relevance (24 %), while poverty-rate accounts for 15.4 % of the weight and the unemployment indicator being the one with the highest marginal effect on the aggregate measure (37.6 %).

Longevity at birth ■ School droputs ■ LT Unemployment ■ Poverty rate

	"Survey" Min-max	"Data" Min-max
■ Longevity at birth	24%	55%
▨ School droputs	24%	10%
▨ LT Unemployment	38%	25%
■ Poverty rate	15%	10%
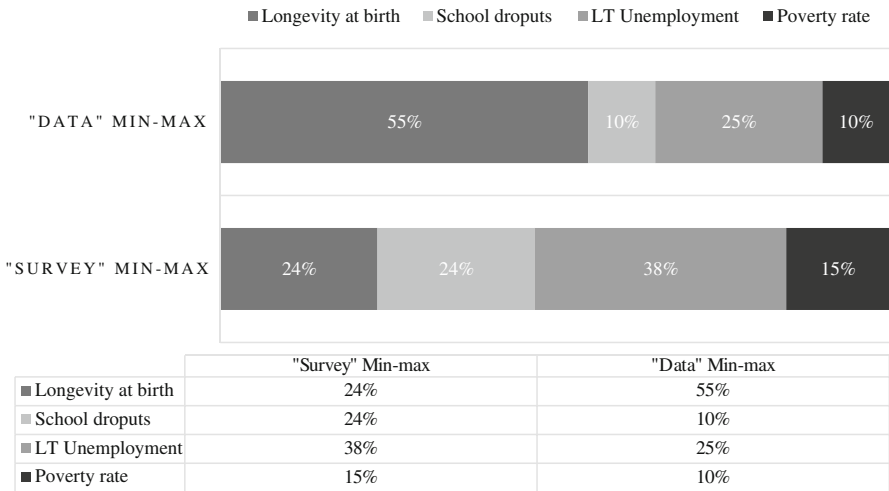
Fig. 2 Relative weights in different normalisation strategies

Table 6 Marginal rates of substitution in the linear model with data-driven min–max

	II. Can be compensated by the following change in…			
	Longevity years	% Points early school leavers	% Points l.t. unemployment	% Points poverty rate
I. An increase of...				
1 year in longevity		5.3	2.3	6.2
1 % Point in early school leavers	0.2		−0.4	−1.2
1 % Point in l.t. unemployment	0.4	−2.3		−2.7
1 % Point in poverty rate	0.2	−0.8	−0.4	

Moreover, using (4), the marginal rates of substitution between any pairs of indicators x_j, x_k can be computed for each of the three aggregation models. Results are reported in the following Tables 6 and 7.

As expected, the marginal rates of substitution mirror the heterogeneity in the relative weights, and yet convey a more pragmatic evidence on the relevance of the hidden, and partially unintended, trade-offs lying behind the apparently simple and neutral aggregation framework adopted. Just to make an example, in the data-driven min–max model, one additional year of longevity increases the synthetic index of social inclusion as it would a reduction of at least 5.3 percentage points in school dropouts, around 2.5 percentage points in long-term unemployment, and around 6.2 points in poverty rate. In the survey-based min–max model, the marginal rates of substitution for a unitary increase in life-expectancy are much lower: namely, 1 percentage point change in early-school leavers, 0.62 points of long-term unemployment, 1.5 points of poverty rate.

Table 7 Marginal rates of substitution in the linear model with the survey-driven min–max

	II. Can be compensated by the following change in…			
	Longevity years	% Points early school leavers	% Points l.t. unemployment	% Points poverty rate
I. An increase of...				
1 year in longevity		1	0.62	1.5
1 % point in early school leavers	1		−0.62	−1.5
1 % point in l.t. unemployment	1.6	−1.6		−2.4
1 % point in poverty rate	0.66	−0.66	−0.41	

These MRS are only valid when the variables take values within the boundaries [max(xj) − min(xj)]. Else ways, the MRS would be either zero, or infinite, or indeterminate, according to whether the numerator, the denominator, or both, in the ratio $v'_k(x_k)/v'_k(x_k)v'_j(x_j).v'_j(x_j)$ in (4) is zero

5.1 Discussion: Positive vs Normative Analysis

The aforementioned heterogeneous relative weights and trade-offs, both *within* and *between* the aggregation models, arise because of the (differences in the) adopted normalisation strategies, and will strongly influence the resulting indices, as Sect. 6 will show. Moreover, this happens in the context of aggregation frameworks granting "equal weights" to their components. As already stated, such label can be partially misleading, since the equal weighting pertains just to the normalised attributes. Indeed, to the extent to which rescaling is a requirement for composite measures, the actual aggregation concerns the transformed variables, in place of the observed performances, and yet there is an unavoidable and intrinsic difference between the interpretation of original and normalised performances. The transformed unit of measurement (e.g., between zero and one, if the min–max rescaling is adopted) can be interpreted as a sort of degree of fulfilment of some criterion. Whether this criterion should be purely statistical (e.g., being far or close to the observed minimum or maximum achievements), or whether it should encompass some informed value judgements related to the topic at hand (as in the expert elicitation or in the adoption of policy benchmarks), relies on the researcher's choice. Both directions are, in principle, correct.

What we would like to stress at this point is not whether such trade-offs are more acceptable under a data-driven or a survey-driven strategy: no normalisation is "safe" from the emergence of such hidden weights. Rather, we highlight that (i) normalisation-generated trade-offs are inevitable, (ii) the ground on which they are justified can differ greatly, depending on the normalisation strategy adopted, and, therefore, (iii) the resulting aggregate indices should be interpreted accordingly.

When a data-driven approach is selected, debating on the acceptability of the underlying marginal rates of substitution is a marginal issue: the justification, and therefore the interpretation, of such coefficients is inherently statistical. It follows that the interpretation of the resulting composite indices should be of the same nature, that is, statistical, which constitutes a strong and solid ground for a *positive* analysis of a composite phenomenon. That being said, the absence of value-judgements in the con-

struction of a data-driven index does not neutralise the hidden trade-offs shown in the previous tables. It is still true that, with a data-driven min–max and with a data-selection as described in Sect. 2, life-expectancy carries a weight which is more than twice what is assigned to the remaining three variables, and that one additional year of longevity is implicitly made equivalent to, e.g., 6.77 percentage points of school dropouts.

It follows that, by construction, such relative weights and marginal rates of substitution are sensitive to the choice of the data-sample and to distribution of the original variables. Indeed, and especially for the data-driven min–max, the presence of outliers in the data would stretch the range over which the normalisation is performed, therefore altering the original variable's marginal contribution to the overall index (as noted, we prudently excluded the extremely high values for the Spanish autonomous cities of Ceuta and Medilla in computing the data-driven benchmarks). As an additional warning, such transformations—again, especially the min–max—are not stable when data for new years or new regions become available, which could sensibly affect the distribution of data (Lefebvre et al. 2010). Similarly, a shift in the territorial dimension of the analysis (e.g., from a national to a regional or provincial level) will cause similar changes, since the provincial data are likely to exhibit higher variability than the regional ones.

Under the strategy of expert-elicitation of the transformation parameters, standard properties as strong non-satiation and continuity of the normalisation function are not guaranteed (indeed, in our example, the min–max becomes weakly monotonic when the elicited constraints are binding for some observed variable). Dimensions' trade-offs reflect the preferences of an actual group of experts, and are therefore independent from the selection of data and from the territorial dimension of the analysis. It is not possible to claim that opinions of experts would lead to more suitable benchmarks compared to the ones revealed from the data. Indeed, what differentiates such benchmarks and trade-offs from the data-driven ones is that the former are easier to be interpreted under an economic perspective, in terms of social desirability. Therefore, the resulting aggregate measure would constitute a tool for *normative* analysis.

However, such an elicitation method suffers from the arbitrariness embedded in any survey exercise, e.g., choice of the population, bias in the framing of questions, and is by definition sensitive to such choices. As an example, it is likely that a panel of experts from another Italian or European university would lead to different normalisation parameters. However, as far as the choice of the panel is kept homogeneous, the economic justification for the expert-based strategy is always preserved (i.e., normalisation as a social desirability function), whereas the variability in the experts' answers among different panels is analogous to the aforementioned sensitivity in the data-driven method, when new territorial units or years are added to the data.

6 Results

Results for the linear model with data-driven normalisation (LD model) are summarized at country levels in Table 8 (population-weighted averages at country level),

Table 8 Social inclusion measure and coefficients of variation, data-driven normalisation

	2004	2005	2006	2007	2008	2009	2010	2011	2012
Average									
Belgium	63.5	62.6	64.8	67.0	67.9	69.3	68.8	70.4	70.1
Germany	61.1	61.7	63.9	66.7	69.0	70.5	70.9	72.8	74.9
Denmark	–	–	–	68.7	71.6	71.7	70.7	73.4	74.4
Spain	53.2	56.1	59.8	60.1	59.5	58.1	55.5	54.4	52.1
Italy	62.4	62.6	65.9	67.5	68.4	68.8	69.8	68.6	66.2
Coefficient of variation within country									
Belgium	0.22	0.25	0.25	0.25	0.25	0.23	0.24	0.25	0.26
Germany	0.15	0.15	0.14	0.13	0.12	0.11	0.11	0.10	0.10
Denmark	–	–	–	0.03	0.03	0.03	0.04	0.04	0.02
Spain	0.26	0.22	0.22	0.20	0.21	0.22	0.26	0.28	0.29
Italy	0.28	0.28	0.26	0.25	0.23	0.23	0.22	0.25	0.26

together with coefficients of variation within countries. Full results are available in the Appendix C.[15] Data for Denmark are not available before 2007.

Results for the linear model of social inclusion with survey-driven normalisation (LS model) are summarized in Table 9 (aggregated at countries' level), and fully reported in the Appendix D.

The left graph in Fig. 3 provides an overall view of the social inclusion trends at country level, under the data-driven model. Differences between countries appear rather limited until 2007. Denmark and Italy report the highest levels of Inclusion, followed by Belgium, Germany and Spain. There is a general increase in the index for all countries until 2008. Since then, the Spanish performance declines and eventually reaches in 2012 levels of Inclusion close to those of 2004. Italy's index is apparently slightly affected by the economic crisis (roughly, from 2010 onward): its aggregate Inclusion ceases to improve and starts to decline, reaching in 2012 the same levels of 2006. Belgium and Germany show a general continuous increase in their levels of Inclusion. In particular, since 2008 Germany overcomes Italy, while Belgium does it in 2011. Overall, the situation in 2012 appear to be more heterogeneous than it was in the early years in our sample: Italy and Spain show a negative trend (increasing exclusion), while Belgium and Germany continue to improve their aggregate performance. Moreover, while many Italian regions score very well throughout the available time span, some others are consistent bad performers. Most of the top-10 regions between 2004 and 2012 are Italians, yet both Campania and Sicilia constantly rank at the very bottom of the tables. Thus, the social inclusion index emphasizes the well-known dichotomous socio-economic picture of Italy as well as the contradictions of Belgium

---

[15] Throughout this discussions of our results, we will concentrate more on commenting the levels of the composite indices, rather than the rankings of the regions, following the recommendations in Atkinson et al. (2004). The authors argue that the ultimate concern of the policy should lie on performance levels, since rankings might conceal the actual distances between territorial units, thus leading the reader to misleading conclusions.

Table 9 Social inclusion measure and coefficients of variation, expert-based normalisation

	2004	2005	2006	2007	2008	2009	2010	2011	2012
Average									
Belgium	61.6	60.2	62.3	64.0	64.6	67.4	66.6	68.0	67.3
Germany	55	56.2	58.2	63.0	67.0	69.5	69.4	70.8	73.8
Denmark	–	–	–	70.0	73.9	74.1	74.1	78.4	79.2
Spain	46.9	48.8	52.2	51.9	51.5	47.0	36.4	32.2	30.3
Italy	50.3	51.0	55.4	58.4	57.8	58.0	58.1	57.1	54.5
Coefficient of variation									
Belgium	0.43	0.49	0.48	0.47	0.49	0.40	0.44	0.44	0.46
Germany	0.30	0.30	0.28	0.26	0.24	0.20	0.20	0.18	0.14
Denmark	–	–	–	0.06	0.04	0.06	0.05	0.04	0.02
Spain	0.28	0.20	0.20	0.18	0.19	0.27	0.42	0.42	0.40
Italy	0.41	0.41	0.36	0.34	0.34	0.35	0.36	0.40	0.42
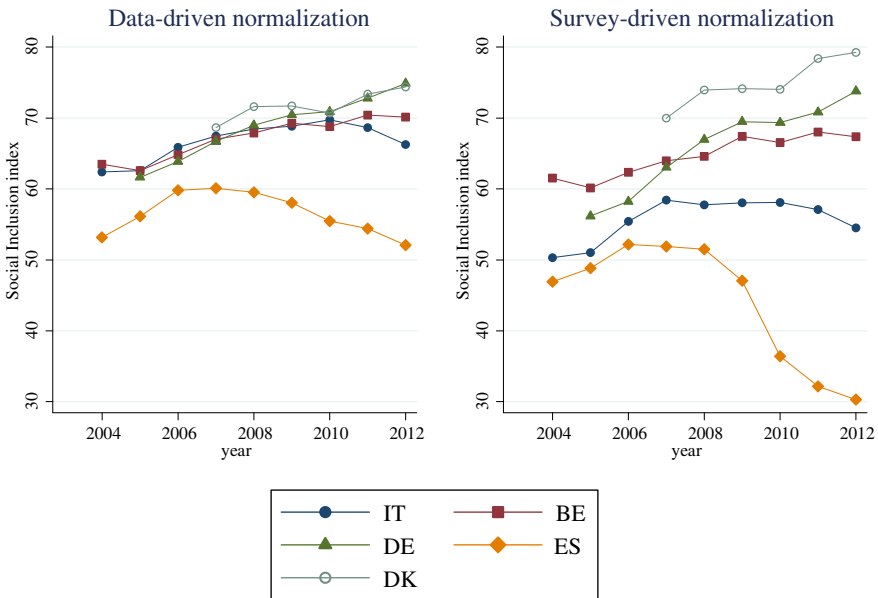
Data from Eurostat Regional Database 2016. Country averages weighted through regional population.

Fig. 3 Time-trend of the social inclusion index under different normalisations

where important differences exist between the Flemish region, the Bruxelles region and Wallonia.

When adopting the survey-driven normalisation, (right graph in Fig. 3) country trends are confirmed yet between-country differences are more evident. Social inclusion increases in Belgium, Germany and Denmark, while Italy and Spain experience a continuous decline, which starts since the early years of the economic crisis. The over-

all picture is quite different from the one commented before, for at least three reasons. First, social inclusion in levels is lower for Italy, Spain and Belgium, and Italy lies now always below both Belgium and Germany for all the time-interval, while Denmark is by far the best performing country. In particular, Italy (blue circle markers) and Germany (green triangle markers) show similar levels of Inclusion in 2006. After that, the index continues to increase for Germany while it remains constant (and then declines) in Italy. Thus, there is a clear phenomenon of rank reversal between the two models: in terms of "desirability" (as defined by the expert-panel), the aggregated Italian picture is worse than the German one, and we will discuss this effect in a Sect. 6.1. Second, the German regions, together with Belgium's Flanders, achieve more top-10 rankings than they were under the previous specification, especially after 2007. Second, Spain exhibits a decline in social inclusion which is much more dramatic than it appeared before. The negative trend starts after 2006 but the drop in performance is substantial after 2008, leading to final levels of Inclusion much lower than they were in 2004. Again, we stress the fact that, albeit the negative trend for Spain was already visible from the baseline data-driven model, this picture conveys a much stronger need for intervention. Third, the heterogeneity within each country is much higher, as noticeable from the coefficients of variation in Table 9. Spain, Italy and Belgium still report the highest coefficients, but heterogeneities are rather constant in the latter country while they are increasing in Italy and especially in Spain. An opposite trend appears in Germany and Denmark, where convergence of social inclusion between regions seems to occur.

Changing the normalisation strategy has large implications on the distribution of the social inclusion index in our regional sample, as we show by plotting a kernel-density estimation for the distribution of the data-driven and the survey-driven indices, for 2004 and 2012 (Fig. 4). The left graph (LD model) highlights how the distribution became less disperse and more uni-modal between the starting and the final period. Conversely, the one on the right reports a more heterogeneous starting distribution, which becomes clearly bi-modal in the final period of the analysis, confirming the previously commented trends. Indeed, such differences in trends and levels could lead to very different policy implications in terms of how to assuage and prevent social exclusion.

In order to test whether the two specifications convey similar rankings, we perform a Kendall's tau[16] tests between the ranking of the data-driven model and the one coming from the survey-driven model, for each year. Results are reported in Table 10, excluding the Danish regions since they have no Inclusion index before 2007. Coefficients' magnitude indicate that rakings' correspondence is far from perfect, yet we can always reject (at 99 %) the null-hypothesis of no correlation between the models' rankings.

Finally, let us now focus on the results for the year 2012, the last available year in our sample. Some similarities emerge from the two specifications, i.e., Germany

---

[16] The Kendall-τ test is a non parametric method that allows to measure the degree of correspondence between two rankings. In particular, the Kendall-τ b allows for the possibility of ties in the rankings. Command in STATA: ktau. A resulting test-value of zero would indicate that no correlation exist between the two rankings, while a value of 1 would indicate perfect correlation. Conversely, negative values (down to a minimum of -1) would indicate that rankings are inverted.
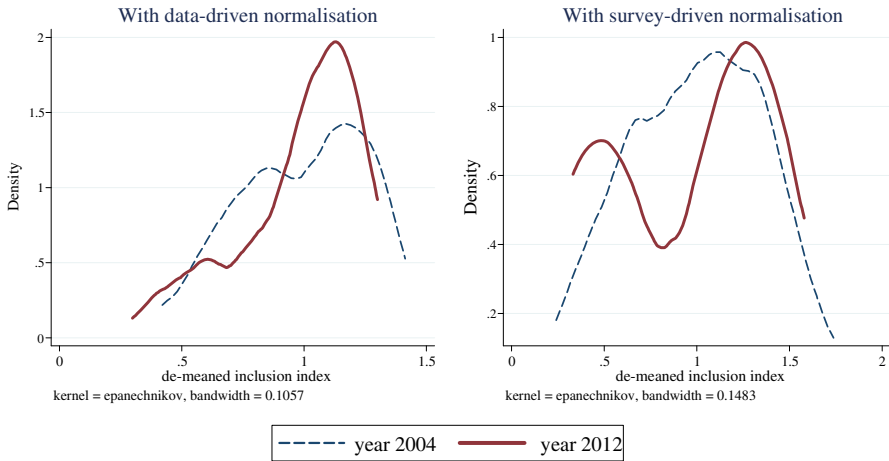
Fig. 4 Distributions of inclusion indices

Table 10 Kendall's τ correlation coefficients (Danish regions are excluded)

	2004	2005	2006	2007	2008	2009	2010	2011	2012
Kendall's τ	0.79	0.78	0.80	0.78	0.75	0.73	0.73	0.78	0.78

and Denmark stand out as the top-two countries for inclusion levels, regardless of the adopted normalisation strategy. Nevertheless, their ranking is reversed when switching from the data-driven approach, where Germany has a small advantage, to the survey-driven strategy, where the index for Denmark is considerably higher. Belgium is third-placed under every specification. Yet, its relative Inclusion-loss with respect to both Germany and Denmark is much lower under the two data-driven strategies. Similarly, average values for Italy and Spain are, respectively, fourth and fifth placed. Nevertheless, the spread between them and the remaining countries is extremely higher when the survey-based normalisation is implemented.

As the coefficients for standard deviation suggest, country averages conceal a substantial degree of heterogeneity between regions. Spain, Italy and Belgium, in particular, exhibit high levels of variability, with their standard deviation being around 40 % of national averages in the survey-driven model. Moreover, the choice of the normalisation strategy strongly affects the extent of such heterogeneity. This leads us to conclude that looking at national averages cannot provide a fully informative tool to evaluate the phenomenon at study. The graphical representations in Fig. 5 allow us to visualise the regional-dimension of the indices and to draw further valuable insights on both levels and variability between and within countries and models. We report such figures for 2012 as well as for the starting year 2004.

A comparison of the graphs for 2012 (the two graphs at the bottom) in Fig. 5 illustrates how, switching from the survey-to the data-based model: (i) Spanish and Italian regions experience a substantial increase in Inclusion, particularly with respect
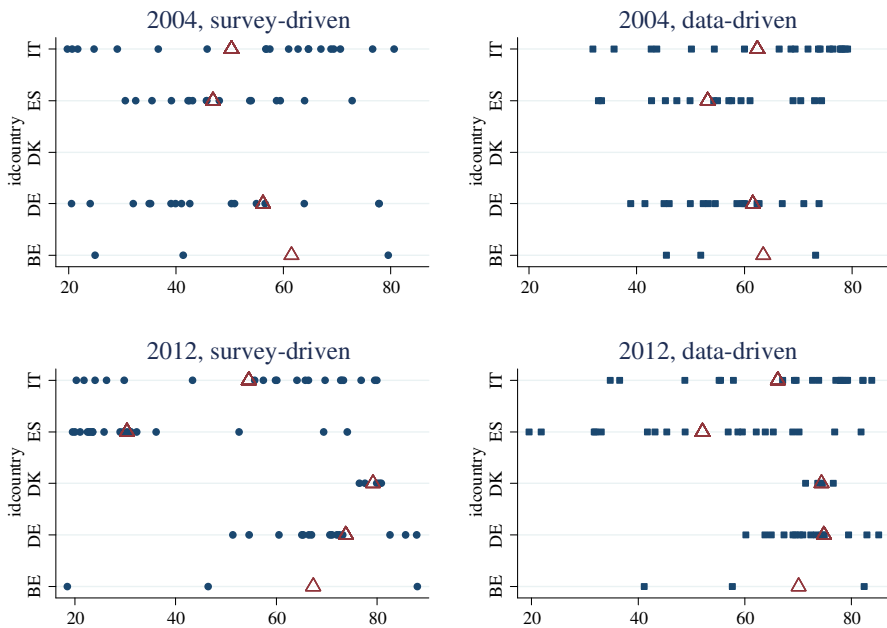
Fig. 5 Heterogeneity within countries due to normalisation, 2004 and 2012

to Germany and Denmark; (ii) Danish territories are relatively worse off; (iii) all countries, except for Spain, exhibit a smaller degree of regional variation.

The case of Spain in 2012 is of particular interest for our methodological approach. Indeed, even though at country level the regional ranking is consistent, the Spanish picture conveyed by the survey-driven model is more troublesome: only three regions appear to have levels of Inclusion comparable with Germany and Italy (Navarra, Paìs Vasco, Cantabria), with the others lying far below on the metric scale, placed (almost exclusively, together with Italian territories) in the bottom-20 of the ranking. This denotes a notable welfare loss with respect to the remaining regions, as well as a distinctive skewedness of the distribution. Both of these features are completely absent from the data-driven results, where Spain exhibits a high degree of heterogeneity, with some territories performing relatively well, others relatively bad, and a group lying in between. In particular, under these frameworks, a substantial group of ten regions (therefore, a majority share) appears to be roughly in line with the German and Italian distribution of the index (Navarra, Paìs Vasco, Cantabria, Madrid, Castilla y León, Aragón, La Rioja, Galicia, Asturias, Cataluña). Evidently, the survey-driven model conveys a much stronger early-warning message than the data-driven ones, which would potentially lead to very different policy implications.

We can exploit Fig. 6 to spot the normalisation-induced variation in European territorial rankings. The graph summarizes the rankings obtained for each region under each normalisation for 2012, and territories are sorted by their ranking in the linear data-driven model ("X" marker, while the survey-driven ranking has a "circle" mark). The labels on the X-axis report the NUTS code for the administrative regions, whose
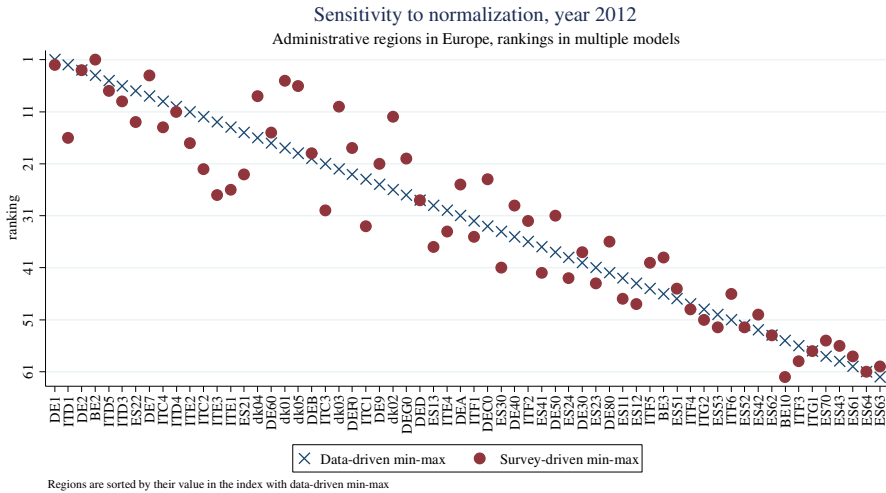
Fig. 6 Rankings' sensitivity

first two letters identify the country (e.g., "ES01" refers to a Spanish region). The correspondence between labels and region names is illustrated in Appendix A (Tables 11, 12).

By focusing on the X and the circle marks, it is relatively easy to identify the regions which are "penalized" by the data-driven min–max normalisation with respect to the survey-driven one (i.e., the "x" lies below the "circle"), as well as the opposite (when the "x" lies above the "circle"). There are eight Spanish regions among the last eleven under all of the three specifications, and their ranking is basically normalisation-invariant (Valencia, Castilla-la Mancha, Murcia, Canarias, Extremadura, Andalucía, Melilla, Ceuta).[17] Conversely, almost all of the remaining regions' rankings are strongly affected by normalisation choices, and numerous substantial rank-reversals occur, as we illustrate through few examples. The Comunidad Foral de Navarra (ES22) ranks 7th under the data-driven min–max, while dropping to 13th place under the survey-based model. Meanwhile, the German Land Hessen, ranked just below Navarra in the data-driven model (8th), reaches a much higher rank in the survey-driven model (4th). An even more dramatic set of rank-reversals occurs when comparing Navarra with all of the five Danish regions: although the latter group appears at far distance in data-driven model (the best ranked Danish region being Midtjylland at 16th), these territories largely outrank Navarra in the survey-based framework. Incidentally, the exact same reversal affects also the region of País Vasco (ES21). Spanish regions with lower positions in the table are similarly outranked by other European territories when the analysis is performed through survey-driven normalisation: e.g., Cantabria, Madrid, Castilla y León, overtaken by, among others, Nordrhein-Westfalen, Molise, Bremen.

---

[17] Still, we know from Fig. 5 that their distance from the more virtuous territories is lower under the data-driven approach.

As well as it is for Spain, the relative performance of Italian regions changes distinctly when comparing methods. Under the data-min–max, the Italian regional picture appears extremely heterogeneous, with some territories performing worse, and the remaining others being equal, or better off, with respect to regions in Germany and Denmark. Moreover, the levels and distribution of Italian social inclusion appear very similar to those in Spain, when excluding the two "outlier" autonomous cities of Ceuta and Medilla. Switching to the survey-based normalisation makes the picture change radically, and in a different fashion with respect to what was discussed for Spain. Italy appears as a highly severed country: the number of low-performing territories is sensibly higher, and a much larger distance separates these territories with the remaining relatively-good performing ones, which are—in turn—worse placed with respect to both Germany and Denmark than they were under the data-driven strategy. Nevertheless, the overall Italian picture looks substantially better than the Spanish one, again, contrarily to what could be inferred from the data-driven model. Figure 6 helps in identifying the numerous rank-reversals affecting Italian, German and Denmark regions, which happen almost exclusively to the benefit of the latter two countries. A notable example is the northern region of Trentino-Alto Adige (ITD1), ranked 2nd in the data-driven approach, which drops to 17th under the survey-based approach, being overtaken by German, Danish and Italian regions, as well as by Madrid. Similarly, the central regions Toscana (ITE1) and Marche (ITE3) as well as the industrial north-west region Piemonte (ITC1), who all rank mid-high in the data-driven models, lose numerous positions once the normalisation switches.

Belgium is affected by the aforementioned "Italian effect", yet to a lower extent. Although the ranking of its three regions appears extremely robust, both the within-country heterogeneity and the between-countries relative rankings are substantially modified by the methodological choices. The Inclusion level for Flanders, as well as its ranking, is constantly very high, and is increased when the survey-normalisation is adopted. The French-speaking region Wallonia, however, has a significantly lower performance under the survey-based model, which results in a wider gap with the Flemish region. Finally, the Bruxelles region is already among the worst-ranked in the data-driven model, and yet it drops to bottom of the ranking under the survey-model. As a consequence, the emerging Belgian picture conveys much more heterogeneity in the survey-based framework.

A similar pattern can be found for Germany: although its overall levels of social inclusion are higher than Belgium and Italy, the survey-driven approach returns a degree of within-country variability that is absent in the data-driven models.

The distribution of the social inclusion index for Denmark is characterised by a low degree of dispersion, regardless of the adopted normalisation. As for levels and rankings in the data-driven models, Danish regions score relatively high values. Yet, their best performing territory is overcome by one Belgian, nine Italian, three German, and two Spanish regions. Conversely, in the survey-driven framework their ranking significantly improves: e.g., the worst performing Danish region is overcome by just three regions from Germany, three from Italy, and one from Belgium.

6.1 Weights and Rank-Reversals

It is useful to briefly recap how different normalisation strategies can lead to substantially different scenarios of social inclusion in Europe. The key factors to consider are: (i) the relative-advantage that each country has in a specific dimension; (ii) as well as the *within-country* heterogeneity between the performances of the four selected raw-variables. Indeed, Spain and Italy present, on average, a strongly unbalanced dashboard: the longevity dimension is particularly high, relative to the other countries, while education and socio-economic variables show much worse relative-values. Since the two normalisation strategies (data-driven vs survey-driven) give opposite relative weights to these dimensions (a prevalent weight to longevity in the data-driven models; a more equal set of weights in the survey-driven one, with some prevalence to unemployment) this leads to the aforementioned difference in-levels, with many Spanish and Italian regions falling to lower-ranked positions. Countries like Denmark and Germany are less affected by the change, given that their dashboard of indicators is uniform. Moreover, heterogeneity *within* countries is very different: Spain, Italy and Belgium show coefficients of standard deviations much higher than Denmark and Germany in each of the included variables. This means that the Inclusion-mix can differ greatly between regions, within the same country. This kind of heterogeneity is somehow softened in the data-driven models, given that a single dimension gets such a large relative weight. In the survey-driven model, conversely, because of the higher weight given to the remaining dimensions other than longevity, such heterogeneity is enhanced.

This, in turn, explains the numerous rank reversals discussed in this section. Let us consider, for instance, the case of Italy and Germany, who exhibit, since 2004, similar trends in early school-leaving, life expectancy at birth and poverty rate. Nevertheless, the levels of these variables are quite different: there are much more school-dropouts and poverty rates in Italy, which also presents substantially higher longevity. When it comes to long-term unemployment, the country-trends are crossing: Italy experienced a consistent decline in its labour market performance, while Germany saw a constant improvement (according to many observers, a consequence of the Hartz Reforms). Such trends are summarised in Fig. 7.

In the data-driven model for Italy, the increase in normalised life-expectancy more than counterbalances, due to its substantially higher relative weight, the worsening conditions in the labour market, therefore allowing the overall measure to increase slightly. Almost no role is played by early school-leaving, which has very little variations. In Germany, all the dimensions improve, thus leading to a regular increase in the composite measure, yet the Inclusion index is lower than for Italy, exactly because of the weight given to life expectancy.

When the elicited benchmarks are implemented, life expectancy becomes the best performing dimension for Italy, while early school-leaving is heavily penalized. In Germany, both these normalised-attributes are much higher than under the data-driven model. Finally, although being rather small in 2004, the spread between the countries' long-term unemployment normalised-rates becomes much more evident. Given the reduced relative weight given to longevity in the expert-based model, and the higher
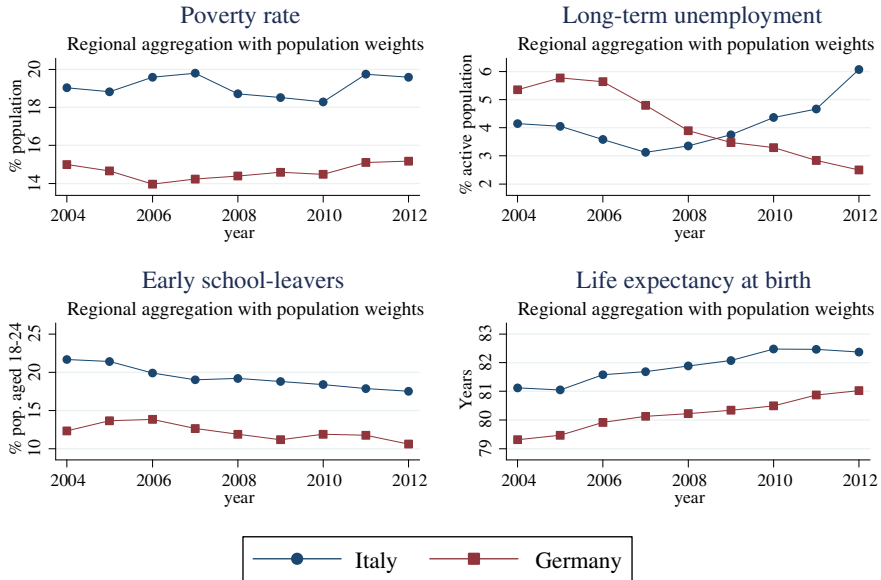
Data: Eurostat Regional Database 2016

Fig. 7 Time trends of indicators in Germany and Italy

one given to long-term unemployment, the German index is (1) higher than Italy, while it was lower in the previous specification; and (2) increasing at a higher rate.

We can further illustrate similar rank-reversals by focusing on the results for the year 2012. The Italian region Trentino-Alto Adige, for instance, scores slightly better than Belgium's Flanders under the data-driven model. Although both regions admittedly exhibit virtuous performance in each of the four raw variables, the Italian region is relatively better off in longevity-at-birth (83.6 vs 81.4 years) while Flanders has an edge in unemployment (1.5 vs 1.6), education (8.7 vs 15.8) and poverty (9.8 vs 12.5). Under the data-min–max, Trentino's losses in three over four dimensions are more than compensated by the gain in life-expectancy, so that the two territories end up having very similar ranking and levels (83.8 for Trentino, 82.4 for Flanders, in the data-driven min–max). Conversely, the survey-based normalisation implies larger weights to the three dimensions where Trentino trails Flanders, thus enhancing the score and the ranking of the Flemish region, while depressing Trentino's ones (88 for Flanders, 73 for Trentino).

Another example comes from the comparison between Spanish País Vasco and Danish Syddanmark. In the data-driven min–max model they rank 15th (76.9) and 22nd (73.7), respectively. Under the survey-based framework the Danish region climbs to 10th place (77.6), while País Vasco drops to 23rd (69.4). Again, the reason for this shift relies on the heterogeneity in the two regions' dashboards, enhanced by the new expert-based weights. País Vasco performs very well in life-expectancy (83.1, well in school dropouts (11.5 %, close to the 9 % of Syddanmark), yet it loses ground in long-term unemployment (6.4 %, while Syddanmark's rate is just 2.4 %).

7 Conclusion

The unavoidable subjectivity of composite measures of Well-being are cause of controversies in this field of economic analysis. In this paper, we argued that the lack of transparency on methodological choices can turn out to be more troublesome than subjectivity per se: specifically, we focused on the choices of the normalisation function. In the context of building a synthetic Index of social inclusion for 63 European regions between 2004 and 2012, we showed the consequences of adopting different normalisation methods while keeping constant the (linear) aggregation model, with equal weights allocated to the normalised dimensions.

To the extent to which rescaling is a requirement for composite measures, the actual aggregation involves the transformed variables, rather than the observed performances. There is an unavoidable and intrinsic difference between the interpretation of original and normalised performances, and yet social researchers are ultimately interested in the contribution of the *original* variables to the aggregate index. The rescaled unit of measurement (e.g., between zero and one-hundred) can be interpreted as a sort of degree of fulfilment of some criterion. Whether this criterion should be purely statistical (e.g., being far or close to the observed minimum or maximum observed achievements), or whether it should encompass some informed value judgements related to the topic at hand (as in the expert elicitation or in the adoption of policy benchmarks), relies on the researcher's choice.

In the former case, the *agnostic* choice of "letting the data talk", standard properties as strong non-satiation and continuity of the normalisation function are guaranteed, yet dimensions' trade-offs are hard to interpret in economic terms or from a social desirability perspective. Therefore, the resulting aggregate measure would be characterized as a tool for mainly "positive" analysis.

In the latter case, conversely, the normalisation function may become weakly monotonic, when the elicited constraints are binding for some observed variable. Moreover, the elicitation method suffers from the arbitrariness embedded in any survey exercise (selection of the experts, biases in the framing of questions). Finally, dimensions' trade-offs reflect the preferences of an actual group of experts, thus are independent from the data-selection, and allow to characterize the final measure with a "normative" connotation.

The main result of our analysis is that neither method is neutral. Indeed, in our case study, data-driven normalisation softens the aftermaths of the recent economic crisis as well as the differences between territories, since it puts a consistent weight on the longevity variable which follows dynamics that are only partially related to socioeconomic contingencies. Conversely, the survey-driven normalisation emphasizes the worse performance in long-term unemployment and early-school leaving of Italy and Spain, thus producing a bi-modal picture of Inclusion in Europe, with a cluster of region scoring very high and another scoring very low. As a result, numerous rank-reversals occur between regions when switching the normalisation methods.

Conceptually, both normalisation strategies are acceptable, and it is hard to label one as "preferred" to the other. Although the picture obtained with the use of the expert-based approach seems to better reflect the most recent economic trends, this might change if another expert-panel is selected. In other words, the opinions of

experts are not guaranteed to lead to more suitable benchmarks compared to the ones revealed from the data. Indeed, what differentiates the two strategies is the nature of the economic justification lying behind the resulting benchmarks and trade-offs: social welfare preferences, in the case of the expert-elicitation, as opposed to frequency-based parameters, in the case of the data-driven models.

As mentioned in the introduction, building a synthetic index of social inclusion requires that the concept's indeterminate multiplicity be made determinate through a specification of its contents, and of their relationship. In this paper we do not offer a real solution, as far as the normalisation process is concerned. Rather, we raise the awareness that, when a "real solution" is presented, it may not be the unique one, and its premises may hide peculiar trade-offs which should be made transparent to the reader. E.g., in our social inclusion study, a policy maker should be aware of the weights and trade-offs lying behind the data-driven results, and should be presented with the alternative picture coming from the expert-based strategy, in order to be able to draw more informed and efficient conclusions on the topic at hand.

Appendix A: Data-Description

See Fig. 8 and Tables 11, 12.
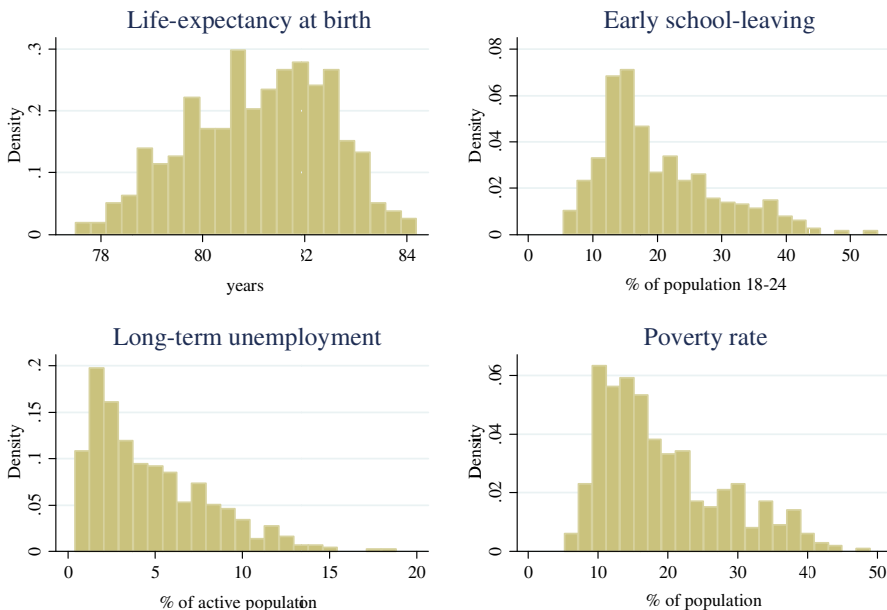
Fig. 8 Variables' distribution for data-driven normalisation for Belgium, Denmark, Germany, Italy and Spain from 2004 to 2012. *Data* Eurostat Regional Database 2016, Administrative regions of Belgium, Denmark, Germany, Italy and Spain, years 2004–2012

Table 11 Correlation between the four variables of social inclusion

	Longevity	Early school leaving	Long-term unemployment	At-risk-of-poverty rate
Longevity	1			
Early school leaving	0.28	1		
Long-term unemployment	−0.02	0.14	1	
At-risk-of-poverty rate	0.1	0.53	0.54	1

Data: Eurostat Database 2016, Administrative regions of Belgium, Denmark, Germany, Italy and Spain, years 2004–2012

Table 12 Administrative regions, names and NUTS codes

Country	NUTS	Administrative region
BE	BE10	Région de Bruxelles-Capitale/Brussels Hoofdstedelijk Gewest
BE	BE2	Vlaams Gewest
BE	BE3	Région wallonne
DE	DE1	Baden-Württemberg
DE	DE2	Bayern
DE	DE30	Berlin
DE	DE40	Brandenburg
DE	DE50	Bremen
DE	DE60	Hamburg
DE	DE7	Hessen
DE	DE80	Mecklenburg-Vorpommern
DE	DE9	Niedersachsen
DE	DEA	Nordrhein-Westfalen
DE	DEB	Rheinland-Pfalz
DE	DEC0	Saarland
DE	DED	Sachsen
DE	DEE0	Sachsen-Anhalt
DE	DEF0	Schleswig-Holstein
DE	DEG0	Thüringen
DK	dk01	Hovedstaden
DK	dk02	Sjælland
DK	dk03	Syddanmark
DK	dk04	Midtjylland
DK	dk05	Nordjylland
ES	ES11	Galicia
ES	ES12	Principado de Asturias
ES	ES13	Cantabria
ES	ES21	País Vasco

Table 12 continued

Country	NUTS	Administrative region
ES	ES22	Comunidad Foral de Navarra
ES	ES23	La Rioja
ES	ES24	Aragón
ES	ES30	Comunidad de Madrid
ES	ES41	Castilla y León
ES	ES42	Castilla-la Mancha
ES	ES43	Extremadura
ES	ES51	Cataluña
ES	ES52	Comunidad Valenciana
ES	ES53	Illes Balears
ES	ES61	Andalucía
ES	ES62	Región de Murcia
ES	ES63	Ciudad Autónoma de Ceuta (ES)
ES	ES64	Ciudad Autónoma de Melilla (ES)
ES	ES70	Canarias (ES)
IT	ITC1	Piemonte
IT	ITC2	VDA
IT	ITC3	Liguria
IT	ITC4	Lombardia
IT	ITD1	TAA
IT	ITD3	Veneto
IT	ITD4	FVG
IT	ITD5	ER
IT	ITE1	Toscana
IT	ITE2	Umbria
IT	ITE3	Marche
IT	ITE4	Lazio
IT	ITF1	Abruzzo
IT	ITF2	Molise
IT	ITF3	Campania
IT	ITF4	Puglia
IT	ITF5	Basilicata
IT	ITF6	Calabria
IT	ITG1	Sicilia
IT	ITG2	Sardegna

Appendix B: Description of the Survey

The Survey was structured as follows:

- An introductory section discussed the topics, the purpose and the contents of the survey.
- Respondents were asked to select the variables (amongst the four described in Sect. 2) for which they would be willing to perform an evaluation.
- A randomization led the respondent to a page devoted to one of the selected variables. All pages were homogeneously designed with a consistent phrasing.
- The EUROSTAT definition of the variable at hand was offered, and descriptive statistics were shown through a bar graph, for 25 European countries (years 2000 and 2012).
- The main task of the survey was then detailed. Respondents should identify, according to their own opinion, two main thresholds for the variable at hand: a negative threshold, defined as a "value of the selected variable which conveys a certainly undesirable and problematic condition", and a positive defined as "a value conveying a certainly desirable and virtuous condition". The threshold had to be chosen by dragging a slider (using the mouse left-click) on a predetermined discrete interval of values,[18] and releasing it to identify the preferred value (see Fig. 9 for a snapshot of the negative-threshold choice for life expectancy).
- An example involving a mock variable "X" explained how to deal with the Qualtrics layout in order to identify the thresholds.
- After choosing the positive and the negative thresholds, a confirmation was required by clicking on "confirm and proceed" button, which would lead the respondent to the next variable-specific page, or to the last section of the survey (if no variables were left).
- The last section of the survey included questions on respondents' age, gender and affiliation (either Economics or Management).

As an example, let us consider the survey-page devoted to the life-expectancy-at-birth indicator. First, a definition of life expectancy was provided. Then, data for 25 European countries (years 2000 and 2012) were shown. At this point, respondents are faced with the summary of what they will be asked to do, i.e., identifying both a favourable and a harmful threshold for life-expectancy-at-birth, according to their own opinion. The harmful threshold is defined as a "level of longevity which represents a certainly negative and undesirable condition". The favour threshold is defined as a "level of longevity which represents a certainly positive and desirable condition". Before reaching the actual question, a full example was provided with a generic variable "X". Respondents had, then, to determine the harmful threshold by dragging a slider on an interval of values (with the left mouse-click), and dropping it at the point that corresponded to their view of a certainly undesirable level of longevity.

---

[18] Fixed intervals of values were imposed in order to avoid extreme and implausible choices (like 0 years old of longevity as "harmful" threshold). The predetermined intervals were: [90–60 years] for longevity; [0, 50 %] for early-school-leaving; [0, 50 %] for long-term unemployment; and [0, 50 %] for poverty-rate. No respondents chose one of the non-zero extremes as their preferred threshold.

SOGLIA DI RISCHIO: qual è un livello di speranza di vita alla nascita che Lei valuterebbe come sicuramente negativo, non desiderabile e problematico in un generico contesto territoriale?

Per individuare la soglia nella tabella seguente occorrerà trascinare il cursore verticale (che assume colore blu non appena lo si sfiori) fino a raggiungere il numero di anni desiderato.

aspettativa di vita media alla nascita (misurata in anni)

| | 60 | 65 | 70 | 75 | 80 | 85 | 90 |

SOGLIA DI RISCHIO: giudizio SICURAMENTE NEGATIVO per valori inferiori a

Fig. 9 Choice of the negative threshold for life expectancy

Table 13 Descriptive statistics on the survey' sample

	Economics	Management	Overall
Respondents	59	28	88
Age			
Less than 40 (%)	36.1	13	28.6
Between 40 and 49 (%)	29.9	43.5	34.3
Between 50 and 59 (%)	23.4	26.1	24.3
60 or more (%)	10.6	17.4	12.6
Female respondents	40.4	34.8	38.6

Figure 9 illustrates the choice that respondents were facing for the harmful threshold of longevity. The choice was not entirely free, since we constrained respondents to select a level of life expectancy within a predetermined interval ranging from 60 to 90 years old, in order to avoid extremely implausible choices (like 0 years old). Similar steps characterized the choice of the favourable threshold, where respondents had to select their answer in the same interval between 60 and 90 years old. A cautionary disclaimer was emphasized at this point, stressing the fact that the favourable threshold should, by construction, be higher than- or equal to- the harmful threshold previously selected.[19]

Out of 149 invitations, we received 88 responses, thus resulting in a response rate of almost 60 %. The completion rate is 100 %, meaning that all those who took part to the survey managed to actually complete it. Out of 88 respondents, 59

---

[19] The disclaimer aimed at avoiding inconsistent choices, e.g., a respondent who would choose, say, 81 years old as a harmful threshold, and subsequently choose 80 as a favourable threshold. No such patterns occurred.

Table 14 Survey-elicited benchmarks

	Median elicited minimum (25–75p)	Median elicited maximum (25p–75p)
Longevity	73 years (70–75)	83 years (80–85)
Early school leaving	10 % (5–10)	20 % (15–25)
Long-term unemployment	3 % (2–4)	9 % (5.25–10)
At-risk-of-poverty rate	5 % (3–7)	20 % (17–21.5)

On-line survey (QUALTRICS software), 88 responses from professors in Economics or Management at the Ca' Foscari University of Venice
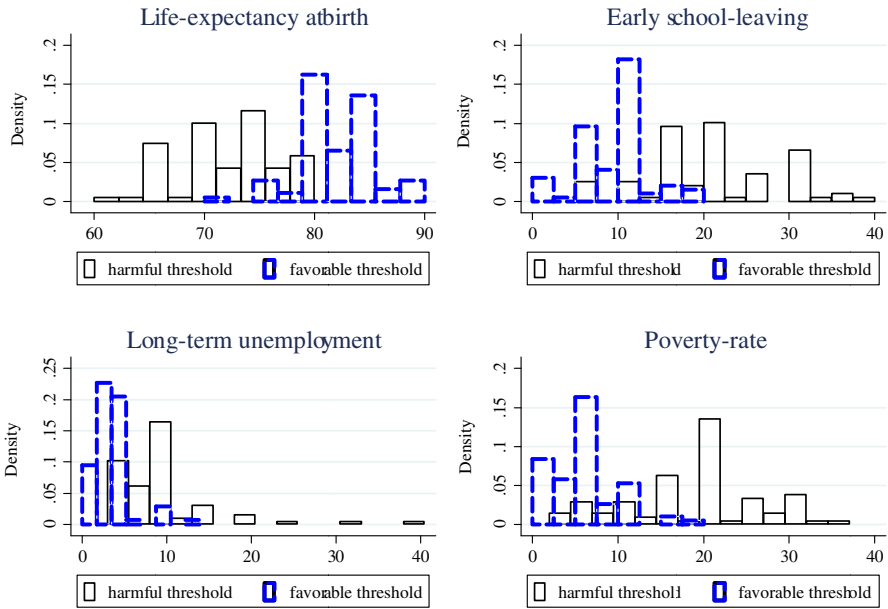
Fig. 10 Distribution of survey responses

were faculty members of the Department of Economics, 29 from the Department of Management. The following table provide brief descriptive statistics on our sample (Table 13).

Median responses and interquartile range are reported in Table 14,[20] while Fig. 10 illustrates the histograms for the responses' distribution. The blue thick-dashed lines represent the answers for the favourable thresholds.

---

[20] We chose the median response as a measure of central tendency to summarize a representative answer, as often done in the literature (e.g., Hoskins and Mascherini 2009) because of its lower sensitivity to outliers, especially when the sample size is small.

Appendix C: Results: Model with Data-Driven normalisation

The following coefficients are obtained by implementing the LD model (8) (Table 15).

Table 15 Aggregate measure of social inclusion, baseline model with data-driven normalisation

Nation	Region	2004	2005	2006	2007	2008	2009	2010	2011	2012
BE	Bruxelles	45.5	41.1	42.8	43.9	45.3	49.4	45.9	44.2	41.2
BE	Vlaams Gewest	73.2	73.2	75.8	78.5	79.9	80.4	80.0	82.4	82.4
BE	Région wallonne	51.9	50.7	52.5	54.0	53.9	55.9	56.3	57.6	57.7
DE	Baden-Württemberg	72.4	73.9	77.3	79.2	81.3	81.4	81.3	83.7	85.2
DE	Bayern	69.7	71.1	71.9	75.4	77.7	78.9	79.6	80.7	83.0
DE	Berlin	45.2	46.1	47.1	53.8	53.6	58.1	58.7	62.7	63.8
DE	Brandenburg	49.0	50.0	52.0	56.6	62.0	63.4	65.7	67.2	67.4
DE	Bremen	44.2	45.1	52.3	55.1	57.1	60.6	61.8	60.9	65.0
DE	Hamburg	57.4	58.6	63.4	64.0	69.5	71.6	72.5	72.8	75.0
DE	Hessen	65.8	67.1	69.4	71.1	72.9	74.9	76.7	78.5	79.5
DE	Mecklenburg-Vorpommern	38.1	38.9	44.8	47.3	53.7	55.4	57.8	60.2	60.2
DE	Niedersachsen	58.8	60.0	62.0	65.2	66.4	68.2	68.6	69.7	72.4
DE	Nordrhein-Westfalen	58.5	59.7	61.3	63.6	65.5	66.8	66.3	68.4	69.5
DE	Rheinland-Pfalz	61.2	62.4	65.5	67.9	70.8	70.1	70.0	72.5	74.5
DE	Saarland	51.3	52.3	53.8	62.1	62.8	63.6	66.8	65.3	69.1
DE	Sachsen	52.2	53.3	56.4	58.7	63.5	64.8	63.1	68.2	70.3
DE	Sachsen-Anhalt	40.8	41.6	45.5	48.9	49.5	53.4	57.1	57.4	57.8
DE	Schleswig-Holstein	61.4	62.7	65.2	67.9	68.4	70.3	69.9	72.3	73.4
DE	Thüringen	53.5	54.6	54.7	57.2	62.7	66.4	68.8	69.5	70.9
DK	Hovedstaden					69.2	69.5	71.4	67.6	73.4
DK	Sjælland					65.7	66.1	68.6	70.1	68.7
DK	Syddanmark					68.9	73.4	69.8	71.4	72.3
DK	Midtjylland					70.4	72.8	74.9	73.9	77.4
DK	Nordjylland					66.8	68.1	69.4	72.1	72.9
ES	Galicia	55.0	61.0	60.9	63.8	64.2	62.8	63.7	62.4	59.6
ES	Principado de Asturias	57.1	61.1	64.2	66.7	69.0	66.3	63.5	64.2	58.7
ES	Cantabria	61.1	68.4	69.7	69.3	73.9	71.3	64.1	65.4	70.2
ES	País Vasco	73.0	76.1	79.0	77.7	80.2	79.4	79.9	77.9	76.9
ES	Comunidad Foral de Navarra	74.4	78.2	83.8	85.1	83.4	83.1	85.5	84.8	81.9
ES	La Rioja	59	59.6	65.3	63.5	58.7	63.0	63.3	56.9	62.2
ES	Aragón	69.1	66.7	70.6	68.2	70.8	72.4	66.8	64.6	63.9
ES	Comunidad de Madrid	70.5	70.4	74.1	73.8	72.8	71.5	72.5	72.6	69.0
ES	Castilla y León	59.4	60.5	65.2	66.3	65.1	66.0	65.5	62.1	65.4
ES	Castilla-la Mancha	47.5	48.1	51.8	51.7	51.4	52.6	49.6	45.2	43.1
ES	Extremadura	33.5	35.8	40.7	39.9	43.7	42.8	37.6	43.8	31.9
ES	Cataluña	57.6	60.9	67.9	65.6	65.6	61.9	59.0	59.0	56.9

Table 15 continued

Nation	Region	2004	2005	2006	2007	2008	2009	2010	2011	2012
ES	Comunidad Valenciana	49.9	52.4	58.1	59.4	56.3	55.7	49.6	50.2	45.4
ES	Illes Balears	54.3	55.4	62.1	58.2	57.8	53.3	51.0	54.9	48.8
ES	Andalucía	32.9	39.5	41.0	43.6	41.7	39.9	36.7	33.3	31.8
ES	Región de Murcia	42.7	46.2	48.1	48.1	47.1	44.4	42.2	44.4	41.7
ES	Ciudad Autónoma de Ceuta		15.4	18.6	17.6	22.7	23.8	20.1	23.2	19.5
ES	Ciudad Autónoma de Melilla			33.7	34.7			38.1	33.3	21.8
ES	Canarias	45.3	46.4	46.5	47.2	46.6	45.1	37.3	32.3	33.1
IT	Piemonte	68.7	70.8	72.7	76.1	74.1	73.9	74.9	74.7	72.7
IT	VDA	71.8	71.0	76.8	70.8	72.5	76.3	77.3	77.6	78.0
IT	Liguria	73.8	71.4	74.9	75.7	78.3	80.9	78.7	78.2	73.9
IT	Lombardia	74.1	74.1	77.6	77.7	78.5	78.1	79.3	81.5	79.3
IT	TAA	76.5	79.1	82.3	85.1	85.9	87.0	87.5	88.7	83.9
IT	Veneto	78.0	77.0	81.0	82.1	81.6	83.0	82.3	81.9	82.2
IT	FVG	78.8	76.7	75.5	82.9	79.2	81.0	81.4	81.5	79.1
IT	ER	77.8	77.9	80.4	81.8	82.5	83.5	84.4	85.2	82.2
IT	Toscana	75.8	78.8	80.6	79.4	80.4	80.9	79.3	78.8	77.1
IT	Umbria	78.2	73.7	77.4	80.7	78.3	81.9	81.9	83.5	78.4
IT	Marche	79.2	76.3	78.6	82.4	82.3	82.0	84.3	82.7	78.0
IT	Lazio	66.5	68.4	71.2	74.4	74.9	74.3	74.2	70.4	69.7
IT	Abruzzo	69.4	69.8	73.1	71.2	73.2	70.0	72.4	71.9	69.3
IT	Molise	60.0	60.1	58.6	63.1	61.5	62.6	69.8	67.6	67.1
IT	Campania	35.8	35.9	39.4	38.4	40.0	43.4	42.8	40.0	36.5
IT	Puglia	42.7	42.9	44.5	50.2	56.2	56.7	58.4	56.4	55.5
IT	Basilicata	54.4	52.2	59.8	61.9	60.0	60.6	63.5	59.1	57.8
IT	Calabria	43.7	44.1	49.0	50.7	52.3	54.6	58.0	55.0	48.7
IT	Sicilia	31.9	31.9	37.3	37.9	41.9	38.8	42.2	38.5	34.7
IT	Sardegna	50.2	47.2	56.0	62.2	58.7	59.4	64.8	57.6	55.2
	Belgium	63.5	62.6	64.8	67.0	67.9	69.3	68.8	70.4	70.1
	Germany	61.1	61.7	63.9	66.7	69.0	70.5	70.9	72.8	74.9
	Denmark	–	–	–	68.7	71.6	71.7	70.7	73.4	74.4
	Spain	53.2	56.1	59.8	60.1	59.5	58.1	55.5	54.4	52.1
	Italy	62.4	62.6	65.9	67.5	68.4	68.8	69.8	68.6	66.2

Appendix D: Results: Model with Survey-Driven normalisation

The following coefficients are obtained by implementing the LS model (9) (Table 16).

Table 16 Social inclusion measure, baseline model with expert-based normalisation

Nation	Region	2004	2005	2006	2007	2008	2009	2010	2011	2012
BE	Bruxelles	24.9	17.0	17.7	18.3	17.4	31.4	22.5	22.8	18.5
BE	Vlaams Gewest	79.6	80.0	82.8	84.4	85.9	86.3	86.5	88.5	88.0
BE	Région wallonne	41.3	38.8	40.3	42.4	41.9	45.5	45.3	46.3	46.4
DE	Baden-Württemberg	76.2	77.8	83.1	84.4	87.3	86.7	86.7	87.4	87.8
DE	Bayern	76.3	77.9	76.0	82.5	84.3	84.3	85.3	85.0	85.7
DE	Berlin	23.5	24.0	25.8	34.2	30.7	39.3	39.6	49.0	51.3
DE	Brandenburg	38.3	39.1	40.6	45.7	56.0	58.5	64.4	65.7	65.1
DE	Bremen	20.1	20.5	28.7	37.1	44.7	51.7	53.1	48.3	60.5
DE	Hamburg	40.2	41.0	51.7	51.3	65.4	73.5	72.9	70.8	73.2
DE	Hessen	62.6	63.9	65.4	71.2	75.2	78.9	79.4	81.4	82.6
DE	Mecklenburg-Vorpommern	31.4	32.0	35.5	38.5	44.8	47.6	52.0	56.3	54.6
DE	Niedersachsen	49.9	50.9	53.8	60.1	63.8	68.0	65.9	65.2	70.7
DE	Nordrhein-Westfalen	49.3	50.3	51.3	56.9	61.0	64.1	61.6	63.8	66.4
DE	Rheinland-Pfalz	53.9	55.0	60.5	65.1	71.2	68.4	63.9	68.4	72.0
DE	Saarland	34.3	35.0	37.2	59.7	59.8	59.8	66.8	59.5	67.0
DE	Sachsen	41.7	42.6	45.0	43.4	51.0	53.7	55.8	60.6	65.3
DE	Sachsen-Anhalt	34.6	35.3	34.8	34.5	34.5	43.4	47.3	45.4	46.3
DE	Schleswig-Holstein	55.5	56.6	59.8	67.6	68.5	72.5	70.7	71.6	72.5
DE	Thüringen	39.1	39.9	41.9	43.8	56.0	61.1	65.8	68.7	71.1
DK	Hovedstaden					74.8	72.4	76.1	68.9	80.2
DK	Sjælland					65.8	68.9	74.1	77.8	73.8
DK	Syddanmark					68.0	77.3	68.8	74.4	76.3
DK	Midtjylland					69.8	72.8	77.9	76.4	81.6
DK	Nordjylland					66.3	69.5	71.3	78.1	77.5
ES	Galicia	32.5	44.7	45.3	46.3	46.5	48.0	42.2	32.9	29.6
ES	Principado de Asturias	45.7	45.3	55.2	58.4	56.5	52.4	46.2	45.3	28.9
ES	Cantabria	48.1	54.9	59.2	59.0	61.7	59.2	42.3	38.5	52.6
ES	País Vasco	72.8	74.6	76.8	73.3	76.8	77.3	79.1	72.3	69.4
ES	Comunidad Foral de Navarra	59.4	67.3	80.6	80.5	74.5	71.2	78.5	81.5	74.1
ES	La Rioja	44.6	45.5	48.3	47.2	48.0	50.3	42.6	34.7	31.0
ES	Aragón	58.8	49.4	57.8	51.5	55.4	61.8	47.7	39.8	31.1
ES	Comunidad de Madrid	64.0	59.1	62.5	61.0	58.3	57.7	48.7	42.2	36.1
ES	Castilla y León	43.1	46.8	48.5	48.3	49.0	48.3	41.3	34.7	32.3
ES	Castilla-la Mancha	45.8	45.5	47.8	46.8	47.8	44.3	33.4	25.8	25.8
ES	Extremadura	35.5	36.4	43.1	43.3	43.9	37.2	26.3	23.3	22.5
ES	Cataluña	53.8	55.6	57.8	57.1	58.5	51.9	37.1	29.3	30.3

Table 16 continued

Nation	Region	2004	2005	2006	2007	2008	2009	2010	2011	2012
ES	Comunidad Valenciana	42.5	42.2	44.9	48.6	45.5	42.3	23.5	23.3	23.5
ES	Illes Balears	54.0	49.2	57.2	53.9	55.6	50.7	35.0	31.3	23.5
ES	Andalucía	30.5	38.4	42.5	42.5	41.2	30.5	20.5	21.0	21.0
ES	Región de Murcia	42.2	42.5	44.3	44.3	44.5	40.8	25.7	23.3	23.0
ES	Ciudad Autónoma de Ceuta		15.0	19.9	14.8	26.0	16.6	15.8	17.5	20.0
ES	Ciudad Autónoma de Melilla			30.6	25.8			24.5	19.0	19.5
ES	Canarias	39.2	41.6	43.5	43.8	42.1	28.8	23.3	22.5	22.8
IT	Piemonte	56.8	60.1	60.4	68.3	63.8	63.3	64.3	65.9	59.8
IT	VDA	61.0	65.8	70.4	64.1	63.9	64.8	66.7	68.6	69.7
IT	Liguria	68.9	60.8	66.7	65.1	75.8	83.5	73.8	73.5	64.1
IT	Lombardia	62.8	63.8	66.8	65.9	66.0	65.8	69.3	74.8	73.3
IT	TAA	64.7	67.8	75.7	79.2	80.4	78.9	78.3	82.1	73.0
IT	Veneto	69.4	67.7	76.7	79.9	75.0	74.7	75.8	73.3	79.5
IT	FVG	80.7	73.4	63.3	83.1	74.1	77.4	81.8	81.0	76.8
IT	ER	67.0	67.7	71.2	73.6	75.5	79.7	82.3	84.9	79.9
IT	Toscana	64.7	72.9	75.4	69.8	74.5	73.8	70.0	67.2	66.4
IT	Umbria	76.7	68.3	68.7	78.1	69.8	78.4	79.7	81.8	72.8
IT	Marche	70.7	61.8	64.9	73.8	76.0	74.8	76.3	77.3	65.8
IT	Lazio	56.7	61.5	66.8	73.0	70.6	69.9	65.9	55.7	57.4
IT	Abruzzo	57.5	58.0	64.8	60.3	62.0	58.7	60.3	61.8	55.7
IT	Molise	45.8	48.7	45.8	53.5	49.1	50.9	60.5	56.2	60.1
IT	Campania	20.6	18.4	25.4	30.4	27.3	26.2	23.3	20.0	20.3
IT	Puglia	21.7	26.1	30.0	35.0	36.5	36.3	33.4	32.7	26.3
IT	Basilicata	36.7	35.6	46.1	53.2	50.4	55.4	43.8	45.1	43.4
IT	Calabria	24.7	26.6	30.4	33.3	37.3	41.6	44.1	34.4	29.8
IT	Sicilia	19.8	19.0	25.2	26.1	26.2	24.3	25.4	25.3	21.8
IT	Sardegna	29.0	28.9	36.1	41.0	35.5	35.3	42.3	31.8	24.0
	Belgium	61.6	60.2	62.3	64.0	64.6	67.4	66.6	68.0	67.3
	Germany	55	56.2	58.2	63.0	67.0	69.5	69.4	70.8	73.8
	Denmark	–	–	–	70.0	73.9	74.1	74.1	78.4	79.2
	Spain	46.9	48.8	52.2	51.9	51.5	47.0	36.4	32.2	30.3
	Italy	50.3	51.0	55.4	58.4	57.8	58.0	58.1	57.1	54.5

References

Anand S, Sen A (1994) Human development index: methodology and measurement. Human Development Report Office (HDRO), United Nations Development Programme (UNDP). Published as background research for the 1994 Human Development Report

Atkinson AB, Cantillon B, Marlier E, Nolan B (2002) Social indicators: The EU and social inclusion: The EU and social inclusion. Oxford University Press, Oxford

Atkinson AB, Marlier E, Nolan B (2004) Indicators and targets for social inclusion in the European Union. JCMS 42(1):47–75

Blackorby C, Donaldson D (1982) Ratio-scale and translation-scale full interpersonal comparability without domain restrictions: Admissible social-evaluation functions. Int Econ Rev 23(2):249–268

Boarini R, D'Ercole MM (2013) Going beyond GDP: An OECD Perspective*. Fiscal Stud 34(3):289–314. doi:10.1111/j.1475-5890.2013.12007.x

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

Cherchye L, Knox Lovell CA, Moesen W, Van Puyenbroeck T (2007) One market, one number? A composite indicator assessment of EU internal market dynamics. Eur Econ Rev 51(3):749–779. doi:10.1016/j.euroecorev.2006.03.011

Chowdhury S, Squire L (2006) Setting weights for aggregate indices: An application to the commitment to development index and human development index. J Dev Stud 42(5):761–771

Decancq K, Lugo MA (2008) Setting weights in multidimensional indices of well-being. OPHI working papers, Queen Elizabeth House, University of Oxford

Decancq K, Lugo MA (2009) Measuring inequality of well-being with a correlation-sensitive multidimensional Gini index. University of Oxford Department of Economics Discussion Paper Series, vol 459

Decancq K, Lugo MA (2013) Weights in multidimensional indices of wellbeing: an overview. Econ Rev 32(1):7–34

European Commission (2009) Portfolio of indicators for the monitoring of the European strategy for social protection and social inclusion Bruxelles. Brussels: European Commission—Employment, social affairs and equal opportunities DG social protection and social integration, September 2009

European Communities Commission (1992) Towards a Europe of solidarity. Intensifying the fight against social exclusion, fostering integration. Communication from the Commission. COM(92) 542 final. Commission of the European Communities, Bruxelles

European Council (2001) Laeken European Council. Presidency conclusions and annexes. Bulletin 17.12.2001, PE 313.424

Eurostat (2010) Combating poverty and social exclusion. A statistical portrait of the European Union. Publications Office of the European Union, Luxembourg

Giovannini E, Nardo M, Saisana M, Saltelli A, Tarantola A, Hoffman A (2008) Handbook on constructing composite indicators: methodology and user guide

Hoskins B, Mascherini M (2009) Measuring active citizenship through the development of a composite indicator. Soc Indic Res 90(3):459–488. doi:10.1007/s11205-008-9271-2

Kasparian J, Rolland A (2012) OECD's 'Better Life Index': can any country be well ranked? J Appl Stat 39(10):2223–2230

Kim Y, Kee Y, Lee S (2015) An analysis of the relative importance of components in measuring community wellbeing: perspectives of citizens, public officials, and experts. Soc Ind Res 121(2):345–369. doi:10.1007/s11205-014-0652-4

Klugman J, Rodríguez F, Choi H-J (2011) The HDI 2010: new controversies, old critiques. J Econ Inequal 9(2):249–288

Lefebvre M, Coelli T, Pestieau P (2010) On the convergence of social protection performance in the European Union. CESifo Econ Stud 56(2):300–322

Maggino F, Nuvolati G (2012) Quality of life in Italy: research and reflections, (vol 48). Springer, New York

Martinetti EC, von Jacobi N (2012) Light and shade of multidimensional indexes: how methodological choices impact on empirical results. Quality of life in Italy: research and reflections. Social Indicators Research Series, p 48

Mascherini M, Hoskins B (2008) Retrieving expert opinion on weights for the Active Citizenship Composite Indicator. European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen

Mazziotta M, Pareto A (2015) On a Generalized non-compensatory composite index for measuring socio-economic phenomena. Soc Indic Res:1–21. doi:10.1007/s11205-015-0998-2

Meyer P, Ponthière G (2011) Eliciting preferences on multiattribute societies with a Choquet integral. Comput Econ 37(2):133–168

Murias P, Novello S, Martinez F (2012) The regions of economic well-being in Italy and Spain. Reg Stud 46(6):793–816

Nguefack-Tsague G, Klasen S, Zucchini W (2011) On weighting the components of the human development index: a statistical justification. J Human Dev Capab 12(2):183–202

Pestieau P (2009) Assessing the performance of the public sector. Ann Public Coop Econ 80(1):133–161

Pinar M, Cruciani C, Giove S, Sostero M (2014) Constructing the FEEM sustainability index: a Choquet integral application. Ecol Ind 39:189–202

Ravallion M (2011) On multidimensional indices of poverty. J Econ Inequal 9(2):235–248

Ravallion M (2012a) Mashup indices of development. World Bank Res Obs 27(1):1–32

Ravallion M (2012b) Troubling tradeoffs in the human development index. J Dev Econ 99(2):201–209

Saisana M, Saltelli A, Tarantola S (2005) Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. J R Stat Soc Ser A (Stat Soc) 168(2):307–323

Sen A, Anand S (1997) Concepts of human development and poverty: a multidimensional perspective. In: Poverty and human development: human development papers. United Nations Development Programme, New York, pp 1–20

Silva R, Ferreira-Lopes A (2013) A regional development index for Portugal. Soc Indic Res 118(3):1055–1085

Stiglitz JE, Sen A, Fitoussi JP (2010) Report by the commission on the measurement of economic performance and social progress. Commission on the Measurement of Economic Performance and Social Progress, Paris