Are governments matching citizens' demand for better lives? A new approach comparing subjective and objective welfare measures

This paper proposes two composite indices of well-being to examine the gap between a government's welfare outcome and its citizens' desired well-being and how this "welfare gap" may influence utility.

Luisa Corrado
University of Rome Tor Vergata

Giuseppe De Michele
European Stability Mechanism

European Stability Mechanism

Are governments matching citizens' demand for better lives? A new approach comparing subjective and objective welfare measures

Luisa Corrado[1] University of Rome Tor Vergata

Giuseppe De Michele[2] European Stability Mechanism

Abstract

We propose a new approach which helps to shed light on the importance of the relationship between a government's welfare outcome and its citizens' desired well-being, defining a concept of "welfare gap". To determine this gap, we build two composite indices of well-being measured at the individual and aggregate level - i.e. subjective and objective welfare measures - assessing overall well-being and its progress over time. To this end, we apply idiosyncratic settings of Structural Equation Models to examine the interrelations and causal relationships across welfare determinants and among the underlying drivers of well-being. By comparing the dimensions' weights and rankings of the objective and subjective welfare measures, we obtain largely opposite results in both analyses, except for the relevance of the health status. Material living conditions are the most important dimensions in the objective ranking, whilst the quality of life indicators lie at the top of the subjective ladder. Moreover, the distance between subjective welfare aspirations and objective outcomes described through the "welfare gap" measure could contribute to explaining the anti-establishment sentiment recently observed in different societies.

Keywords: Structural Equation Modeling, Latent Multidimensional Index, Beyond GDP, Utility Function, Objective Welfare Index, Subjective Welfare Index, Stated Preference, Generalised SEM MIMIC, GSEM, Bootstrapped SEM, Small Sample Size, Weights.

JEL codes: C43, C83, D12, D63, E21, E24, I31, I38, O57

[1] Email: luisa.corrado@uniroma2.it

[2] Email: g.demichele@esm.europa.eu

ISSN 2443-5503
ISBN 978-92-95085-70-1

doi:10.2852/98429
EU catalog number DW-AB-19-005-EN-N

Are governments matching citizens' demand for better lives? A new approach comparing subjective and objective welfare measures*

Luisa Corrado† and Giuseppe De Michele‡

August 5, 2019

Abstract

We propose a new approach which helps to shed light on the importance of the relationship between a government's welfare outcome and its citizens' desired well-being, defining a concept of "welfare gap". To determine this gap, we build two composite indices of well-being measured at the individual and aggregate level - i.e. subjective and objective welfare measures - assessing overall well-being and its progress over time. To this end, we apply idiosyncratic settings of Structural Equation Models to examine the interrelations and causal relationships across welfare determinants and among the underlying drivers of well-being. By comparing the dimensions' weights and rankings of the objective and subjective welfare measures, we obtain largely opposite results in both analyses, except for the relevance of the health status. Material living conditions are the most important dimensions in the objective ranking, whilst the quality of life indicators lie at the top of the subjective ladder. Moreover, the distance between subjective welfare aspirations and objective outcomes described through the "welfare gap" measure could contribute to explaining the anti-establishment sentiment recently observed in different societies.

JEL classification: C43, C83, D12, D63, E21, E24, I31, I38, O57.

Keywords: Structural Equation Modeling, Latent Multidimensional Index, Beyond GDP, Utility Function, Objective Welfare Index, Subjective Welfare Index, Stated Preference, Generalised SEM MIMIC, GSEM, Bootstrapped SEM, Small Sample Size, Weights.

*The views expressed in this paper are those of the authors and should not be attributed to the European Stability Mechanism.

†University of Rome Tor Vergata, Department of Economics and Finance. E-mail: luisa.corrado@uniroma2.it.

‡European Stability Mechanism. E-mail: g.demichele@esm.europa.eu.

1 Introduction

For more than sixty years the Gross Domestic Product[1] (GDP) has been the benchmark used to measure nations' and people's welfare. The GDP proved to be an effective measure of market-based economic activity and wealth creation but it is only a rough indicator of social welfare and progress. In particular, it fails to capture some of the non-economic factors that make a difference in people's lives, such as security, social relationships, income distribution and the quality of the environment. Moreover, GDP is very limited in accounting for elements which make economic growth sustainable.

A wide gap between the official statistics and people's perceptions contributes to a lack of confidence in those who produce and rely on these figures. Aware of these shortfalls, economists, statisticians and policy makers have devoted their efforts to develop broader measures of well-being. Producing better and more realistic ways of measuring economic, environmental and social performances, is also a critical step in improving the effectiveness of governments' action in matching citizens' welfare aspirations. A second key element is welfare measurement both at the individual and aggregate level. In the last two decades there have been many discussions on how to move 'beyond GDP' with a growing consensus that measuring well-being requires considering broader dimensions (economic and non-economic) of people's achievements and opportunities.
The discussion and research on welfare measures have found expression in some major initiatives, such as the report of the Commission on the Measurement of Economic Performance and Social Progress, the so-called Stiglitz–Sen–Fitoussi Commission, in 2009; the EU Communication (and follow-up action) on 'GDP and Beyond' in the same year, and; the OECD Better Life Initiative, launched in 2011. Along with these international initiatives, a large number of national actions have flourished in the form of public national consultations, involving stakeholders and civil society (i.e. in Australia, Belgium, Finland, France, Germany, Italy, Mexico and the United Kingdom). In some countries the government - and sometimes the parliament - actively engaged in this process: this is the case for Australia (Well-being Framework), Finland (National Strategy far Sustainable Development; Findicator), France (Les Nouveaux Indicateurs de Richesse), Germany (National Sustainable Development Strategy; W3-Indikatoren), Italy (Indicatori di Benessere Equo e Sostenibile) and the United Kingdom (Measuring National Well-being Programme).[2]

In the same period, welfare economics has proliferated in various directions, involving the theory of fair allocation, the theory of social choice, the capability approach, the study of happiness and the psychology of well-being (Fleurbaey and Blanchet, 2013). Starting from the distinction between monetary and nonmonetary aggregates, Fleurbaey (2009) classified four alternative approaches to the measurement of social welfare beyond GDP: i) 'Corrected GDP', that would take into account non-market aspects of well-being[3]; ii) 'Gross National Happiness', which has been

[1]The Gross Domestic Product, the core concept within the System of National Accounts (SNA), measures the aggregate value of economic production in a given year and in a given country.

[2]Among those national efforts to measure wider dimensions of progress, it is worth mentioning the ongoing 'Measuring national well-being' programme of the UK Office for National Statistics (ONS), launched in 2010, for its scientific importance and potential impact. With reference to Italy, in 2010 the Italian National Statistical Office (ISTAT) and the National Council on the Economy and Labour (CNEL) launched the "Equitable and Sustainable Well-being" project (BES –*Benessere Equo e Sostenibile*) with the aim of defining a measurement framework to assess people's welfare in the country. The first BES report was published in 2013 by ISTAT. Building on this, in 2017 the Italian government, as provided by the Law n.163/2016, integrated a set of twelve welfare indicators into policy making, through the public finance process. Italy is the first OECD country to insert well-being indicators in the official budget cycle and to introduce forecasts on a set of twelve selected variables, in order to assess the impact of Government's economic policy measures on different aspects of citizens' lives beyond GDP.

[3]The idea of correcting GDP has been often interpreted, after Nordhaus and Tobin's (1973) seminal work, as adding or subtracting aggregates similar to GDP. In this context, monetary aggregates are valued at market prices

revived by the growing relevance of well-being studies; iii) the 'Capability approach', developed by Sen (1985, 1992), which has become an important reference in the field of alternative indicators to GDP, inspiring a variety of applications; iv) 'Synthetic indicators' that, following the lead of the UNDP Human Development Index (HDI), are constructed as weighted averages of indices of social performance in various domains (Noorbakhsh, 1998). The Better Life Index (BLI) developed by the OECD falls within the last category. More specifically, the OECD BLI framework looks beyond the purely economic aspects of welfare, referring to it as a truly multidimensional concept[4] that addresses the critical limitations of GDP as a welfare metric (Boarini and Mira d'Ercole, 2013).
In our analysis, we propose an innovative utilitarian theoretical approach based on the comparison between objective and subjective welfare measures. We make reference to the multidimensional definition of well-being proposed by the OECD in its Better Life Initiative (OECD, 2011; 2013; 2015; 2017) in order to concretely define these two measures. To this purpose, we utilize two different comparable OECD datasets for the year 2012, one based on average country-level macrodata reflecting welfare outcomes, the other one on microdata reflecting people's stated preferences on welfare domains. We then build an 'objective' welfare measure predicted from the national-level data, whereas a 'subjective' welfare measure is obtained using the new OECD microdata. The construction of these two comparable indices allow us to test the empirical implications from our theoretical model where we assume that individual utility is affected not only by what is desirable for people in terms of subjective welfare but also by the divergence between (average) individual welfare measures and aggregate welfare outcomes achieved by governments.

The methodology at the basis of several composite indices is often silent on the relative weighting of the indicators used to define a single measure of social welfare. In practice, it happens that ad hoc weights often end up being applied implicitly by users or explicitly in published indices, without any in-depth analysis on this topic (Benjamin et al., 2014). The selection of the relative weights for the different dimensions is a crucial step in the construction of a multidimensional index of well-being. The most common approach to weight multidimensional indices of well-being is equal or arbitrary weighting. Equal weighting has often been defended on the ground that all indicators are equally important or by the recognition of an agnostic viewpoint.[5] In order to obtain the objective and subjective welfare measures, as described above, in our work we propose a Structural Equation Modeling (SEM) approach to endogenously estimate the BLI dimension's weights, considering all the available information on the underlying indicators simultaneously.[6] A major point in our analytical strategy is that the SEM approach accounts for all the possible correlations among indicators included in the model since it is based on the analysis of the empirical and estimated population's variance-covariance matrices. This feature allows to overcome one of the major critiques to the social indices, by which they would not account for the covariances of the correlated dimensions of

or at imputed prices if market prices are not available. Within this monetary approach there are also the "Green" accounting and the Net National Product (NNP). Furthermore, a more promising approach for the incorporation of non-monetary aspects of quality of life involves equivalent incomes, in which income can be added and subtracted, reflecting people's willingness-to-pay.

[4]With regard to the multidimensionality, the OECD selected for its BLI a set of eleven underlying dimensions of well-being - ranging from income, jobs and housing to health, education and environment - which people all over the world consider relevant for their quality of life (see Table 6 - Appendix I).

[5]A primal example of equal weighting is the Human Development Index. It is argued that the main motivation for using equal weighting is that three dimensions are deemed equally important. Also the OECD Better Life Index (BLI) adopts equal weights for its eleven underlying dimensions, within a normative approach.

[6]Following the classification proposed by Decanqu and Lugo (2013), our work falls in the class of data-driven weights (cfr. note 6). Notably, more complex explanatory approaches include multiple indicator and multiple causes models (MIMIC) and structural equation models (SEM), which are the methods we apply in this analysis in two innovative forms.

well-being. Through an SEM estimation we can, therefore, obtain better estimates of the weights of the well-being dimensions underlying the multidimensional indices. Within this framework, our work allows us to obtain the objective and subjective welfare measures as two latent constructs, starting from eleven underlying well-being dimensions, and to endogenously estimate the relative weights of those indicators.

This paper is organised as follows. Section two presents the theoretical set-up which shows the interplay between objective and subjective welfare measures and how the discrepancy between them may influence individual utility. In that regard, we consider an environment where individuals state their preferences corresponding to a given level of individual welfare, while governments achieve an aggregate welfare outcome that may or may not coincide with the individual expectations. We also provide a discussion on the empirical method proposed to define the multidimensional welfare indices starting from the macro and microdata related to the underlying welfare dimensions. To this purpose, we gather two data sources by utilizing two different comparable OECD datasets for the year 2012, one based on average country-level macrodata reflecting well-being outcomes, the other one on microdata reflecting people's stated preferences on well-being indicators.

In Section three, we estimate an objective welfare measure. To this end, we consider a Structural Equation Modeling approach combined with a bootstrapping technique to deal with the non-normal distribution of the error terms induced by the small size of the sample utilized. A bootstrapped SEM is applied to estimate an objective welfare index using average country-level data. Given the limited number of observations included in the dataset, we also apply the power analysis to check if our sample of 35 observations and 24 variables is in line with the minimum sample size requirements needed for an appropriate SEM analysis. Moreover, in order to cope with the presence of missing values and to retain all the available information, we did not apply the default listwise deletion but a specific Maximum Likelihood Missing Value estimation method ($MLMV$) within SEM. From the proposed bootstrapped SEM $MLMV$ method we can accurately estimate the relative standardized 'objective' weights and ranking for the eleven dimensions underlying the objective welfare measure. Then, we use these estimates to predict a single objective welfare measure for each country or macroregion and to rank them, as shown in the final section. This framework aims at measuring well-being outcomes, i.e. looking at whether countries are doing well or not, benchmarking against each other, and are making progress over time, in line with the standard approach in this area.

One of the most novel aspects of our work, presented in Section four, is to combine the SEM and the MIMIC model in a single method in order to estimate a 'subjective' welfare index. Then, a distinct feature is to put the SEM method in a Generalized form to deal with non-normality implied by the Likert-type scale of the microdata used for the model estimation. Since data are categorical, we use an ordered probit link function. Besides the BLI, the OECD launched a recent parallel complementary process, Your Better Life Initiative, with the aim of involving people and gathering individual stated preferences on what matters most for their quality of life.[7] Within this unique and wide OECD dataset, which has never been used for complex econometric analysis so far, we selected microdata for 35 countries – 33 OECD and 2 emerging economies – for the year 2012. Apart from the preferences on the eleven welfare determinants, the dataset consisting of more than 12,000 observations includes also individual observations on four geo-demographic variables influencing well-being. In order to account for these control variables, we applied a Multiple Indicators Multiple Causes (MIMIC) model under GSEM to estimate the 'subjective' relative weights and the ranking of the eleven well-being drivers.

[7] The enquiry is carried out through a specific tool, available on the official website *oecdbetterlifeindex.org*, which allows the creation and sharing of the individual rankings and the relative weights of the eleven dimensions underlying BLI.

Finally, in Section five we propose a comparison between the weights of the objective and subjective welfare dimensions. A second comparison, based on country's rankings defined on the basis of the predicted welfare scores, is meant to identify the societal gaps between the objective and subjective welfare measures. We conclude with a brief discussion on the policy implications of the results obtained.

2 Understanding welfare determinants

Multidimensional indices are becoming increasingly important measures to assess social well-being. The idea that well-being is inherently multidimensional[8] is now firmly rooted in the academic and policy-oriented literature (Stiglitz et al., 2009). Those composite indices have the important role of complementing other established measures, such as GDP per capita or life expectancy. One of the reasons why GDP per capita has predominated for so long despite its limitations as a welfare metric, is that it enables observers to monitor nations' economic well-being through one single headline number. Composite indices of welfare measured at the individual and aggregate level also make it possible to assess overall well-being and its progress over time. In this respect, the existing literature had so far a dual approach. On the one hand, some economists either rely on revealed preference indirectly, evaluating policy options by how they affect objective composite indicators that can be viewed as summarizing, under some assumptions, a set of generally-desired government outcomes (for a recent survey, see Fleurbaey, 2009). On the other side, more recent research aims at determining individual-level composite indices that combines together different aspects of well-being that may be measured by stated preferences in survey questions, using the responses to calculate indicators (Benjamin et al, 2012, 2014). Our analysis goes one step further by defining matched realizations of individual and government welfare indicators, defined over the same set of domains, and investigating how the discrepancy between objective and subjective measures affects individual and social welfare. Indeed, we share the fundamental idea that, in addition to economic dimensions, non-economic factors affect welfare. Moreover, any welfare discrepancy between individual *desiderata* and government outcomes, also plays a crucial role in determining utility and social welfare. It is desirable for governments to maximise social welfare evaluated according to citizens' own preferences. A key goal for governments is to achieve a reduction in - and potentially the elimination of - the gap between objective and (average) subjective welfare measures in order to maximise social welfare.

2.1 Theory and main setup

We consider an environment where individuals state their preferences for a given level of individual welfare, η_i, while governments achieve an aggregate welfare outcome, η_{-i}, that may or may not coincide with the individual level, η_i.

Individual welfare is defined by the following utility function:

$$U_i = U(\eta_i) + U\left(\frac{\eta_{-i}}{\eta_i}\right) = \alpha \log \eta_i + \beta \log\left(\frac{\eta_{-i}}{\eta_i}\right) \quad \text{with} \quad \alpha, \beta > 0 \tag{1}$$

[8]Philosophers such as Rawls (1971), Sen (1985) and Nussbaum (2000) support the multidimensional perspective in the notion of well-being. Moreover, the rapidly emerging literature on welfare determinants shows that people's overall satisfaction is affected by many monetary and non-monetary aspects of life, such as their health, employment status, income and marital status (Kahneman and Krueger, 2006).

i.e. utility depends positively on the individual welfare levels corresponding to their own stated preferences on well-being dimensions,[9] η_i. When governments do not match individual preferences ($\frac{\eta_{-i}}{\eta_i} < 1$) utility is reduced by an amount proportional to the negative gap between a government's welfare outcome and (aggregate) individual welfare levels as per people's stated preferences, $\beta \log\left(\frac{\eta_{-i}}{\eta_i}\right) < 0$. The "welfare gap" links what is 'desirable' for people in terms of individual welfare, η_i, to what government policies achieve in reality, η_{-i}. When governments fulfill individual welfare ($\frac{\eta_{-i}}{\eta_i} > 1$), individuals derive extra utility from the positive welfare gap, $\beta \log\left(\frac{\eta_{-i}}{\eta_i}\right) > 0$. If governments exactly meet individuals' welfare levels $\frac{\eta_{-i}}{\eta_i} = 1$ individual utility just equals $\alpha \log \eta_i$. The last nonlinear term delivers an asymmetric response to government outcomes, therefore, if the individual welfare *desiderata* are not met by the government, utility decreases to a larger degree than what it would increase when aspirations are met.

We also assume that both individual, η_i, and aggregate welfare measures, η_{-i}, are latent variables - i.e. they are not observed - and are a function of a set of J different domains.[10] Specifically, the subjective welfare statistic, η_i, is defined as a function on a set of J domains whose preferences are stated at the individual level:

$$\eta_i = \left(y_{i1}\left(\mathbf{x}_i\right), y_{i2}\left(\mathbf{x}_i\right), ..., y_{iJ}\left(\mathbf{x}_i\right)\right) \tag{2}$$

where y_{ij} is the stated preference of the i-th individual on the j-th domain with $j = 1, ..., J$ and $\mathbf{x}_i$ are the individual characteristics that affect the preferences for the specific domain. Each individual will assign his own weights to the various domains that make up subjective welfare, η_i. The weights $\mathbf{\Lambda}^S = \left[\Lambda_1^S, \Lambda_2^S, ..., \Lambda_J^S\right]'$ attached to the set of domain indicators $\mathbf{y}_i = \left[y_{i1}, y_{i2}, ..., y_{iJ}\right]'$, are chosen so that any utility maximizing individual equalizes the marginal welfare in each domain:[11]

$$\frac{\partial \eta_i}{\partial y_{i1}} \frac{\partial y_{i1}}{\partial x_i} = \frac{\partial \eta_i}{\partial y_{i2}} \frac{\partial y_{i2}}{\partial x_i} = \cdots = \frac{\partial \eta_i}{\partial y_{iJ}} \frac{\partial y_{iJ}}{\partial x_i} \tag{3}$$

We now turn to the definition of the aggregate welfare statistic. The (aggregate) objective welfare measure as achieved by the government, η_{-i}, is defined as a function on the same set of $j = 1, ..., J$ domains:

$$\eta_{-i} = \left(y_{-i1}, y_{-i2}, ..., y_{-iJ}\right) \tag{4}$$

where y_{-ij} is the outcome of the i-th country/government on the j-th domain at the aggregate level. The weights $\mathbf{\Lambda}^o = \left[\Lambda_1^o, \Lambda_2^o, ..., \Lambda_J^o\right]'$ implied by the set of domain indicators $\mathbf{y}_{-i} = \left[y_{-i1}, y_{-i2}, ..., y_{-iJ}\right]'$ are chosen so that any welfare maximizing government equalizes the marginal utility in each domain:

$$\frac{\partial \eta_{-i}}{\partial y_{-i1}} = \frac{\partial \eta_{-i}}{\partial y_{-i2}} = \cdots = \frac{\partial \eta_{-i}}{\partial y_{-iJ}} \tag{5}$$

Hence, individual utility can be expressed as a function of subjective and objective welfare measures which are a function of the various domain indicators and related weights:

$$U_i = \alpha \log \eta_i \left[\mathbf{\Lambda}^S\left(\mathbf{y}_i\right), \mathbf{x}_i\right] + \beta \log \frac{\eta_{-i}\left[\mathbf{\Lambda}^o\left(\mathbf{y}_{-i}\right)\right]}{\eta_i\left[\mathbf{\Lambda}^S\left(\mathbf{y}_i\right), \mathbf{x}_i\right]} \tag{6}$$

[9] Utility is increasing in η_i $\left(\partial U_i / \partial \eta_i = (\alpha - \beta)/\eta_i > 0\right)$ and concave $\left(\partial^2 U_i / \partial \eta_i^2 = -(\alpha - \beta)/\eta_i^2 < 0\right)$ when $\alpha > \beta$.

[10] The $J = 11$ indicators (or dimensions) used by the OECD to define its BLI as a latent construct and included in the model are listed in Table 6 - Appendix I (see paragraph 2.2 for details).

[11] In the computation of the marginal utilities $\frac{\partial U_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial y_{i1}} \frac{\partial y_{i1}}{\partial x_i} = \cdots = \frac{\partial U_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial y_{iJ}} \frac{\partial y_{iJ}}{\partial x_i}$ the term $\frac{\partial U_i}{\partial \eta_i}$ appears identically in the partial derivative of each domain, thus we omit this term in (3).

Individuals' utility is affected by what is desirable for people in terms of subjective welfare, $\eta_i[.]$ and, if these preferences are not matched by the relevant government's outcomes, by the distance between subjective and and (aggregate) objective welfare measures $\frac{\eta_{-i}[.]}{\eta_i[.]} \gtrless 1$. Individual welfare $\eta_i[.]$ is a function of a set of stated preferences on the various domains, $\mathbf{y}_i$, and also depends on individual characteristics, $\mathbf{x}_i$. Individual characteristics, $\mathbf{x}_i$, affect people's preferences over the different domains, $\mathbf{y}_i$, and may lead to a level of subjective welfare η_i that may differ from the aggregate (objective) counterpart η_{-i}. Therefore, one reason why objective and subjective welfare measures differ across countries is because individuals attach different preference weights to each domain with respect to their governments. Moreover, the demographic composition of each country may have an impact in determining such a gap because certain categories attach a higher weight to certain indicators, independently from the aggregate (objective) welfare outcome, possibly because they are more keen to a specific domain given their individual characteristics (i.e. gender, age). For some indicators, we know that the link with the socio-demographic characteristics is evident. In this respect, research has found that females are more concerned about the quality of health services (Campbell, 2004). Becchetti et al. (2017), for example, find that females allocate more in health whilst males more in education. In addition, people with low income allocate more in economic well-being, whereas top earners show a higher preference for work, life balance and social relationships. Elders want to invest more in health whilst the youngsters want to invest more in economic well-being and social relationships. It is worth noting, though, that these quoted papers look at the determinants of political preferences (health, environmental concerns), whereas our work deals with all the dimensions of well-being in the definition of the two distinct latent measures of welfare, by identifying endogenously the relative weights that drive people's preferences and government outcomes. Therefore, the existence of a potential gap between objective and subjective welfare measures may reside in the different weight people attach to the domains that make up their own subjective well-being level.

2.2 Building multidimensional objective and subjective welfare measures

In order to define concretely η_i and η_{-i}, we adopt the multidimensional definition of well-being drawing from the framework of the OECD Better Life Initiative. In 2011 the OECD introduced its Better Life Index (BLI) as part of previous efforts at the national and international levels to measure progress and sustainability. The BLI, fully described in the *How's Life?* reports (OECD, 2011; 2013; 2015; 2017), is a key element of the Better Life Initiative. It is devised as a composite multidimensional index, lying on a wide range of elements that contribute to a good life. The eleven well-being dimensions underlying BLI account for material living conditions and quality of life across the population at the aggregate country level. They are broadly consistent with those presented in the Stiglitz-Sen-Fitoussi Commission report (Stiglitz et al., 2009) and with other similar attempts to monitor well-being and progress.[12] In our work, we define an 'objective' (η_{-i}) and a 'subjective' (η_i) welfare measure starting from two different comparable OECD datasets for the year 2012, one based on average country-level data reflecting well-being outcomes, the other one on microdata reflecting people's stated preferences on well-being indicators. We then refer to those two different multidimensional welfare indicators as η_{-i} (objective BLI) and η_i (subjective BLI).

[12] See for example reports from Australia (Measures of Australia's Progress), Germany (Sustainable Development Report), Finland (Findicator- Set of Indicators for Social Progress) Italy (BES Report) and New Zealand (Measuring New Zealand's Progress Using a Sustainable Development Approach) (see also note 2).

The OECD approach in measuring welfare, like many others, shares the view that well-being is multidimensional. Multidimensionality however raises an issue in terms of understanding the interrelations across welfare components, as well as assessing the common underlying drivers. In this framework BLI is thought as a dashboard, therefore the well-being dimensions included in the framework are not aggregated together. However, should this framework be used for policy making, it is important to aggregate the dimensions as well as to identify the common drivers of welfare and to judge what are the most effective levers of well-being.

A related problem in this context is that we do not necessarily know enough about causality and the range of determinants of some welfare components. Many of the well-being components are correlated, and in fact mutually dependent (e.g. income may determine health and health may determine income), but we do not necessarily know the exact structural two-way relationship between these variables.

We also know that some of the well-being components are determined by common factors, for instance higher GDP results in higher investment in education and health, which leads –depending on the degree of efficiency in delivery- to higher education and health outcomes. However, also in this case, we know little about the causal relationships between well-being and its determinants.

Finally, we suspect that there is a strong endogeneity between well-being components and some of its determinants: i.e. higher economic growth results in higher well-being, but higher well-being, as driven by health for instance, results in higher economic growth, too.

Given this imperfect knowledge, the best approach is to model the determinants of welfare by making very soft assumptions on the relationships between the various well-being variables and their common drivers, while at the same time taking into account the possible endogeneity issues of these various relationships. We thus need a way to estimate what mostly contributes to higher well-being, taking into consideration that: (i) there are several dimensions of well-being and we do not necessarily know or want to specify what is the relationship between these components and an overall well-being variable; (ii) there are many interrelations across well-being components; (iii) there are interrelations across underlying drivers of well-being. The Structural Equation Modeling (SEM), in the full-information version, is a good method to analyze interrelations among indicators underlying multidimensional topics, as well-being is. This method, based on the analysis of variance-covariance matrices, allow us to study the interrelations and causal relationships across welfare determinants and across the underlying drivers of well-being (Nachtigall et al., 2003; Pearl, 2012; Bollen and Pearl, 2013). SEM, a factor-analytic approach, provides a flexible framework to analyze and develop complex relationships among multiple variables and latent constructs (Bollen, 1989; Ullmann, 2006; Bentler and Ullmann, 2016).[13] When the phenomena of interest are complex and multidimensional, SEM is the only analytical toolkit that allows complete and simultaneous tests of all the relationships in a non-parametric way. It also allows to identify what are the components that mostly drive well-being as well as what drives these components, without imposing strict assumptions upon the nature and strength of any possible interrelation across the model's variables.

Next, we describe the two OECD datasets, illustrate the model specification of SEM and derive the two synthetic measures of well-being (objective and subjective).

[13]SEM examines both direct and indirect, unidirectional and bidirectional relationships between measured and latent variables. Notably, SEM allows to analyse a set of relationships between one or more independent variables (IVs), and one or more dependent variables (DVs), either continuous or discrete. Both IVs and DVs can be either factors or measured variables.

3 Measuring well-being and progress: Defining an objective welfare measure

3.1 Data issues, model specification and estimation

In this Section we describe the estimation procedure to obtain the multidimensional objective welfare measure η_{-i} using SEM. The paper's estimation strategy consists in finding the best fit from an unobserved common factor to the various outcomes. The first step in the SEM approach is the specification of a conceptual model defining how the observed variables are causally related to one another and to the latent variable(s). In our model, drawing from the OECD Better Life Initiative conceptual framework, we included all the eleven well-being dimensions underlying the objective welfare measure, also referred to as the objective BLI hereinafter.
Structural Equation Modeling builds the BLI as a factor. This latent variable is obtained on the basis of the eleven observed, underlying dimensions of well-being. We also consider the correlation between the obtained BLI latent index and GDP, capturing inclusive growth effects. In Figure 1, the proposed causal model and all the relationships among variables are represented by a path diagram. The path diagrams are fundamental in the SEM approach because they allow us to illustrate the hypothesized set of relationships and interrelations in the model.[14]

In the SEM model illustrated in the Figure below, the measurement equation specifies how in each country the latent variable η_{-i} determines the set of observed indicators ($\mathbf{y}_{-i}$) subject to disturbances or errors ($\mathbf{e}_{-i}$). The model can be expressed in matrix form as:

$$\mathbf{y}_{-i} = \mathbf{\Lambda}^o \eta_{-i} + \mathbf{e}_{-i} \qquad \text{for} \qquad -i = 1, ..., C \tag{7}$$

where $\mathbf{y}_{-i} = [y_{-i1}, y_{-i2}, ..., y_{-iJ}]'$ are the (aggregate) domain indicators, $\mathbf{\Lambda}^o = [\Lambda_1^o, \Lambda_2^o, ..., \Lambda_J^o]'$ are the weights which depend on the relative importance that governments attach to the various domains, η_{-i} is the latent factor for objective well-being and $\mathbf{e}_{-i} = [e_{-i1}, e_{-i2}, ..., e_{-iJ}]'$ is a vector of disturbances.[15] The variance of each indicator is used to determine its own weight in the estimation of the latent factor. After the specification, the model is estimated with the goal of minimizing the difference between the observed and estimated population covariance matrices.

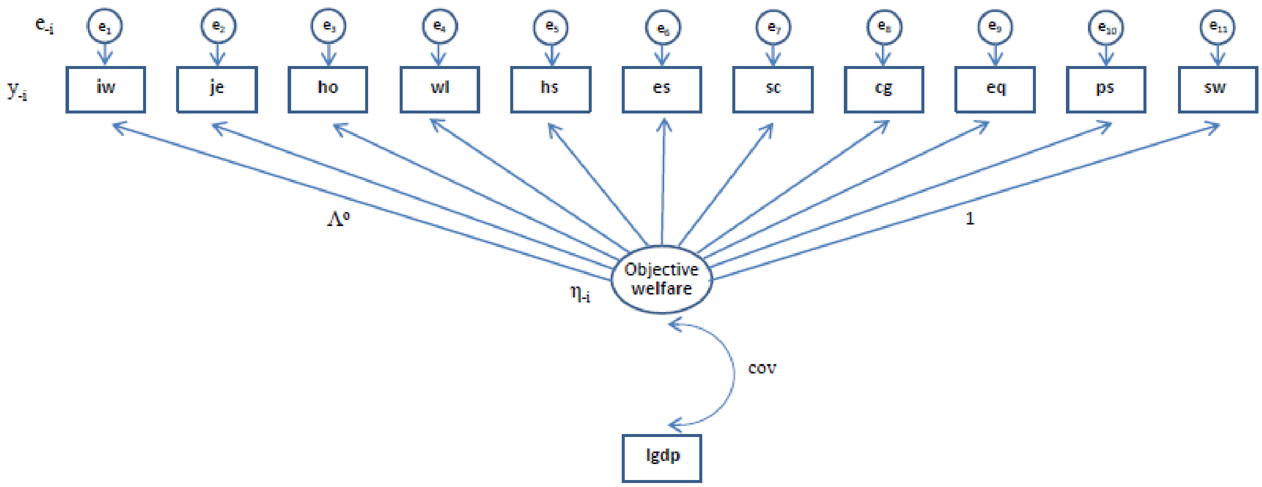
The dataset includes aggregate country-level (average) observations for the eleven selected dimensions of BLI for 35 countries - 33 OECD countries and two emerging economies (Brazil and

[14] By convention, in SEM the direction of the line linking together a latent variable with a measured variable is pointed towards the latter. The rationale behind this convention is that the latent variable - or factor - is a construct derived from the simultaneous contribution of each underlying variable, which in turn are predicted by the factor itself. In that sense, the factor can be viewed as a resulting variable which in turn drives, or 'creates', all the underlying indicators. In the path diagram, the latent variable (BLI) is represented with an ellipse, the measured variables with squares and the errors with circles. Each arrow represents a causal connection between variables, or a causal path. A line ending with an arrow indicates a hypothesized direct relationship -unidirectional causation- between the variables. A line with a two-headed arrow indicates a covariance between the two variables with no implied direction of effect -no specification of the direction of causality- which may be interpreted also as reverse causality. The direction of the arrow does not necessarily indicate the direction of causation (Bentler and Ullman, 2013).

[15] The dependent variables $\mathbf{y}_{-i}$ have residuals indicated by errors ($\mathbf{e}_{-i}$) pointing to the measured indicators in the graph. It is assumed that $E(\mathbf{e}_{-i}) = \mathbf{0}$ and $\text{cov}(\mathbf{e}_{-i}, \eta_{-i}) = \mathbf{0}$. The parameters of the model to be estimated are the regression coefficients for the paths between variables and variances/covariances of independent variables (IVs). Based on the sample data, the parameters are estimated and then used to obtain the estimated population variance-covariance matrix.

Russian Federation) for the year 2012. We then refer to it as 'Objective' OECD BLI dataset.[16]

Figure 1: Structural Equation Model for the Objective Welfare Measure - Path diagram



Note: The variables' notation in Figure 1 is the following: Subjective well-being (sw), Income and wealth (iw), Jobs and earnings (je), Housing condition (ho), Health status (hs), Social connections (sc), Education and skills (es), Environmental quality (eq), Personal security (ps), Work-life balance (wl), Civic engagement (cg), GDP logarithm (lgdp).

We started our analysis from the original OECD BLI dataset, including 24 variables underlying the eleven dimensions of BLI (see Table 6 - Appendix I). In order to utilize the Maximum Likekihood Missing Values ($MLMV$) method within SEM[17], we excluded from this dataset all the imputations made by the OECD, thus retaining the missing data of the OECD BLI dataset. After that, we obtained each of the eleven BLI dimensions by aggregating 1 to 4 variables of interest from 24 underlying indicators.[18] Concerning GDP, we refer to the year 2010 data drawn from the IMF World Economic Outlook database, 'October 2014 edition'. For our calculations, we use the logarithm of GDP (lgdp).

In spite of the small sample of 35 observations, the SEM analysis we produced allowed us to obtain reliable and robust results, as confirmed by goodness-of-fit indicators and tested through a specific power analysis we have conducted, based on Westland's (2010) algorithm (see Appendix II and Appendix III).

In our analysis we opted for the SEM $MLMV$ estimation along with non-parametric bootstrapping (1,000 replications). For model identification, we first need to identify the number of data points

[16]The SEM analysis was conducted based on the official OECD Better Life index (BLI) dataset using the statistical software STATA v. 13.1. Originally, the full OECD dataset included 36 countries but, in order to ensure a better fit of the model, after inspection of scatterplots for dimensions and country, we opted to drop from the original dataset the outlier represented by Luxembourg. With regard to GDP, we utilized the year 2010 data since they are consistent with the features of the 2012 release of the OECD BLI dataset.

[17]From the simulations we carried out, it emerged that the model fit increases considerably when we use the SEM $MLMV$ method along with the raw dataset (with missing values), instead of the default SEM running on the original BLI dataset (with OECD imputations). The $MLMV$ method, implemented by STATA, aims to retrieve as much information as possible from observations containing missing values (see Appendix II for the description of the $MLMV$ method and for details on bootstrapping).

[18]Following the OECD recommendations, within each dimension, indicators are averaged with equal weights in a normative way, and normalized, when expressed in different units of measure.

and the number of parameters to be estimated. The number of data points is the number of non-redundant sample variances and covariances. The number of parameters is found by adding together the number of regression coefficients, variances and covariances to be estimated. To scale homogeneously all the factors, we fix to 1 the regression coefficient of the Subjective well-being (sw) variable. This constraint implies that the BLI factor has the same variance of the selected measured variable.[19] With reference to our model, we have 78 data points versus 24 parameters to estimate.[20] Given that there are more data points than parameters to be estimated, the model is said to be overidentified, a necessary condition for proceeding with the analysis and the estimation of the parameters of interests. The next step in the identification of the model is to examine its measurement portion, which deals with the relationships between the factor and the measured underlying indicators. If the model is composed of only one factor, the model may be identified if the factor has at least 3 indicators with non-zero loadings and the errors are uncorrelated with one another. In our model, we have one factor and eleven measured indicators loading on it, therefore it can be identified.

Statistically, the fundamental question addressed through SEM includes a comparison between an empirical variance-covariance matrix and an estimated population variance-covariance matrix that is a function of the model parameter estimates. SEM uses an iterative approach to minimize the differences between the sample and the estimated population variance-covariance matrices. Maximum Likelihood (ML) is currently the most frequently used estimation approach in SEM (Ullman, 2007) to derive the structural parameters $\mathbf{\Lambda}^o$. If the model is reliable, the parameter estimates will produce an estimated matrix that is close to the sample variance-covariance matrix. 'Closeness' is evaluated with the chi-squared test statistic (χ^2) and goodness-of-fit indices. Moreover, in order to test the robustness, SEM allows us to compare alternative models assessing the relative model fit (see Appendix III for more details on model estimation and evaluation).

3.2 Objective welfare measure: Results

The main parameters and standard errors from our SEM estimation - standardized and unstandardized - are shown in Table 1. The objective welfare measure (or objective BLI) emerges as a latent variable from the eleven dimensions of well-being. Associated to each of these dimensions, there is a coefficient describing the 'loading' of the considered measured variable on the BLI latent factor. The corresponding p value is marked with asterisks, whilst the relative standard error is reported in round parentheses.

Each structural equation coefficient is computed taking into account the sample variances and covariances. Thus, coefficients are calculated simultaneously for all the endogenous variables rather than sequentially, as in canonical multiple regression models. SEM accounts for the degree to which the various indicators covariate with one another.

[19] Subjective well-being (sw) is probably the best predictor of BLI among the considered components of people well-being, thus its scale should be very close to the BLI one. The choice of taking the sw coefficient as the *numéraire*, allows easier interpretation of the remaining BLI indicators' estimated loadings.

[20] Notably, the number of data points is obtained from $\frac{p(p+1)}{2}$, where p equals the number of measured variables. In our model we have 12 measured variables so that the number of data points is 78, corresponding to 12 variances and 66 covariances among variables. The number of parameters to be estimated in our model equals 24 corresponding to the sum of 11 path coefficients (12 measured variables – 1 constrained term), 11 error's variances, 1 variance for latent BLI and 1 covariance.

The coefficients are based on the direct relationships between the variables. They show the quantitative relationships between variables (unstandardized coefficients) as well as the relative importance of the variables within the model (standardized parameters). Notably, the standardized coefficients represent the change in the dependent variable which results with a one unit change in the independent variable.

Table 1: Bootstrapped SEM MLMV model estimated paramenters

Observed variables	**Standardized**		**Unstandardized**	
Income and wealth (iw)	0.727***	(0.126)	20741.01**	(8642.46)
Jobs and earnings (je)	0.927***	(0.036)	0.145***	(0.045)
Housing (ho)	0.841***	(0.068)	0.134***	(0.037)
Health status (hs)	0.844***	(0.060)	0.202***	(0.055)
Social connections (sc)	0.645***	(0.094)	0.063**	(0.023)
Education and skills (es)	0.581***	(0.162)	0.182	(0.100)
Environmental quality (eq)	0.594***	(0.119)	0.136**	(0.055)
Personal security (ps)	- 0.599***	(0.190)	- 0.123	(0.091)
Work-life balance (wl)	0.506*	(0.263)	0.135	(0.091)
Civic engagment (cg)	0.438***	(0.137)	0.108**	(0.045)
Subjective well-being (sw)	0.696***	(0.121)	1 (constrained)	
Correlations/Covariances				
corr[lgdp, BLI]	0.972***	(0.059)		
cov[lgdp, BLI]	-0.420***	(0.099)		
Observations	35			
logLikelihood	-48.580			
Replications	971			

BLI path coefficients without parentheses
Bootstrapped Standard errors in round parentheses
*p<0.05; **p<0.01; ***p<0.001

Data source: OECD Better Life Index data (year 2012)

The unstandardized parameters reflect the form of the relationship, while a standardized coefficient measures the strength of an association. Both are useful to interpret the results (Acock, 2013). In order to analyze the relative importance of each of the eleven dimensions underlying objective welfare measure, we refer to the standardized estimates of the loadings. Unlike the unstandardized estimates, they allow a comparison among dimensions measured in different scales.

As shown in Table 2, from the analysis of the standardized parameters it emerges that, as expected, the most important dimensions driving the objective welfare measure are Job and earnings (je), Health status (hs) and Housing (ho) followed by Income and wealth (iw). Those are the four topics representing the material conditions underlying people well-being.

On the other hand, the least important dimensions explaining the objective welfare measure are Civic engagement (cg), Work and life balance (wl), Education (es) and Environmental quality (eq). Social connection (sc) and Personal security (ps) lie in the middle of the ladder.

It needs to be stressed that Personal security is negatively linked to objective BLI,[21] as reported

[21] An important result confirming the robustness of our model estimation is that, as expected, the relationship

in Table 1, whilst Work-life balance (wl) is statistically significant but less than all the other dimensions in the standardized estimates. Considering the unstandardized parameters, it emerges that Work and life balance (wl), Personal security (ps) and, in a minor way, Education and skills (es) are not statistically significant.

Table 2: OECD Dimensions' ranking of the objective welfare measure

SEM (standardized)
Jobs and earnings (je)
Health status (hs)
Housing (ho)
Income and wealth (iw)
Subjective well-being (sw)
Social connection (sc)
Personal security (ps)
Environmental quality (eq)
Education and skills (es)
Work-life balance (wl)
Civic engagement (cg)

It should be highlighted that, in the unstandardized model, the path from objective BLI to Subjective well-being (sw) is fixed to 1 for identification, whilst Subjective well-being (sw) lies in the middle of the ladder in the standardized rank.

The covariance/correlation between factor BLI and logGDP is reported at the bottom of Table 1. It is used to account for the two-way (reverse) causality between the two variables and it can be interpreted as a measure of the 'inclusiveness' of the process generating GDP, in line with the concept of inclusive growth. With a correlation value of 0.97, GDP can be considered as a major driver of people's well-being.[22] Moreover, indirect effects among GDP and each of the eleven underlying well-being dimensions can be computed considering the BLI construct also as a 'mediator' variable.[23] As Appendix III shows, considering the combined analysis of the overall goodness-of-fit indices reported in Table 9, we can conclude that our hypothesized model presents a good fit, taking into account the small sample size on which all the estimates are based on.

between Personal security (ps) and objective BLI is negative. The main reason explaining this outcome is that, following the OECD Better Life Index framework, we obtain the Personal security (ps) indicator aggregating two underlying variables - Reported homicides and Self-reported victimisation - which notoriously affect people's well-being negatively (see Table 6 - Appendix I).

[22] This result is in line with the estimation made by Jones and Klenow (2016) using a different method. Comparable results are obtained by those authors also with reference to the country ranking based on welfare levels. Furthermore, in corroboration of the robustness of our estimation, it should be highlighted that the results and parameter estimates remain substantially the same if we do not consider the cov/corr (logGDP, BLI) in our model, other things being equal.

[23] A feature of SEM is the ability to test not only direct effects between variables but also indirect effects which involve one or more mediator variables. Indirect effects are obtained as the product of the estimated coefficients - either standardized or unstandardized - of the two paths connecting the first variable to the mediator variable and the mediator variable to the last variable considered.

4 Measuring well-being and progress: Defining a subjective welfare measure

This Section illustrates the estimation procedure of the subjective welfare measure (or subjective BLI) using a special setting of SEM. Within the OECD Better Life Initiative, the OECD recently launched a complementary project, Your Better Life Index, with the aim of assessing welfare and progress of societies from an individual perspective. A specific tool available on the official OECD website enables every user to assess their well-being according to their own preferences.[24] All these 'subjective' microdata - individual stated preferences - were gathered in order to complement the information provided by the standard objective BLI, based mainly on country-level average data, reflecting 'objective' outcomes from official statistics. This new large dataset of individual stated preferences on the eleven dimensions underlying the subjective BLI, represents an unprecedented international attempt to provide comparative evidence on well-being and progress. It constitutes a valuable aspect of our analysis.[25]

As mentioned above, the BLI conceptual framework –both in the 'objective' and 'subjective' versions[26] - refers to a multidimensional indicator relying on eleven underlying dimensions, without any explicit choice by the OECD on their relative importance for people's well-being. As a consequence, the BLI does not explicitly provide for an official single, concise welfare statistic but just for a dashboard of unweighted indicators for each country. A single welfare measure for BLI measuring the level of progress and well-being of countries and regions in a concise way, could be a very useful policy making tool. To this end, the OECD suggests - as a default setting - to consider identical weights for the eleven underlying dimensions, in order to produce an informal concise measure for BLI, without introducing any hypothesis on on the relative importance of the selected well-being drivers. Using the OECD subjective microdata for 35 countries and considering the Likert-scale (non-normal) structure of the individual responses, we propose a Generalized Structural Equation Model (GSEM) to estimate endogenously the relative weights of the eleven dimensions of BLI. More specifically, we adopt a Multiple Indicators Multiple Causes (MIMIC) model under GSEM to account for the geo-demographic control variables included in the OECD individual microdataset. This econometric method allowed us to obtain more precise estimates of countries' BLI scores than those provided by the OECD using the default setting and equal weighting. In addition, the model provided us with a subjective ranking of the eleven dimensions underlying BLI derived from the individual stated preferences.

In order to overcome an important limitation in the GSEM post-estimation indices, we propose to estimate, in parallel with it, a bootstrapped SEM model, running on the same dataset, to get all the available overall-goodness-of-fit indices for the model.[27]

[24] See www.oecdbetterlifeindex.org for details.

[25] The authors thank Romina Boarini and Marco Mira D'Ercole – OECD General Secretariat and OECD Statistics Directorate – for giving us the possibility to use the OECD Your Better Life Index microdata in our work.

[26] To simplify, with 'objective' BLI - or objective welfare measure - we indicate the multidimensional index obtained from the OECD BLI dataset, based on aggregate country's level data from official sources. On the other hand, with 'subjective' BLI - or subjective welfare measure - we refer to the index obtained from individual level OECD BLI microdata.

[27] When using GSEM instead of SEM, we demonstrate an improvement of about 25-30% in the overall fit of the model, through the comparison of the relative Akaike Information Criterion (AIC), a predictive fit index available for both models. Therefore, the use of GSEM for the estimation of the subjective welfare measure in Section four is justified by those values (see Subsection 4.2 and Appendix IV for more details).

4.1 Individual stated preferences and well-being drivers: A GSEM MIMIC approach using new OECD microdata

Through the OECD BLI official website, thousands of users of Your Better Life Index around the world shared their views on what makes for a better life. Users have been encouraged to create and share their own Better Life Index since its launch in 2011. Up to date, the OECD received about one hundred thousand individual indices from 180 countries and territories, which are included in a unique and comprehensive OECD dataset on BLI users stated preferences. Those individual microdata are at the core of this paper. Table 3 reports the summary statistics of the microdata used in the analysis.

Table 3: Descriptive statistics

Variable	Obs.	Mean	Std. Dev.	Min	Max
Income and wealth (iw)	12728	3.03	1.38	0	5
Jobs and earnings (je)	12728	3.23	1.40	0	5
Housing (ho)	12728	3.21	1.37	0	5
Health status (hs)	12728	3.80	1.39	0	5
Social connections (sc)	12728	3.05	1.45	0	5
Education (es)	12728	3.65	1.43	0	5
Environmental quality (eq)	12728	3.37	1.46	0	5
Personal security (ps)	12728	3.25	1.47	0	5
Work-life balance (wl)	12728	3.43	1.48	0	5
Civic engagment (cg)	12728	2.45	1.40	0	5
Subjective well-being (sw)	12728	3.79	1.43	0	5
gender	12721	0.41	0.49	0	1
age	12704	2.44	1.35	1	7
country	12728	16.13	11.01	1	35
world region	12728	1.16	0.63	1	4

Data source: OECD Your Better Life Index microdata (year 2012)

In order to make this work comparable with the Objective BLI results in Section three, we selected from the OECD BLI dataset 12,728 individual observations from 33 OECD countries and 2 emerging economies -Brazil, Russian Federation- for the year 2012.[28] As mentioned above, weights on the eleven dimensions of BLI are assigned by the users, who build and customize their own Index. Users have to rate each topic assigning a rate ranging from 0 ("not important") to 5 ("very important"). Given the Likert scale structure of individual answers, all the responses have only six possible choices, corresponding to six integers from 0 to 5. Therefore, the microdata gathered

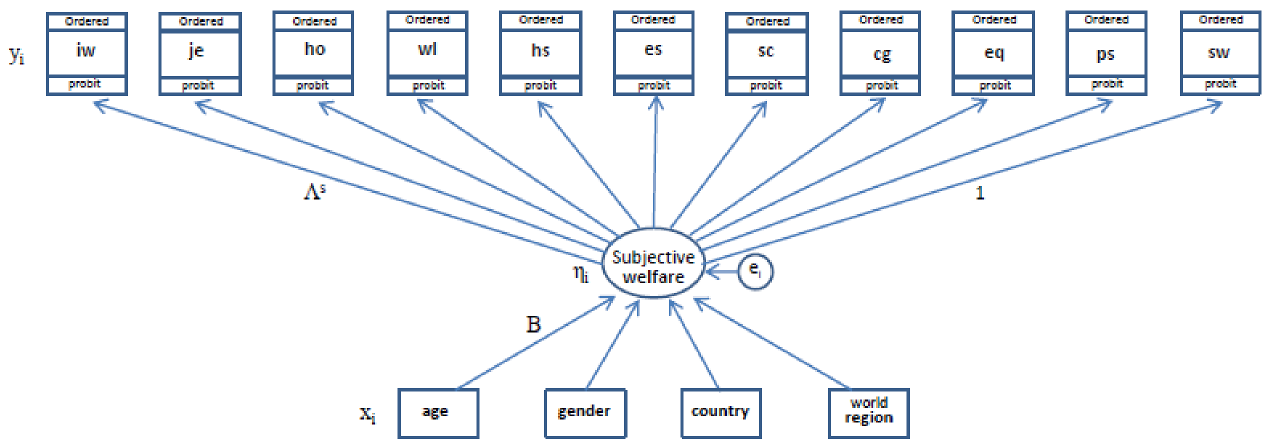
[28] In order to improve the fit of our model, we dropped Luxembourg from the original OECD sample, because its observations emerged as outliers. This choice is also consistent with Section three. It should be noticed that, in our work, the number of observations used by GSEM running on the full OECD sample is lower than 12,728 and equal to 12,703, as reported in Table 12. SEM/GSEM method in STATA 13.1 package makes use of listwise deletion as default setting in presence of missing data. Therefore, missing data are dropped from the dataset leaving only complete rows for each individual.

are categorical (ordinal) and can be defined as individual stated preferences. As expected, the multivariate normality tests confirm that data are multivariate non-normal (see Appendix IV).

4.1.1 Model description

The subjective BLI can be defined as a composite multidimensional construct, based on a large set of underlying variables reflecting material living conditions and quality of life. In line with the OECD BLI framework, we cannot define BLI weights directly, but let them emerge indirectly considering BLI as a latent common factor. Structural equation modeling (SEM) allows to account for causal relationships among indicators. With ordinal categorical responses or polytomous (Likert-type), we need a Generalized model using an ordered probit or logit or complementary log-log link functions to deal with non-normal microdata (Agresti, 2002). Taking into account that ordered probit is considered the best option for latent variable models (Skrondal and Rabe-Hesketh, 2005), we decided to apply it in our GSEM estimation.

Figure 2: Ordered Probit GSEM MIMIC Model for the Subjective Welfare Measure - Path diagram



As mentioned before, besides individual responses on the eleven BLI indicators, the OECD dataset under consideration also includes four control variables which may influence our latent construct. More specifically, these geo-demographic variables are age, gender, country and geographical area - or world region/macroregion - of the respondents. We consider them as 'causes' influencing our latent construct, as shown in the path diagram in Figure 2.

The MIMIC (Multiple Indicators Multiple Causes) model allows us to assess the influence that a set of 'causes' can have directly on the latent BLI or indirectly on the eleven underlying indicators, when BLI operates as a 'mediational' variable. With reference to the 'causes', in the specified GSEM MIMIC model the observed 'causal' variables drive the latent variable which in turn determines the observed indicators. Therefore, methodologically, we propose an ordered probit GSEM MIMIC model to analyse the 'causes' and determinants of well-being and progress measured through the subjective welfare measure. As illustrated in Figure 2, the section of the graph below BLI represents the 'causal' model of the GSEM MIMIC, whereas the section above the latent construct, is the 'measurement' model. Finally, $\mathbf{e}_i$ represent the disturbances.

4.1.2 Model specification and estimation

In the GSEM MIMIC model (Multiple Indicators, Multiple Causes, see Joreskog and Goldberger, 1975; Rabe-Hesketh et al., 2004; Raiser et al., 2007) it is not only assumed that the observed variables are manifestations of a latent concept but also that there are other exogenous variables that 'cause' and influence the latent factor(s). We model subjective welfare for each cross sectional unit (individual) by assuming that the domain indicators, $\mathbf{y}_i$, are related to the latent factor for subjective well-being, η_i, via the measurement equation:

$$\mathbf{y}_i = \mathbf{\Lambda}^s \eta_i + \mathbf{e}_i \qquad \text{for} \quad i = 1, ..., I \tag{8}$$

where $\mathbf{y}_i = [y_{i1}, y_{i2}, ..., y_{iJ}]'$ are the domain indicators, $\mathbf{\Lambda}^s = [\Lambda_1^s, \Lambda_2^s, ..., \Lambda_J^s]'$ the weights (i.e. factor loading matrix) and $\mathbf{S_e}$ is the covariance matrix of $\mathbf{e}_i$=$[e_{i1}, e_{i2}, ..., e_{iJ}]$ which is a vector of disturbances. It is assumed that $E(\mathbf{e}_i) = \mathbf{0}$ and $\text{cov}(\mathbf{e}_i,\eta_i) = \mathbf{0}$. In the MIMIC model, however, besides the measurement equation defined above, there is also a 'causal' equation that expresses the relationships between the latent construct (η_i) and observed variables ($\mathbf{x}_i$) or 'causes':

$$\eta_i = \mathbf{B}\mathbf{x}_i + v_i \tag{9}$$

where $\mathbf{x}_i = [x_{i1}, x_{i2}, ..., x_{ir}]$ are the observed individual characteristics, comprising income and other socio-demographic hallmarks, that are "causes" of η_i subject to disturbances (v_i). $\mathbf{B} = [B_1, ..., B_r]'$ is the corresponding vector of structural parameters relating to the latent dependent variable η_i, whilst $\Theta_\mathbf{v}$ is the variance-covariance matrix of $\mathbf{v}$.

By replacing the measurement equation (8) in the 'causal' equation (9) we obtain:

$$\mathbf{y}_i = \mathbf{\Lambda}^s (\mathbf{B}\mathbf{x}_i + v_i) + \mathbf{e}_i \tag{10}$$

The socio-demographic individual characteristics determine the weight $\mathbf{\Lambda}^S = \mathbf{\Lambda}^s\mathbf{B}$ attached to each each domain indicator underlying the the subjective welfare factor, η_i.

In the OECD microdataset, the observed discrete variables for the welfare domains are generalized responses where the response for y_{ij} is assumed to take one of k_k unique values[29] with $k_0 = -\infty$, $k_y < k_{y+1}$, $k_k = +\infty$. The probability that y_{ij} takes the observed value k_y is:

$$\Pr(y_{ij} = k_y) = \Pr(y_{ij}^* < k_y - z) - \Pr(y_i < k_{y+1} - z) \tag{11}$$

where y_{ij}^* is the latent component for y_{ij} whilst the expected value of y_{ij} is indicated by z.[30]

Since our data are either binomial or categorical (Lykert-type scale), we use a generalised model (GSEM) in order to deal with non-normality and the idiosyncratic structure of the data. Unlike the case of continuous responses, maximum likelihood estimation (ML) cannot be based on the empirical covariance matrix of the observed responses. Indeed, the likelihood is obtained by integrating out the latent variable(s).[31] Let $\boldsymbol{\theta}$ be the vector of independent parameters, $\mathbf{y}$ be the vector

[29] In our model, $k = 6$. As described in paragraph 4.1, the individual discrete response y_{ij} associated to the eleven indicators underlying BLI, are expressed in a Likert-type scale through six integers, ranging from 0 to 5.

[30] The distribution for y_{ij} is determined by the link function. Typical choice of link function for categorical responses is the probit link. Within GSEM, the probit link assigns to y_{ij} the standard normal distribution. Except for the ordinal family, the link function defines the transformation between the mean and the linear prediction for a given response. GSEM fits generalized linear models with latent variables via Maximum Likelihood (ML).

[31] Within STATA 13.1, log-likelihood calculations for fitting any model with latent variables require integrating out the latent variables. The default numerical integration method implemented in GSEM is the Mean-variance adaptive Gauss-Hermite quadrature (MVAGH). This method is based on Rabe-Hesketh et al. (2005).

of observed response variables, $\mathbf{x}$ be the vector of observed exogenous variables or 'causes', and $\boldsymbol{\eta}$ be the latent construct. Then the marginal likelihood can be computed as:

$$\mathcal{L}(\boldsymbol{\theta}) = \int_{\Re^q} f\ \left(\mathbf{y}|\mathbf{x}, \boldsymbol{\eta}, \boldsymbol{\theta}\right)\ \phi\left(\boldsymbol{\eta}|\mu_{\boldsymbol{\eta}},\ \boldsymbol{\Omega}\right) \partial\boldsymbol{\eta} \tag{12}$$

where $\Re$ denotes the set of values on the real line, $\Re^q$ is the analog in a $q-$dimensional space, $f\ (.)$ is the conditional probability density for the observed responses $\mathbf{y}$, $\phi\,(.)$ is the multivariate normal density for $\boldsymbol{\eta}$, $\mu_{\boldsymbol{\eta}}$ is the expected value of $\boldsymbol{\eta}$ and $\boldsymbol{\Omega}$ is the covariance matrix of $\boldsymbol{\eta}$. If we have J indicators, the conditional joint density function for a given observation is:

$$f\ \left(\mathbf{y}|\mathbf{x}, \boldsymbol{\eta}, \boldsymbol{\theta}\right) = \Pi_{j=1}^{J}\ f_j\ \left(y_j|\mathbf{x}, \boldsymbol{\eta}, \boldsymbol{\theta}\right) \tag{13}$$

The advantage of Structural Equation Modeling -also in its generalized form- compared with standard econometric methods, is that SEM uses the full information on causes and indicators of the latent dependent variable. Therefore, the latent construct relates directly to the causes and to the indicators used to specify the model which simultaneously estimates the underlying system of equations.

4.2 Subjective welfare measure: Results

Given the availability of a rich microdataset, we perform the ordered probit GSEM MIMIC model for various groups of countries and macroregions along with the OECD area as a whole.[32] The GSEM estimated parameters are unstandardized. Actually, the unstandardized loadings are fully comparable among them in relative terms (Hoyle, 1995) and they can be used to rank the BLI indicators and 'causes'. In this Section, we focus on the five major European Union (EU) countries and on the United States (US) because of the larger number of observations available for these sub-samples.[33] We then compare the five European Union (EU) countries and the EU as a whole[34] to the United States (US), to show differences between people preferences in those two developed areas.

We start our analysis considering the OECD dimensions' ranking as the benchmark against which countries and continents are to be compared with. As shown in Table 4,[35] we consider the subjective BLI dimensions' ranking from the OECD, the EU and the sub-samples for the six selected countries -United States (US), France (FR), Germany (DE), Italy (IT), Spain (ES), United Kingdom (UK). We observe that, overall, there is some stability at the top and at the bottom of our rankings. More specifically, Health status (hs), Education and skills (es), Enivironmental quality (eq) and Personal security (ps) are generally at the top, whilst Income and wealth (iw), Jobs and earnings (je) and Housing condition (ho) are at the bottom. The relative positions for the other dimensions vary from country to country. Furthermore, we can observe that Social connection (sc), Work-life balance (wl) and Civic engagement (cg) are often in the middle of the ladder for all the considered countries and macroregions.

[32] See Tables 12, 13 and 14 in Appendix V reporting the GSEM MIMIC (and bootstrapped SEM) estimates of coefficients and fit indices for the subjective welfare measure.

[33] The five EU countries sub-samples selected for our analysis are France, Germany, Italy, Spain and UK, respectively. Each sub-sample comprises at least 250 observations, as reported in the Appendix V tables.

[34] For Europe we aggregate individual observations from 21 EU countries within the OECD. Countries included are Austria, Belgium, Czech Republic, Germany, Denmark, Estonia, Finland, France, Greece, Hungary, Ireland, Iceland, Italy, The Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, UK.

[35] The full set of results by country, macroregion and gender are available in Tables 12, 13 and 14 of Appendix V.

It should be stressed that, as expected, Income and wealth (iw) and Jobs and earnings (je) - i.e. the materialistic dimensions underlying BLI - tend to stay very low in the individual ranking based on people's stated preferences, probably because the BLI is perceived as a measure of well-being other than GDP and other materialistic components of life. This explanation could be extended to Housing condition (ho) as well. As a consequence, Income and wealth (iw), Jobs and earnings (je) and Housing condition (ho) tend to be systematically penalized in this kind of surveys. Therefore, an important message emerging from our analysis is that income buys only 'some' happiness. More specifically, with reference to the top of the ranking, it comes out that Health status (hs) is always the most important dimension in explaining subjective BLI, except for Italy (IT), where Environmental quality (eq) is the most important component. We can state that Education and skills (es) and Environmental quality (eq) are the second and third most important components of subjective BLI, followed by Personal security (ps). At the bottom of the ladder, Income and wealth (iw) is always the last dimension - except for Spain where Jobs and earnings (je) is the last component - followed by Jobs and earnings (je) and Housing condition (ho), respectively. In the middle of the ranks lie Civic engagement (cg), Work-life balance (wl) and Social connection (sc) in different orders.

If we compare the Europaen Union as a whole with selected EU countries, taking into account the above-mentioned considerations, we can observe that for Germany, Environmental quality (eq) and Work-life balance (wl) rank low; for Italy, Environmental quality (eq) and Civic engagement (cg) rank high; for France, Work-life balance (wl) and Housing condition (ho) rank high whilst Personal security (ps) and Social connection (sc) rank low; for Spain, Work-life balance (wl) ranks high whilst Environmental quality (eq) and Social connection (sc) rank low in people's preferences. When comparing the United States with the European Union, we can observe that Social connections (sc) rank high in the US, whilst Education and skills (es) and Civic engagement (cg) rank low compared to EU, the remaining dimensions being in similar positions. If we compare the rankings of the EU and the OECD, we notice that the top and the bottom of the ladder are the same, whereas in the middle we have the same dimensions but placed in a different order. Notably, Civic engagement (cg) and Social connection (sc) are in inverted order, with Civic engagement (cg) higher in EU than in the OECD ladder.

In order to carry out an analysis of the relative importance of the BLI dimensions by gender, we split the OECD full sample in two sub-samples for males and females. The most important difference between the two sub-populations is that age has an influence on women's well-being, but not on men's quality of life, whilst the opposite happens with reference to country level analyses. When we compare the two distinct GSEM estimates for males and females, we can observe that the top of the ladder does not vary – Health status (hs), Education and skills (es), Environmental quality (eq) and Personal security (ps) being the most important dimensions.

Also the bottom of the ranking is rather stable with Income and wealth (iw) and Jobs and earnings (je). The remaining dimensions change their relative positions. Notably, Work-life balance (wl) and Housing condition (ho) are more important for women than men whilst the opposite happens to Civic engagement (cg) and Social connections (sc), which are more important for men compared to women.[36]

We finally estimate two comparable models running on the same microdataset, an ordered probit GSEM MIMIC model and a SEM model with bootstrapped robust standard errors, in order to obtain all the available post-estimation indices and the Akaike Information Criteria (AICs) reported in the tables of Appendix V.

[36]It should be highlighted that those rankings correspond, broadly speaking, to the relative importance of each dimension in contributing to the explanation of the subjective BLI variance.

Table 4: Dimensions' rankings of the subjective welfare measure for Countries and Macroregions

GSEM - OECD (full sample)	**GSEM - OECD Male**	**GSEM - OECD Female**	**GSEM - European Union**	**GSEM - USA**
Health Status (hs)	Health status (hs)	Health status (hs)	Health status (hs)	Heath status (hs)
Education and skills (es)	Education and skills (es)	Education and skills (es)	Education and skills (es)	Environmental quality (eq)
Environmental quality (eq)	Environmental quality (eq)	Environmental quality (eq)	Environmental quality (eq)	Social connections (sc)
Personal security (ps)	Personal security (ps)	Personal security (ps)	Personal security (ps)	Personal security (ps)
Social connections (sc)	Social connections (sc)	Work-life balance (wl)	Civic engagement (cg)	Education and skills (es)
Work-life balance (wl)	Civic engagement (cg)	Housing (ho)	Work-life balance (wl)	Work-life balance (wl)
Civic engagement (cg)	Work-life balance (wl)	Social connections (sc)	Social connections (sc)	Civic engagement (cg)
Housing (ho)	Jobs and earnings (je)	Jobs and earnings (je)	Housing (ho)	Housing (ho)
Jobs and earnings (je)	Housing (ho)	Civic engagement (cg)	Jobs and earnings (je)	Jobs and earnings (je)
Income and wealth (iw)	Income and wealth (iw)	Income and wealth (iw)	Income and wealth (iw)	Income and wealth (iw)

GSEM - Germany (DE)	**GSEM - Italy (IT)**	**GSEM - France (FR)**	**GSEM - UK**	**GSEM - Spain (ES)**
Health Status (hs)	Environmental quality (eq)	Health Status (hs)	Health Status (hs)	Health Status (hs)
Education and skills (es)	Health Status (hs)	Education and skills (es)	Environmental quality (eq)	Education and skills (es)
Personal security (ps)	Education and skills (es)	Environmental quality (eq)	Education and skills (es)	Personal security (ps)
Social connections (sc)	Civic engagement (cg)	Work-life balance (wl)	Personal security (ps)	Work-life balance (wl)
Environmental quality (eq)	Personal security (ps)	Housing (ho)	Social connections (sc)	Environmental quality (eq)
Civic engagement (cg)	Social connections (sc)	Civic engagement (cg)	Work-life balance (wl)	Civic engagement (cg)
Housing (ho)	Work-life balance (wl)	Personal security (ps)	Civic engagement (cg)	Housing (ho)
Jobs and earnings (je)	Jobs and earnings (je)	Jobs and earnings (je)	Housing (ho)	Income and wealth (iw)
Work-life balance (wl)	Housing (ho)	Social connections (sc)	Jobs and earnings (je)	Social connections (sc)
Income and wealth (iw)	Income and wealth (iw)	Income and wealth (iw)	Income and wealth (iw)	Jobs and earnings (je)

Notably, with regard to the SEM goodness-of-fit indices for the subjective welfare measure reported in Tables 12, 13 and 14, we observe that the value range for the comparative fit index (CFI) is 0.81-0.90, for the root mean squared error of approximation (RMSEA) is 0.08-0.10, for standardized root mean squared residual (SRMR) is 0.04-0.06, whilst for the overall R^2 (or coefficient of determination, CD) is 0.86-0.91. These indicators show that, overall, the fit of the SEM model to the data is acceptable but not satisfactory, whilst the portion of variance explained by the model and the selected (independent) variables is very high.[37] As explained in Appendix IV, in order to improve the goodness of fit of our estimations, we use a GSEM model – notably, an ordered probit GSEM MIMIC model - accounting for the idiosyncratic structure of the observed categorical data. As a result, using the same dataset, we show that the GSEM AIC is constantly lower of about 20-25%, in absolute values than the SEM AIC. This implies that the overall goodness-of-fit increases significantly in GSEM. Therefore, we can deem that, for all the countries and macroregions considered, the GSEM model overcomes the model goodness-of-fit cut-off criteria specified in Appendix III, providing good and reliable estimates of the model parameters.

5 Comparing objective versus subjective measures of welfare

We have estimated 'objective' and 'subjective' relative weights of well-being determinants using two different settings of SEM on the basis of two OECD BLI datasets, one comprising average country-level observations and the second one individual level microdata for the year 2012.
These two datasets are analyzed using a Structural Equation Modeling (SEM) approach to estimate a welfare measure (BLI) as a latent factor, starting from its underlying indicators. Notably, we applied a bootstrapped SEM MLMV method to estimate the 'objective' weights. An ordered probit GSEM MIMIC model was adopted to estimate the 'subjective' loadings of the eleven underlying dimensions of BLI.

The aim of this section is: (i) to compare the objective and subjective estimated weights of well-being drivers, (ii) to estimate the subjective and objective predicted welfare scores (i.e. predicted BLI scores) for each country and region and compare their relative objective and subjective rankings, and (iii) to draw policy recommendations.

From the comparison of the results presented in the third and the fourth sections, it emerges that there is a wide difference between the welfare dimensions' rankings estimated on the basis of the two OECD datasets. This difference reflects the "welfare gap" between a government's welfare outcome and (country average) individual welfare levels, as per people's stated preferences ($\frac{\eta_{-i}}{\bar{\eta}_i}$) (see equation (1)).

In Table 5, we compare the dimensions' rankings from the SEM standardized estimates and the GSEM unstandardized values[38]. If we look at the SEM and GSEM results, we notice that Health status (hs) is always at the top, whilst Social connections (sc) lies in the middle of the ladder, both in the objective and subjective ranking. All the other dimensions change their relative position.

[37] See Appendix III for an in-depth description of the fit indices.

[38] In GSEM, as the scale of the eleven indicators underlying BLI is the same for all the eleven variables (Likert-type scale), the unstandardized parameters can be interpreted like standardized ones and are fully comparable among them in relative terms (Hoyle, 1995). Moreover, in order to test the robustness of these results, we compared the SEM and GSEM parameter estimates for Spain using its microdata. As expected, we found that the SEM unstandardized parameters are very similar to the standardized ones and that SEM standardized rank correspond exactly to the GSEM (unstandardized) rank.

The comparison between objective and subjective welfare dimensions' rankings shows that, apart from the relevance of Health status (hs) in both analysis, the results are quite diverse. Notably, the material living conditions are the most important dimensions in the objective ranking, whilst the quality of life indicators are at the top of the subjective ladder.

Table 5: SEM 'objective' vs. GSEM 'subjective' BLI dimensions rankings - OECD

SEM OECD objective	**GSEM OECD subjective**
Jobs and earnings (je)	Health status (hs)
Health status (hs)	Education and skills (es)
Housing (ho)	Environmental quality (eq)
Income and wealth (iw)	Personal security (ps)
Subjective well-being (sw)	–
Social connections (sc)	Social connections (sc)
Personal security (ps)	Work-life balance (wl)
Environmental quality (eq)	Civic engagement (cg)
Education and skills (es)	Housing (ho)
Work-life balance (wl)	Jobs and earnings (je)
Civic engagement (cg)	Income and wealth (iw)

The variables expressing material conditions, which are very important on the basis of the aggregate country's outcome, become the least important issues for people's stated preferences. The opposite occurs for Education and skills (es) and Environmental quality (eq), which appear to be the most important dimensions of well-being for people's preferences. Also Personal security (ps), Work-life balance (wl) and Civic engagement (cg) change their relative position, climbing in the individual ladder.

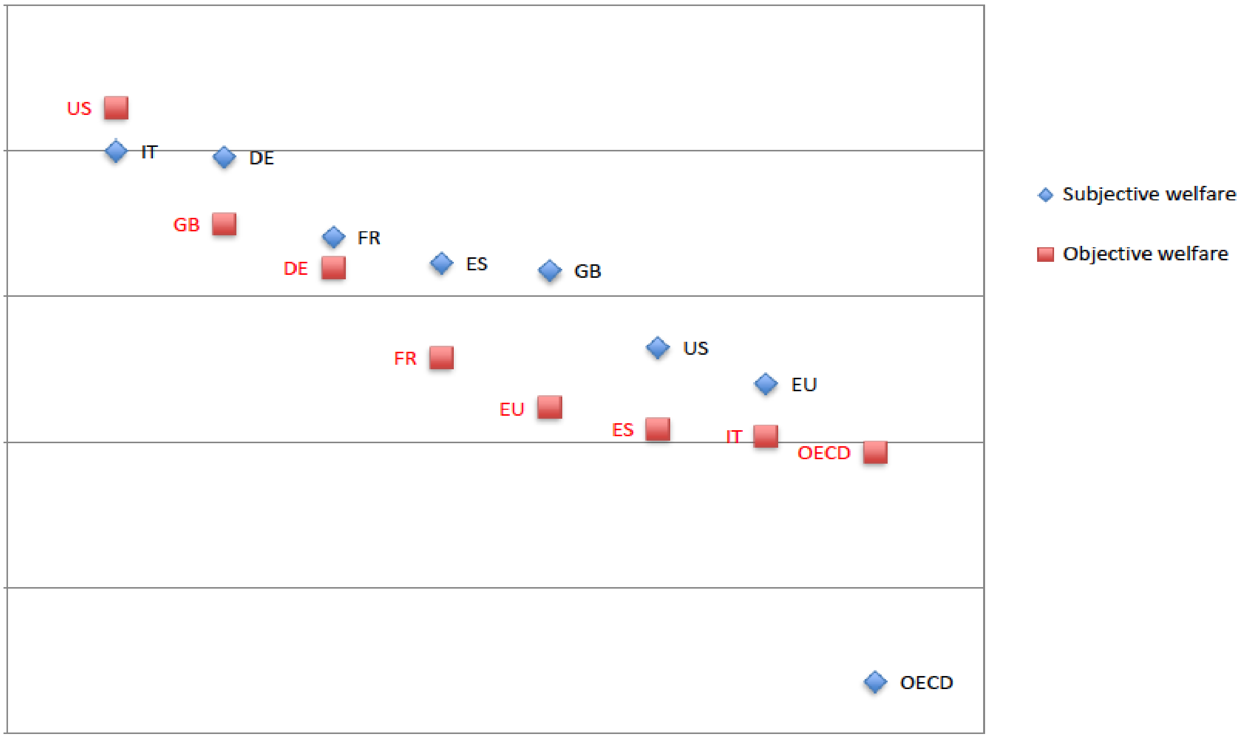
These results are relevant in terms of policy implications because it emerges that material living conditions matter less for people than issues like education and environment. The consequence is that GDP appears as a very important driver for people's well-being, as shown in Section three, but it should be complemented by other elements which decisively contribute to quality of life. In other words, income buys only 'some' happiness. This confirms that it is important to shift the attention and monitoring of governments and policymakers towards different dimensions of people's lives beyond GDP.

After the analysis of both objective and subjective welfare dimensions' rankings, we now focus on the objective and subjective BLI scores calculated at the country and macroregion's level for the year 2012. The predicted BLI score allows to obtain a concise measure of people's well-being for each country and macroregion and to compare them.[39] The results reported in Figure 3 illustrate the comparison between the (country average) subjective predicted BLI scores ($\bar{\eta}_i$) from the GSEM

[39]For the subjective BLI we can obtain a single, headline measure of any country's welfare - the country's factor score - calculating the mean of all the individual BLI factor scores sorted by country. It should be noticed that for the objective welfare measure we cannot directly obtain the country's factor score because of the limited dimension of the OECD BLI 'objective' dataset. Notwithstanding, we obtained the predicted values for each country indirectly computing them as a weighted mean. The latter is obtained, for all the countries of the sample, adding up the

estimation[40] (*Subjective welfare* $(\bar{\eta}_i)$, represented by rhombus) and the objective predicted BLI scores (η_{-i}) from the SEM estimation (*Objective welfare* (η_{-i}), indicated by squares).

Figure 3: Objective and subjective predicted welfare scores by country and macroregion.



Note: *Objective welfare*, measured through the objective predicted BLI score (η_{-i}), represents the government's welfare outcome; *Subjective welfare*, measured through the (country average) subjective predicted BLI scores $(\bar{\eta}_i)$, depicts the aggregated individual welfare aspirations.

For the Objective BLI, the factor scores are estimated with a SEM through a linear regression by using the mean vector and the variance-covariance matrix from the fitted model. As described in Section four, the subjective BLI is estimated with a GSEM MIMIC model. The predicted values - factor scores - are obtained here through an iterative procedure, the empirical Bayes means calculation, also known as posterior means[41] (Skrondal and Rabe-Hesketh, 2004).
In Figure 3 by comparing the United States with the European Union [42] predicted values, it turns out that in the US people are, on average, better off than in the EU, both in objective and subjective terms. If we extend the sample further, by including all the 33 OECD countries - the EU, North

relative value of each dimension multiplied by the specific dimension weight, estimated through the bootstrapped SEM MLMV method.

[40]In line with the theoretical model presented in Section 2.1, the estimated weights are those which maximize individual utility simultaneously, considering the individual stated preferences on the eleven underlying dimensions. Therefore, the Subjective predicted BLI score obtained through GSEM corresponds to the 'desirable' BLI, generating an 'ideal' ranking among the eleven dimensions of well-being for each individual. We can consider the subjective ranking, obtained from the estimated weights, the possible benchmark toward which orienting the objectives of government's socio-economic policies.

[41]Within this method, the iterative procedure makes use of numerical integration whose multivariate integral is approximated by the mean-variance adaptive Gauss-Hermite quadrature (Skrondal and Rabe-Hesketh, 2009).

[42]The EU sample comprises 21 countries, including Eastern Europe (see note 35 for the detailed list of countries).

America, South America and Asia-Pacific regions -, the overall predicted well-being score for the OECD is lower than in the EU and in the US.

We can observe from Figure 3 that for the Unites States (US), the United Kingdom (UK) and European Union (EU), the objective outcomes overcome subjective welfare aspirations in relative terms in 2012. In Germany, France and OECD we can see, instead, that subjective desiderata and government outcomes are in line.

An opposite situation can be observed in Italy and Spain, where people have a very high subjective expectation regarding well-being, but this is associated with very low outcomes achieved by their governments. As described in the theoretical model, such a distance between aspirations and outcomes ($\frac{\eta_{-i}}{\bar{\eta}_i}$) can be defined as a "welfare gap" between what is 'desirable' for people and what government policies achieve in reality. This gap may frustrate citizens' well-being expectations and may contribute to explaining the anti-establishment sentiment that has affected our societies in the latest years, also as a consequence of the economic crisis, as evident in recent elections in Italy and Spain.[43] Coming back to the utility function (1) defined in Section 2, within the European Union (EU) utility tends to be very low for Italy and Spain compared to France and Germany and much lower than in the UK and in the US.

6 Conclusion

The recent economic crisis and the rising inequality affecting our societies over the last decades have stimulated a growing demand for improving the quality of people's lives. However, the pressure on national governments to better living conditions has often been independent of their actual results and policy outcomes. It is desirable for governments to maximise social welfare evaluated according to citizens' own stated preferences.

Social welfare is inherently multidimensional. In this respect, composite indices of well-being, measured at the individual and aggregate level, make it possible to gauge overall welfare and its progress over time. In our analysis, we utilize two different comparable OECD datasets for the year 2012, one based on average country-level macrodata reflecting government's well-being outcomes, the other one on microdata reflecting people's stated preferences on well-being indicators. Drawing from the OECD Better Life Index (BLI) conceptual framework, we then build an 'objective' welfare measure predicted from the national-level data and a 'subjective' index obtained by using OECD microdata. To deal with the idiosyncratic structures of the datasets, we apply two different settings of Structural Equation Models – bootstrapped SEM and Generalised SEM MIMIC - to estimate the relative weights and rankings of the eleven underlying dimensions of well-being.

A key message to be drawn from our objective welfare model is that the material conditions of people's lives, described by Jobs and earnings (je), Health status (hs), Housing (ho) and Income and wealth (iw), are the most relevant dimensions explaining well-being, whilst Civic engagement (cg) is the least important among the eleven considered indicators. The eleven dimensions underlying the objective welfare measure explain 94.1% of the total variance of the latent factor.

On the other hand, the results related to the subjective welfare measure show that the indicators reflecting the quality of life are relatively more important than the variables accounting for the

[43] The predicted BLI score, derived from the individual microdata for the year 2012, provides an indication of people's preferences with respect to the public policy outcomes carried out by their government. From Figure 3, Italians appear to be, overall, more demanding than other EU citizens, therefore we can suppose that Italians exert more 'pressure' on their government to reach objective outcomes. However, in the case of Italy, people's pressure does not correspond to satisfactory government outcomes, as represented by the IT position in the graph. This gap exacerbates the frustration of people and the resulting "welfare gap".

material living conditions in determining people's well-being. It should be stressed that those results are rather stable in all the countries and macroregions considered. An important implication of those subjective outcomes is that income buys only 'some' happiness. This conclusion confirms the importance of devising new methods for measuring well-being and social progress, as recommended by the Stiglitz-Sen-Fitoussi report. This new approach may help governments and policymakers to better design policies, focusing on different dimensions that affects people's well-being. It would complement the information provided by GDP as the leading indicator. In this respect, looking at the relationship between objective and subjective welfare measures is key for a better understanding of social welfare.

From the comparison between the objective and subjective BLI dimensions' weights and rankings, estimated on the basis of the two OECD datasets utilized, it emerges that there is a wide difference between them. This reflects the distance between governments' welfare outcomes (objective measure) and individual welfare levels, as per people's stated preferences (subjective measure). We consider this difference as a mismatch between what people desire and what government policies achieve in terms of welfare outcomes. This gap is at the core of the theoretical model we presented and it could help explain the anti-establishment sentiment that has affected our societies in the latest years, also as a consequence of the recent and acute economic crisis.

The estimation of the predicted welfare scores for different countries and macroregions allows for a geographical comparison in terms of objective and subjective welfare measures, which is used to derive the resulting "welfare gaps" reported in Figure 3. Contrary to the situations recorded in 2012 in the US and the UK, in Italy and Spain very high welfare aspirations were associated with low outcomes achieved by their governments in the same year. This large gap may frustrate citizens' well-being expectations.

7 Bibliography

Acock, A. C. (2013). *Discovering Structural Equation Modeling Using Stata.* Revised Edition. Stata Press.

Agresti, A. (2002). *Categorical Data Analysis.* Second Edition. New York: Wiley

Akaike, H. (1987). *Factor analysis and AIC.* Psychometrika, 52, 317–332.

Becchetti, L., Corrado, L. and Fiaschetti, M. (2017). *The Regional Heterogeneity of Wellbeing 'Expenditure' Preferences: Evidence from a Simulated Allocation Choice on the BES Indicators.* Journal of Economic Geography, 17(4), 857–891.

Benjamin, D. J., Heffetz, O., Kimball, M. S. and Rees-Jones, A. (2012). *What Do You Think Would Make You Happier? What Do You Think You Would Choose?* American Economic Review, 102(5), 2083–2110.

Benjamin, D. J., Heffetz, O., Kimball, M. S. and Szembrot, N. (2014). *Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference.* American Economic Review, 104(9), 2698-2735.

Bentler, P. M. (1990). *Comparative fit indexes in structural models.* Psychological Bulletin, 107, 238–246.

Bentler, P. M. and Ullman, J. B. (2013). *Structural equation modeling.* In Weiner, I. B. (Ed.), Handbook of Psychology, Chapter 23, 661-690, Second edition, New York: Wiley.

Bentler, P. M. and Yuan, K.-H. (1999). *Structural equation modeling with small samples: Test statistics.* Multivariate Behavioral Research, 34(2), 181-197.

Boarini, R. and Mira d'Ercole, M. (2013). *Going beyond GDP: An OECD perspective.* Fiscal studies, 34(3), 289-314.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K. A. (1990). *Overall fit in covariance structure models: Two types of sample size effects.* Psychological Bulletin, 107(2), 256-259.

Bollen, K. A. and Brand, J. E. (2008). *Fixed and random effects in panel data using structural equation models.* California Center for Population Research online Working paper Series.

Bollen, K. A. and Grace, J. B. (2005). *Interpreting the results from multiple regression and structural equation models.* Bulletin of the Ecological Society of America, 86, 283-295.

Bollen, K. A. and Pearl, J. (2013). *Eight myths about causality and Structural Equation Models.* In S.L. Morgan (Ed.), Handbook of Causal Analysis for Social Research, Chapter 15, 301-328, Springer.

Bollen, K. A. and Stine, R. A. (1992). *Bootstrapping goodness-of-fit measures in structural equation models.* Sociological Methods and Research, 21, 205-229.

Campbell, R. (2004). *Gender, Ideology and Issue Preference: Is There such a Thing as a Political Women's Interest in Britain?.* The British Journal of Politics & International Relations, 6, 22-44.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second Edition). Hillsdale, NJ: Lawrence Earlbaum Associates.

Decancq, K, and Lugo, M. A. (2013). *Weights in Multidimensional Indices of Well-Being: An Overview.* Econometric Reviews 32 (1), 7–34.

Doornik, J. A. and Hansen, H. (2008). *An omnibus test for univariate and multivariate normality.* Oxford Bulletin of Economics and Statistics, 70, 927-939.

Efron, B. (1979). *Bootstrap methods: Another look at the jackknife.* Annals of Statistics, 7, 1–26.

Fleurbaey, M. (2009). *Beyond GDP: The Quest for a Measure of Social Welfare.* Journal of Economic Literature 47 (4): 1029–75.

Fleurbaey, M. and Blanchet, D. (2013). *Beyond GDP: Measuring Welfare and Assessing Sustainability*. Oxford: Oxford University Press.

Fisher, R. A. (1925). *Statistical Methods for Research Workers* (First Edition). Oliver & Boyd, Edinburgh, UK.

Hooper, D.,Coughlan, J. and Mullen, M. (2008). *Structural Equation Modelling: Guidelines for determining model fit*. Electronic Journal of Business Research methods, 6(1), 53-60.

Hoyle, R. H. (1995). *The structural equation modeling approach: Basic concepts and fundamental issues*. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications*. 1-15. Thousand Oaks, CA: Sage Publications.

Hoyle, R. H. (2000). *Confirmatory factor analysis*. In Tinsley, H. E. A. and Brown, S. D. (Eds.), *Handbook of applied multivariate statistics and mathematical modeling*, 465–497. New York: Academic Press.

Hu, L. and Bentler, P. M. (1999). *Cutoff criteria for fit indexes in covariance structure analysis: Concentional criteria versus new alternatives*. Structural Equation Modeling, 6(1), 1-55.

Iacobucci, D. (2010). *Structural equation modeling: Fit indices, sample size, and advanced topics*. Journal of Consumer Psychology, 20, 90-98.

Jackson, D. L. (2003). *Revisiting sample size and number of parameter estimates: Some support for the N:q hypothesis*. Structural Equation Modeling, 10(1), 128-141.

Jones, C.I. and Klenow, P.J. (2016). *Beyond GDP? Welfare across Countries and Time*. American Economic Review, 106(9), 2426-2457.

Joreskog, K.G and Goldberger, A. S. (1975). *Estimation of a model with Multiple Indicators and Multiple Causes of a single latent variable*. Journal of American Statistical Association, Vol. 70, No. 351, 631-639.

Kahneman, D. and Krueger, A. B. (2006). *Developments in the Measurement of Subjective Well-Being*. Journal of Economic Perspectives, 20(1), 3–24.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. The Guilford press, Third edition.

Krishnakumar, J. and Nadar, A. (2008). *On exact statistical properties of multidimensional indices based on principal components, factor analysis, MIMIC and structural equation models*. Social Indicators Research, 86(3), 481-496.

Mardia, K. V. (1970). *Measures of multivariate skewness and kurtosis with applications*. Biometrika, 57, 519–530.

Mardia, K. V. (1985). *Mardia's test of multinormality*. In S. Kotz and Johnson, N. L. (Eds.), *Encyclopedia of statistical sciences*, 5, 217–221, New York: Wiley.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). *Sample size in factor analysis*. Psychological Methods, 4, 84–99.

Nachtigall, C., Kroehne, U., Funke, F. and Steyer, R. (2003). *(Why) Should we use SEM? Pros and Cons of Structural equation modeling*. Methods of Psychological Research online, Vol. 8, No.2, pp. 1-22.

Nevitt, J., and Hancock, G. R. (2001). *Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling*. Structural Equation Modeling, 8, 353–377.

Noorbakhsh, F. (1998). *The Human Development Index: Some technical issues and alternative indices*. Journal of International Development, 10, 589-605.

Nordhaus, W. D., and Tobin, J. (1973). *Is Growth Obsolete?*. In Moss, M. (ed.), *The Measurement of Economic and Social Performance*. Studies in Income and Wealth, 38, 509–31. New York: National Bureau of Economic Research.

Nussbaum, M. C. (2000). *Women and Human Development: The Capabilities Approach.* Cambridge, Cambridge University Press.

OECD (2011). *How's Life?: Measuring well-being.* OECD Publishing, Paris.

OECD (2013). *How's Life? 2013: Measuring well-being.* OECD Publishing, Paris.

OECD (2015). *How's Life? 2015: Measuring well-being.* OECD Publishing, Paris.

OECD (2017). *How's Life? 2017: Measuring well-being.* OECD Publishing, Paris.

Olsson, U. H., Foss, T., Troye, S. V. and Howell, R. D. (2000). *The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality.* Structural Equation Modeling, 7, 557–595.

Pearl, J. (2012). *The causal foundations of Structural equation modeling.* In Hoyle, R. H. (Ed.), *Handbook of Structural equation modeling.* Chapter 5, 68-91, New York: Guilford press.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. (2004). *Generalized multilevel structural equation modeling.* Psychometrika, 69, 167–190.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. (2005). *Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects.* Journal of Econometrics 128, 301–323.

Raiser, M., Di Tommaso, M. L. and Weeks, M. (2007). *Home grown or imported? Initial conditions, external anchors and the determinants of institutional reforms in the transition economies.* The Economic Journal, 117, 858-881.

Rawls, J. (1971). A Theory of Justice. Cambridge, Harvard University Press.

Satorra, A., and Bentler, P. M. (2001). *A scaled difference chi-squared test statistic for moment structure analysis.* Psychometrika, 66(4), 507-514.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., and King, J. (2006). *Reporting structural equation modeling and confirmatory factor analysis results: A review.* Journal of Educational Research, 99(6), 323-337.

Sen, A. (1985). *Commodities and Capabilities.* Oxford University Press.

Sen, A. (1992). *Inequality Reexamined.* Cambridge, Harvard University Press.

Skrondal, A., and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Boca Raton, FL: Chapman & Hall/CRC.

Skrondal, A., and Rabe-Hesketh, S. (2005). *Structural Equation Modeling: Categorical variables.* Encyclopedia of Statistics in Behavioral Science, New York: Wiley.

Skrondal, A., and Rabe-Hesketh, S. (2009). *Prediction in multilevel generalized linear models.* Journal of the Royal Statistical Society (A), 172, 659-687.

Stiglitz, J., Sen, A. and Fitoussi, J. (2009). *Report by the Commission on the Measurement of economic performance and social progress.* Available at: http://www.stiglitz-sen-fitoussi.fr

Ullmann, J. B. (2006). *Structural equation modeling: Reviewing the basics and moving forward.* Statistical Developments and Applications, 87(1), 35-50.

Ullman, J. B. (2007). *Structural equation modeling.* In Tabachnick, B. J. and Fidell, L. S. (Eds.), *Using multivariate statistics.* Chapter 14, 676-780, Pearson education, Fifth edition.

Westland, J.C. (2010). *Lower bounds on sample size in structural equation modeling.* Electronic Commerce Research and Applications, 9(6), 476-487.

Westland, J.C. (2012). *Erratum to "Lower bounds on sample size in structural equation modeling".* Electronic Commerce Research and Applications, 11, 445.

Table 6: Appendix I - Objective welfare measure. Headline indicators underlying the 11 dimensions of the objective BLI

Income and wealth (iw)	**Jobs and earnings (je)**	**Housing (ho)**	**Education and skills (es)**
Household net adjusted income	Employment rate	Number of rooms per person	Educational attainment
Household net financial wealth	Personal earnings	Housing expenditure	Years in education
	Job tenure	Dwellings with basic facilities	Students' cognitive skills
	Long-term unemployment rate		
Health status (hs)	**Work-life balance (wl)**	**Civic engagement (cg)**	**Environmental quality (eq)**
Life expectancy at birth	Employees working very long hours	Consultation on rule-making	Satisfaction with water quality
Self-reported health	Time devoted to leisure and personal care	Voter turn-out	Air pollution
Pesonal security (ps)	**Social connections (sc)**	**Subjective well-being (sw)**	
Reported homicides	Social network support	Life Satisfaction	
Self-reported victimisation			

Note: Following the OECD Better Life Index conceptual framework, Income and wealth, Jobs and earnings and Housing are defined as 'Material living conditions' whilst the remaining dimensions fall within the category of 'Quality of life'. See OECD (2011; 2013) for more details on the selected variables.

A Appendix II - Objective welfare measure: Structural Equation Modeling in small samples

In order to estimate the objective welfare measure from the SEM analysis, it is key to establish if our small sample of 35 observations is sufficient to detect the 'effects' or relationships specified in our model, given its complexity. In contrast with some simplistic rules of thumb on this topic, SEM models can perform well, even with small samples (e.g., 50 observations or even less).[44] The best way to determine the minimum sample size required for a specified model is to conduct a power analysis. In this regard, Westland (2010) developed an algorithm[45] to assess the lower bounds on sample size in SEM, as a function of minimum effect size (δ) in estimating the latent variable at a given statistical significance and power level (α; $1-\beta$).[46]

Table 7: Power analysis - SEM a priori sample size lower bound

Number of latent variables = 1
Number of observed variables = 12
Anticipated effect size (δ) = 0.157
Statistical significance (α) = 0.05
Statistical power level (1-β) = 0.8
Minimum sample size to detect effect = 35

Based on Westland's (2010) algorithm, as shown in Table 7,[47] considering 12 observed variables and 1 latent variable included in our SEM model, setting - as usual - a statistical power level at 0.8 ($1-\beta$) and a statistical significance at 0.05 (α), we can state that our small sample of 35 observations allows us to conduct a reliable SEM analysis because the minimum absolute anticipated effect size (δ) detected by our model is 0.157. Notably, an effect size of 0.157 means that our model can detect even small effects and relationships across the considered indicators, so that the resulting SEM estimates can be considered accurate and reliable.[48]

[44]If the variables are reliable, the effects are strong and the model is not overly complex, even smaller samples will suffice (Bollen, 1990). According to some studies, strong and clean measures - as defined by the number of variables loading on each factor and reliable measured variables - would be somewhat compensatory for sample size (Jackson, 2003).

[45]Westland (2010) developed a statistical algorithm to compute a lower bound on sample size in structural equation models assuming that observations were normally distributed. The significance level (α) was set to a default of 0.05, as suggested by Fisher (1925) and power ($1-\beta$) was set to 0.8, as suggested by Cohen (1988). A corrected software implementation of the paper's algorithm has been provided by Soper on his statistical calculator website at www.danielsoper.com/statcalc3/calc.aspx?id=89 (Westland, 2012).

[46]In an a priori form, the Westland algorithm detects the sample size lower bound, given the minimum effect size to detect. The sample size obtained indicates the minimum number of observations required to assure the existence or non-existence of a minimum effect (correlation) on each latent variable in the SEM.

[47]In the Table 7, α is the Sidak-corrected Type I error rate, β is the Type II error rate.

[48]The effect size (δ) is a basic indicator to asses the magnitude of the effects and interrelations that our model is able to detect. Cohen (1988) outlined criteria for interpreting the effect size. According to the thresholds proposed by Cohen, an effect size (correlation) $\delta = 0.10$, $\delta = 0.30$ or $\delta = 0.50$ corresponds to small, medium and large effects. Notice that, the smaller the better.

The presence of missing values in our dataset is managed through the use of the Maximum Likelihood with Missing Values ($MLMV$) method.[49] This method allow us to minimize the loss of information implied by the listwise deletion, the default setting in the standard ML estimation. The majority of the estimation approaches used in SEM assume multivariate normality (i.e., the joint distribution of the variables is distributed normally) and independent errors. In order to test for multivariate normality, we make use of specific tests as shown in Table 8.

Table 8: Multivariate normality (MVN) tests

Mardia mSkewness = 75.20	
$\chi^2(286) = 282.79$	Prob$>\chi^2 = 0.54$
Mardia mKurtosis = 128.68	
$\chi^2(1) = 3.41$	Prob$>\chi^2 = 0.07$
Doornik-Hansen	
$\chi^2(22) = 30.85$	Prob$>\chi^2 = 0.10$

The Doornik-Hansen test (2008) for multivariate normality is based on the skewness and kurtosis of multivariate observations that are transformed to insure independence, and then these are combined into an approximate χ^2 statistic. From the reading of the table above, in our model the Doornik-Hansen test cannot reject the null hypothesis of multivariate normality confirming that the data on which our analysis is based on are multivariate normal. Considering Mardia's test (Mardia,1970; 1985) for multivariate normality reported in Table 8, we can state that the data do not present kurtosis and skewness. Again, we cannot reject the null hypothesis of multivariate normality, confirming the results obtained by the Doornik-Hansen test. Since $p > 0.05$ for all the three tests reported, the null hypothesis that data are multivariate normal cannot be rejected and has to be retained. Even though our data are multivariate normal, a dataset of 35 observations can be considered as a very small sample.
However, the SEM approach is based on covariances which are less stable when estimated from small samples. Parameter estimates and chi-squared tests of fit are also very sensitive to sample size. In order to deal with the limitation deriving from the small sample size, in our analysis we make use of (non-parametric) bootstrapping to improve the stability and robustness of the parameters estimates and reduce the standard errors bias on which many test-statistics are based.[50] Bootstrapping is a computer-based method of resampling developed by Efron (1979). It is an increasingly popular approach to correct standard errors with increasing application in SEM.

[49] The $MLMV$ method within STATA assumes joint normality of all variables and missing values are assumed to be missing at random (MAR).

[50] Resampling (with replacement) of the observed data is called bootstrapping or non-parametric bootstrapping. It assumes that population and sample distributions have the same shape. Parameters, standard errors, and model test statistics are estimated with empirical sampling distributions from large numbers of generated samples, in our case 1,000 replications. The simulation work done by Nevitt and Hancock (2001) suggests that, in terms of bias, a standard 'naïve' bootstrap seems to work at least as well as robust adjustments to standard errors. New test statistics for robust estimation of SEM when based on small samples have been developed by Bollen and Stine (1992), Bentler and Yuan (1999), Satorra and Bentler (2001).

B Appendix III - Objective welfare measure: Model estimation and model evaluation in SEM

Maximum Likelihood (ML) estimation is usually the default method in most programs because of its statistical properties.[51] Most structural equation models described in the literature are analysed with this method, also in the generalized form (Olsson et al., 2000; Krishnakumar and Nadar, 2008). Indeed, the use of an estimation method other than ML requires explicit justification (Hoyle, 2000). The criterion used in the ML estimation -or the fit function -, minimizes the discrepancy between the sample covariances and the population variance-covariance matrix predicted by the research model. The main hypothesis of a structural equation model is that the covariance matrix of the observed variables, $\mathbf{S}$, may be parametrised with a parameter vector $\boldsymbol{\theta}$ based on a given model specification. The ML fit function $F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ to be minimized has the following form:

$$F_{ML}(\mathbf{S}, \boldsymbol{\Sigma}(\boldsymbol{\theta})) = \ln \mid \boldsymbol{\Sigma}(\boldsymbol{\theta}) \mid - \ln \mid \mathbf{S} \mid + tr\left[\mathbf{S}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})\right] - \boldsymbol{\Lambda}^{o} \tag{14}$$

where $\mathbf{S}$ is the sample (observed) variance-covariance matrix of the measured variables, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is the population variance-covariance matrix implied by the model, $\boldsymbol{\theta}$ is the vector of independent parameters and $\boldsymbol{\Lambda}^{o}$ the matrix of structural parameters corresponding to the observed indicators. Most forms of ML estimation in SEM are simultaneous, which means that the estimates of the model parameters are calculated all at once. In our analysis we refer to a full-information ML estimation.[52]

In order to assess the model fit, a chi-squared test (χ^2) is always reported as the default overall goodness-of-fit indicator in SEM analysis.[53] It measures the discrepancy between the sample and the fitted covariance matrices. If the model fits the data, a non-significant χ^2 is desirable. In a good-fitting model the ratio of the chi-squared to the degrees of freedom (χ^2/df) is less than 2 (or even 3) (Schreiber et al., 2006). The model chi-squared test (χ^2_M) has some important limitations.[54] Different fit indices have been developed which look at model fit while eliminating or minimizing the effect of sample size.

There are different classes of fit indices. A bundle of the most popular statistics in the different classes is usually reported for evaluating the model correctly. All the indices described in Table 9 are generally available under default ML estimation (Iacobucci, 2010).

In the class of comparative fit indices, the Root Mean Square Error of Approximation (RMSEA) estimates the lack of fit of a model compared to a perfect (or satured) model. It is scaled in the same way as a badness-of-fit index where a value of zero indicates the best fit. It is also a parsimony-adjusted index. The RMSEA follows a noncentral χ^2 distribution, where the noncen-

[51] When all statistical requirements are satisfied and the model is correctly specified, ML estimates in large samples are asymptotically unbiased, efficient and consistent.

[52] Computer implementation of the ML estimation is typically iterative, which means that, once we derive an initial solution - or starting values - then the method attempts to improve these estimates until convergence. For overidentified models, the fit of model to the data may be imperfect, but iterative estimation will continue until the improvements in the model fit fall below a preset minimum value to achieve convergence.

[53] The basic model test statistic is given by $(N-1)F_{ML}$ where F_{ML} is the value of the statistical criterion (fit function) minimized in the ML estimation and $(N-1)$ is one less than the sample size. In large samples and assuming multivariate normality, the product $(N-1)F_{ML}$ follows a central χ^2 distribution with degrees of freedom given by the model specification, df_M. This statistic is referred to as the model chi-squared (χ^2_M). It is also known as the likelihood ratio χ^2 or generalized likelihood ratio. For an overidentified model, χ^2_M tests the exact-fit hypothesis, or the prediction that there is no discrepancy between the population covariances and those predicted by the model.

[54] Among these limitations, χ^2 values are dependent on the sample size. In models with large samples, trivial differences often cause the χ^2 to be significant solely because of sample size.

trality parameter allows for discrepancies between model-implied and sample covariances up to the level of the expected value of χ^2, or dfs. Values of 0.06 to 0.08 or less indicate a close-fitting model (Schreiber et al., 2006).

Table 9: Goodness-of-fit tests

Likelihood ratio (Absolute fit index)
$\chi^2_M(87) = 109.884$
$p>\chi^2 = 0.049$
Relative $\chi^2(\chi^2/df) < 2:1$
Population error
RMSEA = 0.087
90% CI, lower bound = 0.005; upper bound 0.133
pclose = 0.143 (Probability RMSEA <=0.05)
Baseline comparison
CFI = 0.914
TLI = 0.935
Size of residuals
CD = 0.941

RMSEA=Root mean squared error of approximation; CFI=Comparative fit index; TLI=Tucker-Lewis index; CD=Coefficient of determination =R^2; $\chi^2_M = Model\chi^2$

In the same class, the Bentler's Comparative Fit Index (CFI; Bentler, 1990) assesses the fit of a given model relative to other models. It is an incremental fit index which measures the relative improvement in the fit of the proposed model over that of a baseline model, typically the independence model. The CFI employs the noncentral χ^2 distribution with noncentrality parameters. The larger the CFI the better the fit. The CFI lies in the to $0-1$ range, and it is a good indicator of model fit even in small samples. A CFI value greater than 0.95 is often indicative of good fitting models (Hu and Bentler, 1999).

In this class is also included the Non-Normed Fit Index (NNFI), also known as the Tucker-Lewis index (TLI). Values of TLI greater than 0.95 are indicative of good-fit. In Table 9 we report a collection of the main overall goodness-of-fit tests' values referred to our SEM model.[55]

Considering the relative threshold levels, from the combined analysis of the reported overall goodness-of-fit indices, we can conclude that the SEM model used to estimate objective BLI presents a good fit. This result is particularly positive and significant taking into account the small sample size on which all the estimates are based on. Notably, the relative χ^2 - defined as the ratio of χ^2 over degree of freedom - is less than two, CFI and TLI are close to 0.95 and the RMSEA is 0.087.

As suggested by Kline (2011), one should also inspect the matrix of correlation of the residuals and describe their pattern as part of a diagnostic assessment of fit. In this regard, we make use of equation-level goodness-of-fit statistics to test the reliability of each path considered in our analysis. Their values for our model are reported in Table 10.

[55] The Standardized root mean square residual (SRMR), in the class of residual-based fit indices, is not reported in Table 9 because of missing values.

Table 10: Equation level Goodness-of-fit tests

Observed variables	R^2	mc	mc^2
Subjective well-beiing (sw)	0.485	0.696	0.485
Income and wealth (iw)	0.528	0.727	0.528
Jobs and earnings (je)	0.859	0.927	0.859
Housing (ho)	0.707	0.841	0.707
Work-life balance (wl)	0.256	0.506	0.256
Health status (hs)	0.712	0.844	0.712
Education and skills (es)	0.337	0.581	0.337
Social connections (sc)	0.417	0.645	0.417
Civic engagement (cg)	0.192	0.438	0.192
Environmental quality (eq)	0.352	0.594	0.352
Pesonal security (ps)	0.358	0.599	0.358
overall	0.941		

Note: mc = correlation between the dependent variable and its prediction; mc^2 = $Bentler - Raykovsquared$ multiple correlation coefficient.

Reliability is defined in the classic sense, as the proportion of true variance relative to total variance. Both reliability and the proportion of variance of a measured variable are assessed through squared multiple correlation (mc^2) and R^2, where the measured variable is the independent variable (IV) and the factor is the dependent variable (DV), that is the latent factor for BLI.[56] Notably, each mc^2 is interpreted as the reliability of the measured variable in the analysis and R^2 as the proportion of variance in the variable accounted for by the factor. From the analysis of Table 10, it emerges that the reliability of Civic engagement (cg), Work and life balance (wl), Personal security (ps) and Education and skills (es) is relatively weak in explaining the latent factor for Objective BLI.[57]

The main outcome emerging from the R^2 values in the Table 10 is that the overall variance accounted by our model is 94.1% of the total variance,[58] indicating that the model contains almost all the relevant dimensions explaining people's well being as measured by the latent factor for Objective BLI.

[56] It should be stressed that the equation for mc^2 is applicable only when there are no complex factor loadings or correlated errors.

[57] It should be highlighted that, for the latter three indicators –Work-life balance (wl), Personal security (ps) and Education and skills (es)- this limited reliability is combined with an insufficient statistical significance indicated by high $p - value$ levels for the unstandardized estimation, as reported in Table 1 as opposed to an higher reliability of Jobs and earnings (je), Health status (hs), Housing condition (ho) and Income and Wealth (iw) in the same table.

[58] The overall R^2 value of 94.1% corresponds to the Coefficient of determination (CD) value reported in Table 9, an index accounting for the size of residuals.

C Appendix IV - Subjective welfare measure: Multivariate normality tests and model evaluation

As expected, the Multivariate normality tests reported in Table 11 confirm that data are multivariate non-normal.[59] Since the p-values are <0.05 for all the tests reported, the null hypothesis that data are multivariate normal can be rejected. A generalized method or bootstrapping dealing with non-normality is needed for a good and robust econometric analysis.

Table 11: Multivariate normality (MVN) tests

Mardia mSkewness = 5.350	
$\chi^2(286) = 11351.21$	Prob>$\chi^2 = 0.00$
Mardia mKurtosis = 193.04	
$\chi^2(1) = 27858.91$	Prob>$\chi^2 = 0.00$
Doornik-Hansen	
$\chi^2(22) = 2745.01$	Prob>$\chi^2 = 0.00$

If the data are categorical then the assumption of MVN distribution underlying SEM model is not met. In order to deal with this limitation, we have two possibilities: estimating the model using a SEM with robust standard errors (bootstrapping), as done in the previous section, or estimating the model with a Generalized SEM model (GSEM). The latter is the method we selected for our econometric analysis of Subjective BLI.

After the estimation of our GSEM MIMIC model, we need to make a further step in our analysis related to the model evaluation. In other words, we are interested in assessing if the model estimated through GSEM MIMIC is also a good model in terms of fit. We cannot directly answer this question because of the limitation of goodness-of-fit indices availability under GSEM.[60] Therefore, we propose an indirect method which use two different models running on the same dataset – bootstrapped SEM and GSEM MIMIC – comparing them through their relative Akaike Information Criterion (*AIC*; Akaike, 1987), a predictive fit index available for both methods. Smaller *AIC* values indicate a good-fitting and parsimonious model.

When using a GSEM estimation instead of SEM, we can observe a significant improvement in the overall fit of the model. Taking into account the SEM goodness-of-fit indices reported in the tables of Appendix V, we can state that the fit of the model for the countries and regions considered is slightly under the acceptance thresholds for all of them. But the *AIC* of GSEM is always 25-30% lower than the SEM *AIC*. Therefore, we can reasonably conclude that the GSEM model overcomes the acceptance cut-off values indicated in the literature[61] for all the countries and regions considered. This implies that the GSEM estimations ensure a good fit of the models.

The Akaike Information Criterion (*AIC*), a predictive fit index, falls also within the category of the parsimony-adjusted indices because it may favour simpler models. The *AIC* is applicable

[59] Notice that the MVN tests are based on the full OECD dataset.

[60] Most of SEM post-estimation tests and indices are not available after GSEM because of the assumption of joint-normality of the observed variables.

[61] According to Hooper et al. (2008), the cut-off criteria for acceptable model fit are: values greater 0.9 for CFI; values less than 0.07 for RMSEA; values less than 0.08 for SRMR. Low χ^2 relative to degrees of freedom, with an insignificant p-value, is the criterion to assess the absolute fit of a model.

to models estimated with Maximum Likelihood methods. The AIC formula presented in the SEM literature to which we refer is:

$$AIC = \chi^2_M - 2df_M \tag{15}$$

where χ^2_M is the model chi-squared, known as the likelihood ratio χ^2 or generalized likelihood ratio.[62] The index decreases the χ^2_M by a factor of twice the model degrees of freedom. The χ^2 value is the traditional measure for evaluating the overall model fit described in Appendix III (Hu and Bentler, 1999). If $\chi^2_M = 0$, the model perfectly fits the data (each observed covariance equals its counterpart implied by the model). If the fit of an overidentified model, which is not correctly specified, becomes increasingly worse, then the value of χ^2_M increases. Therefore, χ^2_M is scaled as a 'badness-of-fit' statistic.

The key is that the relative change in the AIC is a function of model complexity. It should be noted that the relative correction for parsimony of the AIC becomes smaller and smaller as the sample size increases (Kline, 2011). Smaller values correspond to a good-fitting and parsimonious model. Specifically, the selected model will present relatively better fit and fewer free parameters, compared with competing models. It should be stressed that there is no fixed threshold value for the AIC. Therefore, 'small' is intended as a relative term to compare with a second model AIC. This method is useful for cross-validation because it is not dependent on sample data (Ullmann, 2007).

[62] The Akaike Information Criterion can also be expressed as follows: $AIC = -2\log L(\boldsymbol{\theta}) + 2df_M$, with $L(\boldsymbol{\theta})$ being the Likelihood function.

Table 12: Appendix V - Subjective welfare measure. GSEM MIMIC and SEM estimated parameters (unstandardized) - OECD Full sample; OECD Male; OECD Female

	OECD				OECD M				OECD F			
Indicators	**GSEM**		**SEM**		**GSEM**		**SEM**		**GSEM**		**SEM**	
Income and wealth (iw)	0.725***	(0.017)	0.777***	(0.015)	0.731***	(0.022)	0.789***	(0.019)	0.734***	(0.028)	0.772***	(0.025)
Jobs and earnings (je)	0.876***	(0.020)	0.903***	(0.015)	0.881***	(0.025)	0.902***	(0.019)	0.872***	(0.032)	0.907***	(0.026)
Housing (ho)	0.878***	(0.020)	0.880***	(0.015)	0.876***	(0.025)	0.884***	(0.019)	0.892***	(0.033)	0.879***	(0.026)
Health status (hs)	1.252***	(0.027)	1.068***	(0.014)	1.264***	(0.036)	1.075***	(0.018)	1.231***	(0.044)	1.051***	(0.024)
Social connections (sc)	0.916***	(0.020)	0.939***	(0.013)	0.929***	(0.026)	0.931***	(0.017)	0.882***	(0.032)	0.935***	(0.023)
Education (es)	1.059***	(0.024)	1.022***	(0.015)	1.068***	(0.030)	1.026***	(0.019)	1.049***	(0.038)	1.017***	(0.025)
Environmental quality (eq)	1.003***	(0.022)	0.997***	(0.014)	1.033***	(0.029)	1.007***	(0.018)	0.973***	(0.035)	0.993***	(0.025)
Personal security (ps)	0.956***	(0.021)	0.985***	(0.015)	0.965***	(0.027)	0.984***	(0.019)	0.946***	(0.034)	0.990***	(0.026)
Work-life balance (wl)	0.906***	(0.020)	0.974***	(0.014)	0.885***	(0.025)	0.957***	(0.017)	0.934***	(0.033)	0.992***	(0.023)
Civic engagment (cg)	0.880***	(0.020)	0.858***	(0.014)	0.893***	(0.025)	0.861***	(0.018)	0.861***	(0.031)	0.861***	(0.026)
Subjective well-being (sw)	1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)	
Control variables	*Causes*		*Cov.*		*Causes*		*Cov.*		*Causes*		*Cov.*	
world region	0.082***	(0.017)	0.024***	(0.006)	0.080***	(0.022)	0.022*	(0.009)	0.084***	(0.026)	0.026**	(0.009)
country	-0.005***	(0.001)	-0.420***	(0.099)	-0.006***	(0.001)	-0.629***	(0.146)	-0.003*	(0.001)	-0.188	(0.145)
age	0.007	(0.007)	-0.0009	(0.011)	-0.007	(0.010)	-0.001	(0.017)	0.030**	(0.011)	0.042*	(0.017)
gender	0.201***	(0.021)	0.050***	(0.004)	-		-		-		-	
Fit indices												
$R^2 overall(CD)$	0.892				0.895				0.884			
χ^2_M	(84), 7044.2				(74), 4086.6				(74), 3025.4			
CFI	0.876				0.882				0.862			
RMSEA	0.081				0.085				0.088			
SRMR	0.043				0.044				0.048			
AIC (SEM)	626843				364547				243876			
AIC (GSEM)	399885				238890				160617			
Observations	12703				7501				5203			
logLikelihood (SEM)	-313385				-182238				-121902			
logLikelihood (GSEM)	-199872				-119376				-80239			

BLI path coefficients without parentheses
Standard errors in round parentheses
*p<0.05; **p<0.01; ***p<0.001

Data source: OECD Your Better Life Index microdata (year 2012)

RMSEA=Root mean squared error of approximation; CFI=Comparative fit index; TLI=Tucker-Lewis index; CD=Coefficient of determination $=R^2$; $\chi^2_M = Model \chi^2$.
Note: The GSEM reported is an ordered probit MIMIC model (Multiple Indicators Multiple Causes), whilst the SEM considered is a bootstrapped SEM (500 reps.) with robust standard errors.

Table 13: Appendix V - Subjective welfare measure. GSEM MIMIC and SEM estimated parameters (unstandardized) - USA; European Union; UK

	USA				EU				UK			
Indicators	**GSEM**		**SEM**		**GSEM**		**SEM**		**GSEM**		**SEM**	
Income and wealth (iw)	0.544***	(0.040)	0.683***	(0.048)	0.719***	(0.023)	0.769***	(0.022)	0.844***	(0.109)	0.880***	(0.098)
Jobs and earnings (je)	0.775***	(0.050)	0.917***	(0.045)	0.866***	(0.027)	0.908***	(0.022)	0.886***	(0.112)	0.947***	(0.109)
Housing (ho)	0.802***	(0.052)	0.893***	(0.049)	0.880***	(0.027)	0.881***	(0.021)	0.947***	(0.118)	0.945***	(0.108)
Health status (hs)	1.170***	(0.070)	1.101***	(0.042)	1.275***	(0.038)	1.075***	(0.020)	1.438***	(0.171)	1.169***	(0.102)
Social connections (sc)	1.004***	(0.062)	1.103***	(0.057)	0.888***	(0.027)	0.936***	(0.020)	1.042***	(0.122)	1.103***	(0.096)
Education (es)	0.951***	(0.060)	1.031***	(0.044)	1.098***	(0.033)	1.052***	(0.021)	1.178***	(0.141)	1.213***	(0.107)
Environmental quality (eq)	1.058***	(0.065)	1.129***	(0.051)	0.997***	(0.030)	0.995***	(0.022)	1.189***	(0.139)	1.109***	(0.095)
Personal security (ps)	0.961***	(0.060)	1.085***	(0.053)	0.968***	(0.029)	0.993***	(0.022)	1.110***	(0.131)	1.120***	(0.108)
Work-life balance (wl)	0.861***	(0.053)	1.041***	(0.048)	0.905***	(0.027)	0.967***	(0.021)	0.956***	(0.113)	1.027***	(0.093)
Civic engagment (cg)	0.850***	(0.054)	0.936***	(0.059)	0.925***	(0.029)	0.904***	(0.022)	0.949***	(0.117)	0.956***	(0.095)
Subjective well-being (sw)	1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)	
Control varables	*Causes*		*Cov.*		*Causes*		*Cov.*		*Causes*		*Cov.*	
country					-0.005**	(0.002)	-0.282***	(0.080)				
gender	0.159***	(0.047)	0.042***	(0.010)	0.195***	(0.024)	0.048***	(0.005)	0.262***	(0.078)	0.068***	(0.019)
age	-0.019	(0.015)	-0.053	(0.030)	0.006	(0.009)	-0.005	(0.015)	-0.003	(0.028)	0.018	(0.047)
Fit indices												
$R^2 overall(CD)$	0.863				0.876				0.862			
χ^2_M	(64), 1449.6				(74), 4092.4				(64), 449.1			
CFI	0.809				0.861				0.810			
RMSEA	0.104				0.085				0.103			
SRMR	0.062				0.048				0.064			
AIC (SEM)	81146				350891				22672			
AIC (GSEM)	63640				239627				18008			
Observations	2016				7551				563			
logLikelihood (SEM)	-40538				-175410				-11301			
logLikelihood (GSEM)	-31752				-119744				-8936			

BLI path coefficients without parentheses
Standard errors in round parentheses
*p<0.05; **p<0.01; ***p<0.001

Data source: OECD Your Better Life Index microdata (year 2012)

RMSEA=Root mean squared error of approximation; CFI=Comparative fit index; TLI=Tucker-Lewis index; CD=Coefficient of determination $=R^2$; $\chi^2_M = Model\chi^2$.
Note: The GSEM reported is an ordered probit MIMIC model (Multiple Indicators Multiple Causes), whilst the SEM considered is a bootstrapped SEM (500 reps.) with robust standard errors.

Table 14: Appendix V - Subjective welfare measure. GSEM MIMIC and SEM estimated parameters (unstandardized) - Germany; France; Italy; Spain

	DE				FR				IT				ES			
Indicators	**GSEM**		**SEM**		**GSEM**		**SEM**		**GSEM**		**SEM**		**GSEM**		**SEM**	
Income and wealth (iw)	0.764***	(0.071)	0.783***	(0.069)	0.715***	(0.040)	0.742***	(0.036)	0.679***	(0.066)	0.791***	(0.061)	0.734***	(0.108)	0.748***	(0.091)
Jobs and earnings (je)	0.965***	(0.085)	0.954***	(0.072)	0.892***	(0.047)	0.923***	(0.039)	0.758***	(0.075)	0.847***	(0.073)	0.559***	(0.090)	0.697***	(0.108)
Housing (ho)	0.960***	(0.083)	0.928***	(0.074)	0.908***	(0.048)	0.869***	(0.038)	0.747***	(0.072)	0.806***	(0.068)	0.770***	(0.113)	0.806***	(0.095)
Health status (hs)	1.395***	(0.117)	1.155***	(0.071)	1.260***	(0.065)	1.052***	(0.037)	1.089***	(0.107)	0.974***	(0.068)	1.125***	(0.164)	0.986***	(0.085)
Social connections (sc)	1.060***	(0.089)	1.022***	(0.064)	0.883***	(0.045)	0.942***	(0.033)	0.887***	(0.083)	0.976***	(0.060)	0.715***	(0.103)	0.778***	(0.095)
Education (es)	1.222***	(0.106)	1.070***	(0.077)	1.190***	(0.061)	1.083***	(0.038)	1.051***	(0.103)	1.010***	(0.065)	1.059***	(0.154)	0.994***	(0.092)
Environmental quality (eq)	1.045***	(0.091)	1.011***	(0.074)	0.918***	(0.048)	0.949***	(0.037)	1.098***	(0.107)	1.049***	(0.065)	0.878***	(0.126)	0.862***	(0.091)
Personal security (ps)	1.079***	(0.093)	1.050***	(0.075)	0.900***	(0.046)	0.924***	(0.037)	0.963***	(0.091)	1.014***	(0.074)	0.987***	(0.132)	0.973***	(0.070)
Work-life balance (wl)	0.912***	(0.078)	0.977***	(0.070)	0.915***	(0.047)	0.967***	(0.035)	0.862***	(0.082)	0.965***	(0.056)	0.897***	(0.122)	0.923***	(0.086)
Civic engagment (cg)	1.043***	(0.090)	0.943***	(0.066)	0.906***	(0.047)	0.880***	(0.039)	0.970***	(0.093)	1.048***	(0.076)	0.800***	(0.113)	0.804***	(0.067)
Subjective well-being (sw)	1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)		1 (constr.)	
Control variables	*Causes*		*Cov.*		*Causes*		*Cov.*		*Causes*		*Cov.*		*Causes*		*Cov.*	
gender	0.160**	(0.055)	0.040***	(0.012)	0.159***	(0.036)	0.044***	(0.008)	0,175	(0.099)	0.046*	(0.018)	0,047	(0.173)	0,0008	(0.034)
age	-0.023	(0.022)	-0.043	(0.032)	-0.009	(0.013)	-0.037	(0.026)	-0.019	(0.043)	-0.025	(0.040)	0.081	(0.062)	0.099	(0.088)
Fit indices																
$R^2 overall(CD)$	0.865				0.855				0.894				0.905			
χ^2_M	(64), 725.5				(64), 1607.2				(64), 474.7				(64), 200.5			
CFI	0.831				0.836				0.868				0.898			
RMSEA	0.097				0.091				0.096				0.087			
SRMR	0.058				0.056				0.052				0.051			
AIC (SEM)	42341				115842				26780				11290			
AIC (GSEM)	34426				93214				20626				8834			
Observations	1088				2881				691				279			
logLikelihood (SEM)	-21136				-57886				-13355				-5610			
logLikelihood (GSEM)	-17145				-46539				-10245				-4349			

BLI path coefficients without parentheses
Standard errors in round parentheses
*p<0.05; **p<0.01; ***p<0.001

Data source: OECD Your Better Life Index microdata (year 2012)

RMSEA=Root mean squared error of approximation; CFI=Comparative fit index; TLI=Tucker-Lewis index; CD=Coefficient of determination $=R^2; \chi^2_M = Model\chi^2$.
Note: The GSEM reported is an ordered probit MIMIC model (Multiple Indicators Multiple Causes), whilst the SEM considered is a bootstrapped SEM (500 reps.) with robust standard errors.

European Stability Mechanism

6a Circuit de la Foire Internationale
L-1347 Luxembourg

Tel: +352 260 292 0

www.esm.europa.eu

info@esm.europa.eu