New Frontiers in Measuring the Well-being in the Big Data Era

Daniele Toninelli[1], Michela Cameletti[1], Stephan Schlosser[2]
[1]University of Bergamo, via dei Caniana, 2, 24127 Bergamo, Italy; email: daniele.toninelli@unibg.it, michela.cameletti@unibg.it;
[2]Universität Göttingen, Goßlerstraße 19, 37073 Göttingen, Germany; email: sschlos1@uni-goettingen.de.

Abstract
The measurement of well-being (WB) is extremely challenging especially due to the multidimensional, country-specific and latent nature of this concept. Despite these difficulties, a timely estimation of WB is essential in order to support policy makers' decision processes and to obtain reliable measurements able to assess, a-posteriori, policies' effectiveness. In the last years a lot of effort has been dedicated to develop research projects aimed at measuring citizen well-being. In the recent era of "big data", many sources of information (e.g. the web or social networks) can contribute in producing enhanced, timely and less expensive estimates and indicators of WB that could integrate or update the available official statistics.

This paper represents a preliminary step towards this objective. The analysis focuses on 18 European countries, by using data from the European Social Survey (ESS) collected in 2016 and by applying SEM (Structural Equation Modelling) to variables covering the main subjective well-being (SWB) dimensions. In particular, we evaluate if a country-specific (i.e., local) model rather than a global European model is able to provide reliable estimates of the relative importance of well-being dimensions. This allows us to evaluate, by country, if and how much the main dimensions affect the SWB of citizens. Finally, we test if a "local approach" can also enhance the goodness of fit obtained estimating a global European model.

Key Words: subjective well-being, European Social Survey, cross-country study, official statistics, European countries, Structural Equation Modelling

1. Introduction

Nowadays Well-Being (WB) is considered one of the key variables for statistical agencies both at the national and international level. It is not just a proxy of the development and of the progress of a country, but it can also be used to evaluate the effectiveness of societal policies. In this framework, the WB concept is generally regarded as "a guiding principle for policymaking that pursues economic, social and ecological objectives simultaneously"[1]. Thus, recently the target of national statistical institutes has been set in order to fill the gap between WB metrics, policy-making process and policy intervention and evaluation: a challenging issue. In this context, the WB measurement became a key

[1] German Federal Government (2017), p. 4.

issue within the framework of official statistics. Consequently, in the last years more and more research programs were implemented, in order to obtain an accurate measure of the phenomenon, both at the national level and internationally.

For example, since 2011 the OECD has been developed the project "Better live initiative"[2], aimed at measuring the societal progress through the evaluation of improvements in the WB and in the living conditions of people and households. The society's progress measurement is based on eleven domains of WB, including jobs, health and housing, civic engagement and environment. The different dimensions of the phenomenon are further grouped by three distinct domains: material conditions, quality of life and sustainability. In particular, the OECD activity is oriented to measure the Subjective WB (SWB), nowadays widely considered "an essential part of measuring quality of life alongside other social and economic dimensions" (OECD, 2013a). In order to encourage this approach, the OECD is also producing several international guidelines providing advices that ranges from the data collection phase to the data analysis, as well as the use and publication of SWB measures[3].

There are other examples of national projects focused on the WB measurement, such as: several activities promoted by the Office for National Statistics[4], in UK; the Istat BES project[5], developed in Italy; the Programme of New Zealand's Official Social Statistics[6]; the works made by the Scientific Advisory Board, set by the German government in 2015[7].

Following the recommendations of the Commission on the Measurement of Economic Performance and Social Progress in 2009 (published in Stiglitz, Sen & Fitoussi, 2010), all these national WB projects have in common a similar approach. They moved from a "traditional" macro-economic statistics approach (based on macro-economic indicators - such as the GDP - and almost completely focused on the evaluation of economic systems), towards a more complete view, able to provide a reliable picture of living conditions and societal progress. This approach is multidimensional as it is based on the use of several indicators (covering different aspects of societal life) that, together, can provide an accurate measure of the multifaceted phenomenon of WB.

In this work, we focus on SWB, an extremely important concept able to "provide key information about people's quality of life" (Stiglitz, Sen & Fitoussi, 2010, p. 58). SWB is usually broadly defined as "good mental states, including all of the various evaluations, positive and negative, that people make of their lives, and the affective reactions of people to their experiences" (see, OECD, 2013b, p. 5).

The importance of research focused on SWB grew a lot in the last decades. According to Diener (2013), in 1981 there were just 131 publications about this topic, introducing mainly

[2] For further information about the project, see: http://www.oecd.org/statistics/measuring-well-being-and-progress.htm.
[3] See, for example: http://www.oecd.org/statistics/oecd-guidelines-on-measuring-subjective-well-being-9789264191655-en.htm.
[4] E.g., see: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing.
[5] BES Project's official website: http://www.misuredelbenessere.it/index.php?id=51.
[6] For further information see: (http://archive.stats.govt.nz/browse_for_stats/people_and_communities/Well-being.aspx?_ga=2.256140555.1024852905.1536395288-639848462.1536395288.
[7] In this regard, see the German Essays by the Scientific Advisory Board (https://www.gut-leben-in-deutschland.de/SiteGlobals/PL/23350043) and the documentation of Government Strategy on Wellbeing in Germany (https://www.gut-leben-in-deutschland.de/SiteGlobals/PL/21426409).

descriptive and cross-sectional studies, mainly focused on demographic correlates. In 2012 the number of studies increased to about 12,000, and there was a shift towards studies often developed with a longitudinal perspective, based on broadly representative samples (e.g., the Gallup World Poll[8]) and aimed at evaluating citizen experience not only based on the use of self-report scales. Nowadays most SWB measurement projects are based on very large-scale surveys, involving representative samples of one or more countries. Nevertheless, these big survey projects require a lot of time and resources to be implemented. However, in the last years new potentially interesting data sources have appeared that can be used for the purpose of SWB estimation. Huge amount of data can be collected, for example, through social networks or using new data collection techniques such as web-scraping. Despite a lot of research is still needed to face the criticalities, these relatively inexpensive and broad sources of data represent a potentially interesting tool for building new, enhanced, cheap and more timely indicators. However, we believe we should still be very careful, in dealing with such sources. Web and social media data, at present, cannot probably be considered, per-se, a completely reliable method for estimating the WB of a population (e.g. for problems linked to the population coverage and to the representativeness). Nevertheless, the integrated use of these new data together with more "traditional" data collected through survey projects can provide policy makers with more timely indicators about both society critical areas and the impact of public policies.

In this framework, our current research aims at testing the potentialities of this integrated data sources approach in estimating the level (local or national) of a latent concept such as the SWB. For this aim, we believe that Structural Equation Modelling (SEM) could be a reliable tool, since we aim at studying a latent concept. Nevertheless, before starting to work directly on web/social network data, in this paper we use ESS data to test the potentiality of SEM for SWB estimation. Moreover, we assess if at a local level (i.e., by country) the SWB estimates can be enhanced.

In particular, in this paper we study several indicators obtained from groups of items of the ESS questionnaire that are supposed to cover almost all the main dimensions of the SWB. In order to check if the use of SEM, applied to such indicators, allows us to estimate the SWB of the studied countries, we compare our estimates with direct measures of WB given by other items of the ESS questionnaire (that represent our benchmarks).

In particular, our main objective is to evaluate, estimating local structural equation models, how each dimension is affecting the latent concept of SWB, and if there is any significant difference between the studied countries. Preliminary results on a broader perspective (that is across all European countries) were encouraging us, despite we observed a general lack of fit of the obtained model (Toninelli, Cameletti & Schlosser, 2018). Nevertheless, in this paper we want also to assess if we can obtain a relative improvement, in terms of goodness of fit, shifting from a global European model to a series of local models.

2. Background

Although the nature of the variables (latent) harmonizes well with the objective of the method, SWB has been studied by means of SEM in only few studies.

---

[8] For further information about the Gallup Global Poll, see https://www.gallup.com/services/170945/worldpoll.aspx; for more details about the Gallup global well-being research, see https://news.gallup.com/poll/126965/gallup-global-wellbeing.aspx.

For example, Warner & Rasco (2014) applied SEM to data collected on a group of college students (*n*=847) focusing on predictors of positive and negative affect, in addition to predictors of satisfaction with life. They found that the model fit is enhanced, when the negative affect latent variable was treated as separate outcome, instead as one of the several indicators of SWB (once the scores are reversed). In our study we do not focus on the relative impact of negative vs positive affect; we rather rescale all the negative oriented scales in order to have a complete coverage of the main WB dimensions, using as much indicators as possible (depending on the availability of usable items within the ESS questionnaire). Moreover Warner & Rasco (2014) studied a very homogeneous sample (in terms of age, ethnicity and education), whereas we can count on data from a wider sample (34,836 respondents), considered being representative of the countries' population[9].

Other papers focused on very specific topics and relied on relatively small samples and/or on cross-sectional studies. For example, Oliver et al. (2009) used SEM to study the effect of actions and policies towards elderly adults in Dominican Republic. Zaidi et al. (2017) investigated well-being among older people in the UK by using SEM. Turashvili & Turashvili (2015) applied SEM in order to estimate the psychological well-being on a group of Georgian students. Lin & Yeh (2014) examined the link between gratitude, social support, coping style and well-being, studying 750 undergraduate students.
In comparison to the listed studies, our approach is broader and based on a large-scale survey involving several European countries[10]. In addition, we use a wide range of items contained in the ESS questionnaire (which allow us to cover almost all the WB dimensions). Moreover, we can potentially make use of a longitudinal approach (comparing results of the different waves of the ESS) in order to detect how SWB varies over time country by country and at the European level.

The first step of our research project (discussed in Toninelli, Cameletti & Schlosser, 2018) aimed at testing the capability of SEM in measuring the SWB. For this purpose, we studied data coming from the 8th wave of the ESS, collected in 2016 and published in the following year. The study involved 18 European countries. One of the advantages of studying ESS data is that two items of the questionnaire directly measure (on a 1 to 10 scale) the level of perceived SWB in terms of happiness (question C1) and satisfaction about respondent's own life (question B27). We merged these two variables into one variable (called *H-SAvg*), that is simply their average. A preliminary analysis showed, for *H-SAvg*, a high variability across countries (see also Figure 1), ranging from a minimum level of 5.95 (observed for Russia) to a maximum of 8.16 (for Switzerland); the general average for is 7.52 (st.dev.=0.565).

The advantage of having variables directly measuring the SWB is that in our study they can be used as benchmark. This means that we can check if and how much the estimates obtained using SEM are able to reproduce the relative level of SWB of the different European countries. The results of this first phase (see Toninelli, Cameletti & Schlosser, 2018) can be summarized as follows. On the one hand, SEM was able to estimate the level

[9] ESS samples are "representative of all persons aged 15 and over (no upper age limit) resident within private households in each country, regardless of their nationality, citizenship or language" (source: http://www.europeansocialsurvey.org/methodology/ess_methodology/sampling.html; accessed Sept. 25th, 2018).
[10] There are also other studies focused on a wider application of SEM, currently under development; e.g., see: http://aimed-mi3.com/abstract/categories-of-well-being-modeled-through-structural-equation-modeling/.

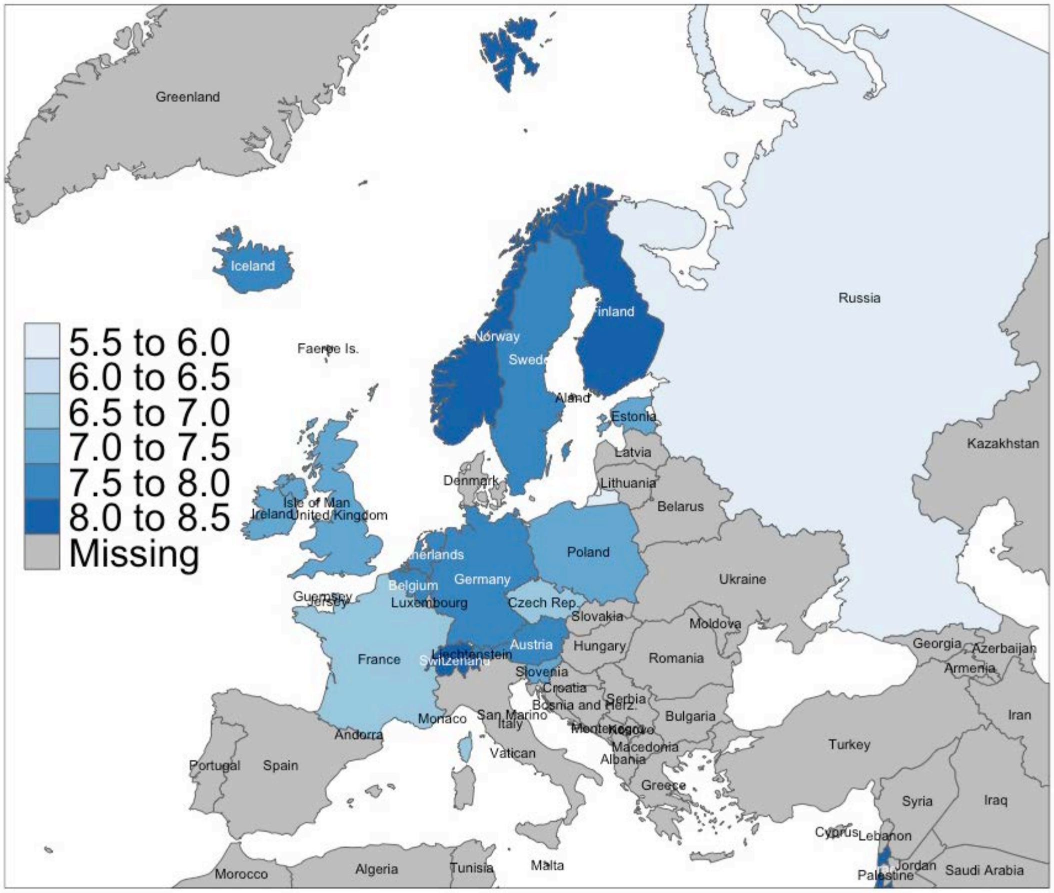

Figure 1: WB in 18 European countries (source: ESS, wave 8; variable: *H-SAvg*, average of items C1 and B27)

of WB in the different countries: both the Pearson correlation coefficient between estimated and observed WB scores (ρ=0.859; p<.001) and the Spearman rank correlation coefficient (ρ=0.835; p<.01) were high and significantly different from zero. Thus, both the estimated scores and the corresponding ranks fitted well the actual observed SWB level.

On the other hand, the estimated structural equation model showed a poor fit to data, according to the chi-square test ($\chi^2(163) = 16.429$, p<.001). Despite this is likely to happen with very large samples (Kline, 2011, p. 201 and p. 209), the Comparative Fit Index (CFI=.764) also showed a not satisficing improvement in the model fit over a baseline independence model[11]. The lack of fit is also highlighted by the Root Mean Square Error of Approximation (RMSEA =.061): this index should better be smaller than .10 (or 0.05 according to Browne and Cudeck (1993). All indexes suggested a need for further exploration in order to detect and study the lack of fit problems of the full model.

Nevertheless, there are also some reasons that can justify this lack of fit between the model and data. Firstly, we used a global European model in order to estimate the SWB level of specific countries, where the relative importance of the different dimensions could be quite different; thus, the global model could not be able to represent the actual relative importance of different WB dimensions in a specific country. Secondly, in estimating the

[11] The fit of the model is usually considered good when the index is higher than 0.90.

global structural equation model we did not apply weights, thus estimates are probably driven by the most populated countries of the group.

These preliminary results showed that a deeper study based on models estimated at the country level could probably attenuate the fitting issues. The research presented in this paper starts from this point.

3. Data & Method

For this research, we used the data collected in 2016 within the 8th wave of the European Social Survey (ESS). The ESS is a broad survey involving representative samples of several European countries. A total of 18 European countries were involved in this wave[12]. In particular, the results of this work are based on the analysis of 34,836 records. Nevertheless, the model estimates rely on fewer records. For the global model estimation, for example, we used only 26,455 units (i.e., 75.9% of the total number of respondents). This drop is explained by the listwise deletion method used by the Lavaan `R` package: it caused the exclusion of records with partial nonresponse, mainly observed for sensitive items such as the ones measuring the household income.

The ESS questionnaire includes several variables that can be linked to different SWB dimensions. Table 1 shows the group of items that we analyzed in order to cover each dimension: in the first column one can find the WB dimensions proposed by Stiglitz, Sen & Fitoussi (2010, p. 14-15), defined as *SSF dimensions*. In the following column we list the name of the dimensions we were able to detect using the ESS questionnaire (*ESS dimensions*). The name of the latent variables as appeared in the estimated structural equation models (*SEM variable names*; note that sometimes two variables were furtherly merged into just one) is shown in the third column. The last column of Table 1 shows the number of ESS questionnaire items (and batteries) used in order to "cover" all the ESS dimensions.[13] Note that we decided to add further dimensions to the ones proposed by Stiglitz, Sen & Fitoussi (2010), that is *Country attachment & people trust*, *Public involvement*, *Discrimination & citizenship* and *Religion* (evaluating the level of religiousness of respondents). Nevertheless, during the stage of model estimation we decided to remove the *Religion* dimension because the original four variables showed a very high correlation, causing estimation problems (it was not even possible to further reduce the correlation combining them together, but in one variable only). However, we wanted to avoid the use of just one variable directly measuring a WB dimension (that is without estimating a latent construct based on more than one observed variable).

All preliminary analyses and SEM estimation were implemented using the `R lavaan` package developed by Rosseel (2012[14]). Before the final model estimation, it was necessary to perform some preliminary steps, in order to fix issues related to ill-scaled matrices, excess of correlation between variables and other problems of model convergence.

---

[12] Austria, Belgium, Switzerland, Czech Rep., Germany, Estonia, Finland, France, United Kingdom, Ireland, Iceland, Israel, Netherlands, Norway, Poland, Russia, Sweden, Slovenia. Data about Italy were published in late spring of 2018, thus not on time for being included in this work.

[13] Thus our approach is different from the one suggested by Diener (1984) and by Andrews & Withey (1976), that studied three components of SWB: life satisfaction, positive experiences and negative experiences.

[14] For further information, see: http://lavaan.ugent.be/.

Table 1: WB dimensions proposed by Stiglitz, Sen & Fitoussi (*SSF*) and covered by selected ESS questionnaire items (*ESS*); number of items and batteries.

SSF dimensions	***ESS dimensions***	***SEM variable name → new merged var.***	***No. of items (batt.=battery)***
Social connection	*Social involvement*	*socinv*	3
Insecurity	*Feeling safe*	*safe → hlthsafe*	2
Health	*Heath conditions*	*health → hlthsafe*	2
---	*Country attachment & people trust*	*atchtrst*	4+2 batt.
Political voice	*Public involvement*	*atchtrst*	2+1 batt.
---	*Religion*	---	4
---	*Discrimination / citizenship*	*discr*	4
Environment	*Worries about the environment*	*env*	2
Material living standards	*Household income perception*	*hinc*	3
Personal activities / work	*Work status*	*job → jobedu*	1+1 batt.
Education	*Education level*	*educ → jobedu*	1

Our analysis is structured in two different phases. The first step is the ideal prosecution of the work presented in Toninelli, Cameletti, Schlosser (2018). By estimating structural equation models by country, we want to detect if the relative importance of the different SWB dimensions is varying across the 18 studied countries. For such a purpose, we will compare the standardized estimated coefficients in terms of both their significance and magnitude.

In the second phase, we will evaluate some goodness of fit measures, in order to check if the local models are able to perform better than the global model estimated with all available data (that is at the European level). In particular, for this second phase we will take into account both model fit indexes (such as the likelihood ratio chi-square, also called model chi-square) and approximated fit indexes (such as the RMSEA, the CFI and the SRMR). The use of model fit indexes enables to check if estimated model covariance is consistent with the observed covariance, whereas approximated fit indexes are measure of model-data correspondence, measuring (directly or indirectly) the goodness of fit of the estimated models. Kline (2011)[15] strongly suggested this integrated study.

On the other hand, we will compare the rankings obtained using the country-specific estimates (based on the SEM local models) against the rankings built on our benchmark item. Following what suggested by Dolan & Metcalfe (2012, p. 420[16]), the benchmark

[15] For an overview about the main characteristics and drawbacks of these indexes, see Kline (2011, pp. 189-214).
[16] It was not possible to completely follow the Dolan & Metcalfe (2012) suggestion, because the ESS questionnaire includes just two out of the four items proposed by these authors.

variable, named *H-SAvg*, is computed as the average of two items that measure directly the SWB, that is the variable *Happy*[17] and the variable *Satisfaction*[18].

4. Results

In this section, we analyze the impact of the different dimensions on the latent WB variable across the various European countries. Then we discuss the structural equation models' goodness of fit, comparing the indexes referred to the global model to the ones obtained with the local models. In the last part of the section, we check if local models can enhance the estimate of WB, providing additional estimates about the relative levels of WB observed in the considered European countries.

4.1 Relative importance of SWB dimensions

Table 2 shows how the importance of different SWB dimensions vary across the 18 studied European countries. It provides the standardized coefficients (and their significance) for the different latent dimensions we took into account. These coefficients were estimated applying the structural equation models at the local level (that is performing a country by country analysis). The last three rows of Table 2 include, respectively, the average and standard deviation of national standardized coefficients, and the coefficients (with their significance) obtained for the global model (i.e., working on all available data, that is at the European level).

First of all Table 2 highlights two dimensions that in some countries are nonsignificant, i.e. the *Discrimination* (*discr*, i.e. how much a respondent is potentially or feels actually discriminated) and the *Environment* (*env*, i.e. how much a respondent worries about the environment). For example, being worried about the environment do not seem to affect the perceived SWB in Austria, Czech Republic, Iceland, Israel and Slovenia, whereas in other countries this latent dimension reaches higher significance values (the highest observed in Belgium and Finland). *Discrimination* and *Environment* are also the only two dimensions negatively affecting the SWB (the average value of standardized parameters are negative, respectively equal to -0.08 and -0.13); moreover, the coefficients for the global models are also negative and equal to -0.13 and -0.17, respectively.

All the other dimensions are significant both for local models and for the global model. The three dimension that mostly affect the SWB are, in decreasing order, *Health & safe* (*hlthsafe*; 0.82 the average of local estimates and 0.89 the coefficient for the global model), the *Household income* (*hinc*) perception (average = 0.81; global model = 0.87) and the *Education level and work status* (average = 0.79; global model = 0.81).

However, the relative importance of these three dimensions varies a lot according to the considered countries. For example *Health and safe* is extremely important for countries such as Russia (estimated standardized coefficient = 1.00), France, Poland, Netherlands (0.97), Czech Republic and Sweden (0.97). Nevertheless, in other countries (Belgium, UK

[17] This variable corresponds to the ESS question #C1 ("*Taking all things together, how happy would you say you are*?"); the response scale includes discrete values within a 0 to 10 range, with extremes labelled as "extremely unhappy" and "extremely happy".
[18] It corresponds to the ESS question #B27 ("*All things considered, how satisfied are you with your life as a whole nowadays?*"); response scale ranging from 0 to 10, with extremes labelled as "Extremely dissatisfied" and "Extremely satisfied".

Table 2: Standardized coefficients for WB dimensions and significance (structural equation models by country and global model).

Model (country)	***socinv***	***atchtrst***	***hinc***	***discr***	***env***	***jobedu***	***hlthsafe***
	Latent dimensions						
Austria	0.39 ****	0.67 ****	0.75 ****	-0.03	0.10	0.64 ****	0.82 ****
Belgium	0.58 ****	0.55 ****	0.74 ****	-0.13 **	-0.33 ****	0.88 ****	0.50 ****
Switzer-land	0.55 ****	0.58 ****	0.72 ****	-0.17 ***	-0.17 *	0.77 ****	0.83 ****
Czech Rep.	0.38 ****	0.33 ****	0.91 ****	-0.12 ***	0.02	0.77 ****	0.96 ****
Germany	0.49 ****	0.68 ****	0.80 ****	-0.07 *	-0.13 ***	0.64 ****	0.72 ****
Estonia	0.57 ****	0.64 ****	0.93 ****	-0.3 ****	-0.09 **	0.95 ****	0.88 ****
Finland	0.35 ****	0.50 ****	0.82 ****	-0.09 *	-0.30 ****	0.95 ****	0.86 ****
France	0.46 ****	0.31 ****	0.64 ****	0.04	-0.28 ****	0.65 ****	0.97 ****
UK	0.38 ****	0.51 ****	0.83 ****	-0.02	-0.24 ****	0.87 ****	0.54 ****
Ireland	0.43 ****	0.58 ****	0.88 ****	0.05	-0.19 **	0.80 ****	0.52 ****
Iceland	0.41 ****	0.57 ****	0.90 ****	-0.3 ****	-0.01	0.57 ****	0.75 ****
Israel	0.40 ****	0.63 ****	0.85 ****	-0.06	-0.04	0.80 ****	0.65 ****
Nether-lands	0.56 ****	0.59 ****	0.75 ****	-0.01	-0.18 **	0.95 ****	0.97 ****
Norway	0.35 ****	0.51 ****	0.74 ****	-0.14 **	-0.12 *	0.86 ****	0.94 ****
Poland	0.43 ****	0.57 ****	0.78 ****	0.07	-0.25 ****	0.56 ****	0.97 ****
Russia	0.19 ****	0.38 ****	0.89 ****	-0.06	0.16 *	0.73 ****	1.00 ****
Sweden	0.44 ****	0.47 ****	0.75 ****	-0.03	-0.19 ****	0.95 ****	0.96 ****
Slovenia	0.41 ****	0.33 ****	0.86 ****	-0.04	-0.05	0.93 ****	0.89 ****
Average *(n. w.)*	**0.43**	**0.52**	**0.81**	**-0.08**	**-0.13**	**0.79**	**0.82**
Std. dev.	**0.093**	**0.113**	**0.077**	**0.101**	**0.133**	**0.132**	**0.163**
Global model	**0.54 ******	**0.63 ******	**0.87 ******	**-0.13 ******	**-0.17 ******	**0.81 ******	**0.89 ******

*Significance: **** = p < .0001; *** = p < .001; ** = p < .01; * = p < .05.*
Note: cells in grey highlight nonsignificant coefficients; n.w. = not weighted.

and Ireland), estimated coefficients are between 0.50 and 0.54. In the latter two countries the relative importance of *Household income* and of *Education level and work status* is noticeably higher, reaching levels higher than 0.80. The *Health and safe* standardized coefficients (average = 0.82) are also the ones that show the highest variability across countries (std. dev. = 0.163). Also *Household income* and *Education level and work status* show a quite big variability across countries: the first dimension ranges from a minimum level of 0.64 in France to a maximum level of 0.93 in Estonia, whereas the second registers the higher value (0.95) in Estonia, Finland, Netherlands and Sweden and the smallest effect (0.56) in Poland. The smallest variability (std. dev. = 0.077) corresponds to the dimension Household income (*hinc*): this highlights how much this dimension is highly important for almost all the considered countries.

The two dimensions that seem to affect less the perceived SWB are *atchtrst* (including the *Country attachment and people trust*, but also the *Public involvement*), with an average of 0.52, and the *Social involvement* (*socinv*, with an average of 0.43). The first dimension (*atchtrst*) shows a quite low level in Russia (0.19) and the highest levels in Belgium (0.58) and Estonia (0.57). The second one (*socinv*) seems to affect the SWB more in Germany (0.68) and Austria (0.67) than in Slovenia, Czech Republic (0.33) and in France (0.31).

Generally speaking, Table 2 demonstrates that the SWB is a phenomenon very well linked with a local background: the relative importance of different dimensions varies a lot across countries. Just to cite an example, *jobedu* is the most influencing variable for Belgium, but in Germany its relative weight is lower than the ones associated to *hinc*, *hlthsafe* and *atchtrst*. Also *hlthsafe*, generally the most influencing dimension, is the most important one in most of the countries (e.g. in Austria, Switzerland, Czech Republic), but is less important than *jobedu*, *hinc*, *socinv* and *atchtrst* in Belgium and less influencing the SWB than *jobedu* and *hinc* in UK.

All these results seem to suggest that a study of WB should not be generalized. Global European models probably would show problems in terms of goodness of fit, because they are an attempt of “averaging” the different effects of WB dimensions over all considered geographical contexts. This probably also explains the low goodness of fit obtained through the global European mode presented in Toninelli, Cameletti, Schlosser (2018). A more local approach (that is studying and estimating models at the country level, or even at some higher geographic level as for example NUTS in UK) should provide with a more precise and useful picture of how the WB is affected by its latent dimensions. Anyway, how much do we actually gain in terms of goodness of fit working at the local rather than at the global (European) level?

4.2 Goodness of fit analysis

In order to evaluate the potential improvement in the models’ goodness of fit, shifting from a global to a local perspective, we follow the method suggested by Kline (2011, p. 193 and following).

Thus, we analyse first the model fit. For each country the results about the model test statistics confirm the findings obtained for the global model. The model chi-square test confirms the null hypothesis rejection: chi-square values show a range from a minimum equal to 749.4 (for Israel) to a maximum equal to 2,087 (for Germany). Degrees of freedom range from 163 to 165 and all p-values are lower than 0.001. Thus, we have always to reject the null hypothesis of exact fit (i.e., there are no discrepancies between population covariances and covariances predicted by the models). These results are anyway reasonable, because model fit test for samples bigger than 5,000 can likely show that the models fail even if the differences between observed and predicted covariances are acceptably small. In this case, a further study of the approximate fit is needed.

Table 3 shows approximate fit test results. We take into consideration, following the suggestion of Kline (2011), three approximate fit statistics: the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI) and the Standardized Root Mean Square Residual (SRMR). RMSEA is a parsimonious adjusted index (i.e. corrected for model complexity). Its values for each model are shown together with the 90% confidence interval bounds and with the *p*-value testing the close-fit null hypothesis. The best fit is usually detected when RMSEA = 0 (thus the index can be considered a direct measure of badness of fit). For all our models we observe very small values, nevertheless, the usual threshold for accepting the close-fit hypothesis is having RMSEA values lower or equal to 0.05. No one of the RMSEA values is smaller than this threshold. Moreover, the *p*-values confirms that in all the cases we reject the null hypotheses of close-fit. The first grey column of Table 3 shows the percentage change observed in the RMSEA shifting from the global model to each specific local model. We would expect a reduction of the index, whereas we actually observe (if we exclude the case of Russia, -2,2%) quite big increase, up to +30.3% of the value observed for the model of Belgium and to the +33% observed for Iceland. The non-weighted average shows an increase of 14.5%. This is against our initial expectations.

CFI provides a partially different view, being part of the group of incremental fit indexes: it measures the improvement in fit of an estimated model compared to a null model (sometimes defined independence model) for which the population covariances among variables are set to 0. The CFI is computed as follows:

$$CFI = 1 - \frac{\chi^2_{\mathrm{M}} - df_{\mathrm{M}}}{\chi^2_{\mathrm{M}} - df_{\mathrm{M}}},$$

where χ^2_{M} and df_{M} are, respectively, the chi-square model fit statistics and its degrees of freedom and χ^2_{B} and df_{B} are the same statistics referred to the baseline model. Thus the CFI varies from 0 to 1 and is a direct measure of goodness of fit (with values of 1 meaning a perfect fit). The percentage change shown in the second grey column of Table 3 shows that the goodness of fit is not at all improved when estimating local models rather than the global model. The Iceland and the Switzerland models show a decrease of more than 20% in the CFI values. The average change is equal to -10.7%: also these results are from our point of view counterintuitive.

Finally, Table 3 shows the values of the Standardized Root Man Square Residual (SRMR), an index based on covariance residuals. This index measures the differences between observed and predicted correlations and, in an ideal situation, should be equal to zero (or close to zero for an acceptable model fit). Thus, using local rather than the global European model, we would expect lower level of the index, whereas Table 3 shows higher value, ranging from +7% for Estonia to +58.3% for Iceland. Generally we observe an average +27.6% change.

Hu & Bentler (1999) suggest that an acceptable fit is detected when the CFI index is $\geq$0.95 and the SRMR is $\leq$.08. Despite some studies findings (e.g. Fan & Sivo, 2005 and Yuan, 2005) are against the use of these thresholds, they are still widely used as a good “rule of thumb”. Considering jointly the two indexes, we can conclude that there is no gain, in terms of goodness of fit, moving from a global model estimated at the European level to local models estimated at country level.

Table 3: Models' approximate fit statistics and % differences vs global model.

	RMSEA					*CFI*		*SRMR*	
Model		***% diff.***	***90% c.i.***		***p-value***		***% diff.***		***% diff.***
GLOBAL MODEL	**0.061**		**0.061**	**0.062**	**0.000**	**0.764**		**0.053**	
Austria	0.073	18.4	0.069	0.076	0.000	0.641	-16.1	0.067	25.5
Belgium	0.080	30.3	0.077	0.083	0.000	0.633	-17.1	0.075	41.2
Switzerl.	0.075	22.7	0.071	0.079	0.000	0.579	-24.3	0.070	31.7
Czech R.	0.066	8.2	0.063	0.070	0.000	0.706	-7.6	0.065	21.8
Germany	0.068	11.5	0.066	0.071	0.000	0.752	-1.6	0.063	17.3
Estonia	0.062	1.1	0.059	0.065	0.000	0.805	5.3	0.057	7.0
Finland	0.069	12.9	0.066	0.073	0.000	0.722	-5.5	0.069	29.8
France	0.068	10.3	0.065	0.071	0.000	0.706	-7.6	0.070	30.8
UK	0.073	19.5	0.070	0.077	0.000	0.677	-11.4	0.072	35.5
Ireland	0.074	20.9	0.071	0.077	0.000	0.639	-16.4	0.069	29.9
Iceland	0.082	33.0	0.078	0.085	0.000	0.595	-22.1	0.084	58.3
Israel	0.069	11.6	0.064	0.074	0.000	0.717	-6.2	0.071	33.1
Netherl.	0.067	9.5	0.064	0.071	0.000	0.776	1.5	0.063	18.0
Norway	0.067	9.9	0.064	0.071	0.000	0.668	-12.5	0.066	23.6
Poland	0.069	11.9	0.065	0.073	0.000	0.627	-18.0	0.066	23.3
Russia	0.060	-2.2	0.056	0.064	0.000	0.664	-13.1	0.060	12.0
Sweden	0.072	18.0	0.069	0.076	0.000	0.680	-11.0	0.069	28.7
Slovenia	0.070	13.8	0.066	0.074	0.000	0.690	-9.7	0.069	29.7
AVG	**0.070**	**14.5**				**0.682**	**-10.7**	**0.068**	**27.6**

4.3 Effectiveness of local model estimates

As suggested by Kline (2011), both model fit and approximate fit indexes (presented in Section 4.2) are subject to specification error (that can cause an excess of correlation between residuals). Moreover, they have also issues linked to very big sample sizes and to lack of distributional assumptions (non-normal distribution of the data or of the residuals, non-centrality in the chi-square distribution of some test statistics). Kline (2011) also suggest to empirically check the capability of the model to provide the researcher with reliable estimate, despite the failure of the main goodness of fit tests. For this reason we want to study how much local models are able to reproduce the relative level of WB. We check this by comparing local WB estimates with the two ESS benchmark questions (summarized in the variable *H-SAvg* introduced in last row of sect. 3).

Table 4 allows us to compare the scores of the *H-SAvg* and the WB scores obtained as estimates using the local structural equation models.

Table 4: Ranks of the 18 European countries according to *H-SAvg* (in increasing order) and to the local models' WB estimates

Rank H-SAvg	***Country***	***WB scores H-SAvg***	***WB estimates Local models***	***Rank Local models***
1	Russian F.	5.949	0.060	17
2	Czech Rep.	6.803	0.024	7
3	France	6.887	0.041	13
4	Estonia	7.172	0.054	16
5	Slovenia	7.299	0.031	11
6	Poland	7.401	0.072	18
7	Ireland	7.404	0.026	8
8	UK	7.555	0.042	14
9	Belgium	7.614	0.024	6
10	Germany	7.644	0.027	9
11	Austria	7.743	0.032	12
12	Iceland	7.827	0.017	2
13	Netherlands	7.876	0.027	10
14	Sweden	7.902	0.019	3
15	Norway	8.043	0.021	5
16	Israel	8.107	0.008	1
17	Finland	8.128	0.053	15
18	Switzerland	8.165	0.020	4

It can be immediately seen from the table that the two ranking (one created using the WB estimates directly measured by the ESS questionnaire and the other one obtained by means of the local models) are very different. This is confirmed by the Pearson coefficient correlation confirms which is negative and quite high (=-.507; p=.032). Also considering the rank correlation coefficients, we obtain similar results: the Kendall's tau is equal to -.401 (p=.021) and the Spearman's rho is equal to -.553 (p=.017).

5. Conclusions

In the first phase of our research, we mainly aimed at understanding if SEM could be a reliable method in order to measure the WB. We base our work on ESS data, a broad survey developed in several European countries and involving a very large sample (representative of the population of 15 years and older). The advantage of using such dataset is that there are two questions directly asking to respondents to self-report the perceived WB: these questions can be considered as benchmarks for our study. SEM seemed to be the perfect tool, in order to estimate the WB, given the latent nature of this concept. Thus, we started from several ESS variables, grouped into some latent WB dimensions, with the objective of covering more than the usual dominions usually taken into account in order to study the WB. We estimated, first, a global European model. Despite the lack of fit with the empirical data, the model seemed to be able to estimate well the relative level of WB observed in the 18 countries involved in this study.

Further research focused on standardized SEM coefficients showed that the relative importance of the different WB dimensions was quite different across the studied European countries. This pushed us to change our perspective. Probably, moving from a global model to local models (i.e., model estimated country by country) would have provided us with estimates that could reflect more closely the actual relative WB levels observed in different contexts.

This, unexpectedly, did not happen. Both model fit and approximate fit indexes highlighted a very poor goodness of fit of SEM models (that can be in part justified by the very big size of our sample). Moreover, estimates obtained using local models were not able at all to reproduce the relative WB level observed in the European countries and measured through the ESS questionnaire.

There could be different reasons behind these unexpected results. A more careful study of the violation of distributional assumptions can give a more precise idea about the reliability of the indexes we used to check the lack of fit and the discrepancies between data and estimated models. Moreover, a further deeper study of residuals (in terms of potential residual patterns and of correlation between them) could suggest to differently specify our estimated models. In this first phase, we did not want to estimate specific models for different countries, for two main reasons: first, to allow a direct comparability across countries, second, in order to be able to cover and study the importance of all the WB dimensions using always the same set of variables. A further more specific variable selection that follows the suggestions of a preliminary correlation study and of the residual features could provide us with more reliable (despite “personalized”) structural models. Another potential limit for this study was the big number of missing that caused the loss of about 25% of respondents. A more detailed study of this missing pattern could bring to a further restriction of the variables specified within the local models on a country-by-country basis.

In the future, this research can also focus on a longitudinal perspective, detecting if data of other ESS waves confirm our findings. Should a further and more detailed study, eventually focused also on other waves, confirm results shown in this paper, we could even opt to start experimenting other statistical tools in order to estimate the WB.

The final goal of the whole process would still be to find a way to merge information collected through big surveys (such as the ESS) with data collected automatically (on a systematic basis) from the web and from the social networks (more precisely from Twitter). The final aim is to estimate timely and with limited costs the behaviour of WB in a complex context like the one made by different European countries, where the determinants of the WB are so different in affecting such a relevant and strategic latent variable and its evolutions over time.

Acknowledgements

This research is supported by funding of the Department of Business, Economics and Quantitative Methods of the University of Bergamo in the framework of the project "Strumenti e Metodi Statistici per la Rilevazione e la Modellizzazione di Fenomeni Sociali Complessi".

References

Andrews, F.M., Withey, S.B. (1976), "Social indicators of well-being: American's perceptions of life quality". Plenum Press, New York.

Browne, M.W., Cudeck, R. (1993), "Alternative ways of assessing model fit". In Bollen, K.A., Long, J.S (Eds.), *Testing structural equation models*, 136-162. Sage, Newbury Park, CA.

Diener, E. (1984), "Subjective well-being". *Psychological Bulletin* 95, 542-575.

Diener, E. (2013), "The remarkable changes in the science of subjective well-being". *Perspectives on Psychological Science*, 8, 663-666

Dolan, P., Metcalfe, R. (2012), "Measuring subjective wellbeing: recommendations on measures for use by national governments". *Journal of social policy* 41(2), 409-427.

Fan, X., Sivo, S.A. (2005), "Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of the two-index strategy revisited". *Structural Equation Modelling* 12, 343-367.

German Federal Government (2017), "Government Report on Wellbeing in Germany". Federal Press Office, Berlin. Link: https://buergerdialog.gut-leben-in-deutschland.de/SharedDocs/Downloads/EN/LB/Government-Report-on-Wellbeing-in-Germany.pdf?__blob=publicationFile

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

Kline, R.B. (2011), "Principles and Practice of Structural Equation Modeling". The Guildford Press, New York.

Lin, C.C., Yeh, Y.C. (2014), "How gratitude influences well-being: a structural equation modeling approach". *Soc Indic Res* 118(1), 205-217.

OECD (2013a), "Guidelines on Measuring Subjective Well-being". Link: http://www.oecd.org/statistics/Guidelines%20on%20Measuring%20Subjective%20Well-being.pdf; retrieved on Sept. 8th, 2018.

OECD (2013b), "Concept and validity", in *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing. DOI: http://dx.doi.org/10.1787/9789264191655-5-en

Oliver, A., Navarro, E., Meléndez, J.C., Molina, C, Tomás, J.M. (2009), "Structural equation model for predicting well-being and functional dependency of the elderly in the Dominican Republic". *Rev Panam Salud Publica* 26(3), 189-196.

Rosseel, Y. (2012) "lavaan: An R Package for Structural Equation Modeling". *Journal of Statistical Software* 48(2), 1-36.

Stiglitz, J.E., Sen, A., Fitoussi, J.-P. (2010), Report by the Commission on the Measurement of Economic Performance and Social Progress. http://hdl.voced.edu.au/10707/47743. Cited 21 Feb 2018.

Toninelli, D., Cameletti, M., Schlosser, S. (2018 forthcoming), "Is Structural Equation Modelling Able to Predict Well-being?", in SIS 2018 conference proceedings, Abbruzzo, A., Brentari, E., Chiodi, M., Piacentino, D. (Eds), Pearson.

Turashvili, T., Turashvili, M. (2015), “Structural equation model of psychological well-being, a Georgian Exploration”. *Procedia - Social and Behavioral Sciences* 190, 4-9.

Warner, R.M., Rasco, D. (2014), “Structural equation models for prediction of subjective well-being: Modeling negative affect as a separate outcome”. *The Journal of Happiness & Well-Being*, 2(1), 161-176.

Yuan, K.-H. (2005). “Fit indices versus test statistics”. *Multivariate Behavioral Research* 40, 115-148.

Zaidi, A., Green, M., Iparraguirre, J., Davidson, S., Rossall, P. (2017), “Constructing a metric of well-being among older people in the UK”. CASE Social Exclusion Seminar, London School of Economics.