EXTENDED ABSTRACT

Title: The vulnerability of European regions in the face of the post-2020 Cohesion Policy: new methodological approaches for analysis.

Authors and e-mail of all:
Eduardo Jiménez Fernández[1] (jimeneze@uji.es); Angeles Sánchez[2] (sancheza@ugr.es)

Department and University:
(1) Departamento de Economía. Universitat Jaume I, Castellón.
(2) Departamento de Economía Aplicada. Universidad de Granada, Granada.

Subject area: 15.- Gobernanza local y regional, políticas regionales de desarrollo y cohesión

Keywords: European Union; Cohesion Policy; Regional development; Composite index; Fuzzy Metric; MARS.

JEL codes: C43, I31, O15, R12, R58.

1. Justification and aims

The European Union (EU) Cohesion Policy for the next period (2021-2027) aims at fostering a modernised regional development and cohesion policy, by investing in all regions with a tailored approach, and in a more flexible framework. More specifically, the EU Cohesion Policy focuses on five political goals so that EU becomes: (1) smarter, through innovation, digitisation, economic transformation and support to small and medium-sized businesses; (2) greener, fostering carbon free Europe, implementing the Paris Agreement and investing in energy transition, renewable and the fight against climate change; (3) more connected, with strategic transport and digital networks; (4) more social, delivering on the European Pillar of Social Rights and supporting quality employment, education, skills, social inclusion and equal access to healthcare; and (5) closer to citizens, by supporting locally-led development strategies and sustainable urban development across the EU (European Commission 2018c).

Those guidelines represent big challenges for the design of the regional development policies within the scope of "beyond GDP", according to which the

European Commission should develop several indicators that complement the Gross Domestic Product (GDP) to support policy decisions by more comprehensive information (Commission of the European Communities 2009a).

However, and in spite of the push Europe 2020 represents to measure the societal progress as the improvement in various areas of citizens' well-being (see for instance Sánchez & Ruiz-Martos 2018), a macroeconomic index (the gross national income) is again proposed as the predominant criterion for allocating structural funds in 2021-2027. In this vein, EU regions are classified into three groups. The group 1 is integrated by those with a gross national income ratio equal to or above 100% of the EU average; the group 2 by the regions with a gross national income ratio equal to or above 75% and below 100% of the EU average; and the group 3 by the regions with a gross national income ratio below 75% of the EU average (European Commission 2018c). That is, despite economic and social cohesion are core EU objectives since its foundation, the community regional performance is defined in an economic sense by the size and growth of the economy.

Given that the main objective of the EU Regional Policy –or Cohesion Policy- is to reduce disparities between the levels of development of the various regions and the backwardness of the least favoured regions, in this paper we hypothesize that there is the possibility of taking into account new complementary criteria in line with the five objectives outlined above, in order to better reflecting the reality on the ground of the regions. With this in mind, the main aim of this paper is to construct a composite index of Socio Economic Vulnerability that synthesises the position of each EU region in 2017 with respect the five objectives of the EU Cohesion Policy for the next period (2021-2027). In this context, vulnerability can be understood as greater obstacles or in a worse position to achieve these objectives. More specifically, we build a composite index (the Socio Economic Vulnerability Index or SEVI) for the European NUTS 2 of 27 Member States[1] which allows us to empirically analyse the situation of the EU regions with respect the new cohesion policy aims and therefore to discuss how the structural funds should be allocated. In this way, our proposal fits into the mainstream of economists and

[1] We exclude the regions of the United Kingdom, as the European Commission has done in the preparation and proposals for regulations and budgets for the regional community policy of the 2021-2027 period.

policymakers who argue that associating the notion of economic and social progress to a one-dimensional variable of economic activity, such as GDP or income seems debatable (see for instance O'Donnell et al 2014; Stiglitz et al. 2011; Van den Bergh 2009).

2. Methods

2.1. The measurement model

The first step in the construction of a composite index focuses on the study of a theoretical framework or underlying multi-dimensional model that provides the basis for the selection and aggregation of single indicators (Maggino, 2017a). In turn, the choice of mathematical methods for aggregating the indicators into a composite index will depend on the kind of measurement model that best fits the phenomenon being analysed (Maggino, 2017b; Mazziotta & Pareto, 2019).

Our proposal is to measure the socio economic vulnerability of each EU region in terms of the five dimensions or objectives of the EU Cohesion Policy for the next period (2021-2027), namely to promote: (1) an innovative and smart economic transformation, (2) a greener development, (3) the regional ICT (Information Communication Technologies) connectivity, (4) a more social Europe and (5) a Europe closer to citizens. Thus, we select several individual indicators from these domains, then we synthesize them in a single index to inform us of the position or distance of each of the EU regions with respect to those objectives. Specifically, our objective is to measure how far the regions are from those objectives in 2017, so that the further they are (the bigger the SAVI is), we consider that the region is more vulnerable from the socio-economic point of view and, therefore, should attract more attention and budgetary effort of the EU Cohesion Policy for the next period (2021-2027).

Given the concept of our latent variable (regional vulnerability), as well as the relationships between the latent variable and its measures (single indicators), we develop our model under the scope of a formative model. In the formative measurement model, causality flows from the indicator to the latent variable. That is, single indicators are viewed as causes of the latent variable. For instance, in our case, the SEVI of a region includes indicators of education, unemployment, pollution and material well-being. Any change in one or more of these components (even if the other factors did not change) is

likely to cause a change in a region's SEVI score (the latent construct). However, if a region's SEVI decreased, this would not necessarily be accompanied by an improvement in all of the components (indicators).

In the specification of a formative model, individual indicators (z_i) are explanatory variables and the latent variable is the dependent variable:

$$SEVI_i = f(z_i), \qquad \text{being i the region.}$$

A fundamental characteristic of formative models is that the correlation between individual indicators (z_i) may be positive, negative or zero despite capturing the same concept (Maggino 2017a). Correlated indicators are permitted when the mathematical method chosen to construct the composite index allows avoiding the duplicity of information.

2.2. Fuzzy metrics to build the composite index of regional vulnerability

The procedure presented in this study is based on building a composite index of regional vulnerability using a conjunction of fuzzy metrics related to each individual indicator through a iterative process. Fuzzy metrics are a mathematical tool that allow us to measure or quantify a concept (vulnerability) that in principle is not measurable or quantifiable. Multivariate Adaptive Regression Splines (MARS) models are used to approximate the function relation between the output -the product of all fuzzy metrics considered- and the individual indicators. Through of the data set, this non-parametric method extends the linear model by incorporating non-linearities and variable interactions which is implemented by constructing a suitable basis function (polynomials of degree q) according to forward/backward stepwise algorithm. Thus the final model is constituted as a combination of this generated base functions that can be fitted by ordinary least-squares (Friedman 1991). Generalized cross-validation (GCV) was used as a technique for evaluating the previous model. The GCV is implemented by constructing partitions of the data, namely, subsets of input data which are training in the mode, and the complementary subset of the data which are matching in order adjust the coefficients values to best fit the data and also to select the optimal disjoint basis functions (Craven & Wahba 1971). Once we know which is the best approach to the data, we determine the importance or weight of the individual indicators in the chosen model. In our study, we

select the weights of indicators through a function variable importance, using Partial Dependence Plots (PDP) (Greenwell et al. 2018). This score is used as a weight in the fuzzy metric with respect each individual indicator j. Each one of these weights generates in the metric a specific importance for each indicator. Finally, the algorithm provide a composite indicator that collects this characteristics for each observation.

Those techniques allow us to overcome the most troubling issues of composite indices (namely, the normalization, the allocation of weights and the treatment of redundancy of information) in a robust way.

2.3. Data and single indicators

To develop the Socio Economic Vulnerability Index (SEVI) in the EU regions, we use the official statistics of EUROSTAT and OECD at level NUTS 2 which is the basic regions for the application of regional policies (see Appendix A). We work with the newest regional territorial classification, known as NUTS 2016, which entered into force on 1 January 2018, in accordance with the Commission Regulation (EU) 2016/2066. The overseas NUTS 2 territories have not been taken into account in this study (Ceuta and Melilla in Spain; and Guadeloupe, Martinique, Guyane, La Réunion and Mayotte in France). The final number of EU regions or NUTS2 territories studied is 233.

The selection of indicators has been essentially guided by the five objectives set by the European Commission for Cohesion Policy for the next period (2021-2027). The indicators were also chosen carefully to remove the effect size and meet the following technical criteria (Advisory Committee on Official Statistics 2009; Maggino 2017b): relevance, statistically sound, intelligible and easily interpreted, reliability, and allow international comparison. In any case, our selection has been determined by the availability of statistical information, which is quite scarce at NUTS 2 level in several areas such as those related to climate change and self-reported measures.

Table 1 shows the descriptive statistics of the individual indicators for the 233 EU regions over the period 2016-2017. The values of Pearson's coefficient of variation indicate that the largest territorial differences arose in the objective of fostering the innovative and smart economic transformation (indicators 1 and 2) as well as in the Male unemployment rate. The last column informs about the polarity or the relation between

an indicator and the concept to be measured, in this case the socio economic vulnerability. Positive polarity means that an increase in the indicator could also lead to an increase of the vulnerability. Negative polarity means that an increase in the indicator could lead to a reduction of the vulnerability. Since the composite index proposed is a distance index, the baseline

Table 1
Descriptive statistics of socioeconomic vulnerability for the EU27 regions in 2016-2017 (n = 233)

Indicator (name)	Mean	SD	Minimum	Maximum	CV	Polarity
1. R&D expenditure business enterprise sector (rdbussiness)	0.99	0.96	0	8.06	0.97	negative
2. R&D expenditure public sector (rdpublic)	0.61	0.44	0	2.52	0.71	negative
3. Population exposure to fine particles (pm25)	12.89	4.27	4.40	28.28	0.33	positive
4. Internet at home (internet)	97.33	2.69	4.50	92	0.03	negative
5. Online interaction with public authorities (eadmi)	51.42	20.09	4.50	92	0.39	negative
6. Inequality adjusted GDP per capita (gdp_gini)	19,782.30	7,905.24	5,459.60	52,512,45	0.40	negative
7.Youth unemployment rate (yunemp)	20.39	12.89	3.60	57.15	0.63	positive
8. Male unemployment rate (munemp)	7.94	4.80	1.85	39.25	0.79	positive
9. Female unemployment rate (funemp)	9.48	2.11	4.57	15.52	0.22	positive
10. Elderly people (elderly)	9.48	2.11	4.57	15.52	0.22	positive
11. Early leavers from education and training (early)	10.21	4.87	1.35	27.35	0.48	positive
12. Population with terciary education (terciary)	29.12	8.99	11.80	55	0.31	negative

Note. SD: standard deviation; CV: Pearson's coefficient of variation (SD/Mean). Polarity is the relation between the indicator and the regional socio economic vulnerability.

The methodology selected to construct the SEVI of a region considers in its calculation formula the distance at which each of the individual indicators are in relation to a baseline or reference vector. Our SEVI composite index will take higher values (closer to 1) the greater the distance be with respect the best values or the more desirable values of the individual indicators. That is, the greater the SEVI be, the worse the behavior of a region in the different indicators studied. In order to select the reference values, we have taken into account the complete empirical distribution in the 233 EU regions. More specifically, for those individual indicators that have positive polarity, we select the

maximum value of the indicator in the entire sample studied. For individual indicators with negative polarity, the reference value is the minimum value of the sample.

3. Findings and discussion

Figure 1 shows the variable importance and the scores computed in the last iteration represented as a percentage. The final SEVI was computed using the weights corresponding to the previous scores. This metric provides an order that enables represent all regions according to their socio economic vulnerability, where those regions close to 1 are very vulnerable and the regions close to 0 are less vulnerable.

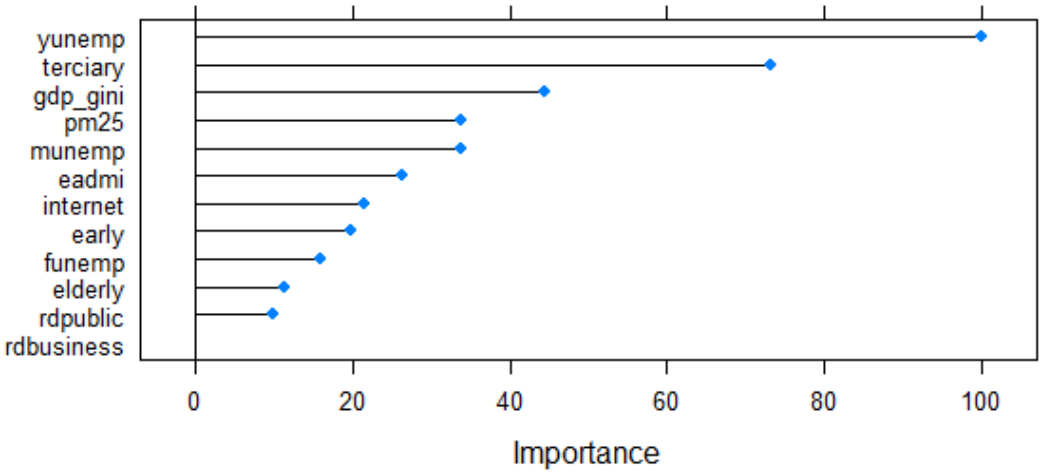

Figure. 1. Variable importance score for MARS model selected.

The results of SEVI for the 233 EU27 regions are shown in Figure 2. According to our results, the regions that are in the most disadvantaged situation to face the challenges of cohesion policy (2021-2027) are Dytiki Makedonia, Ipeiros, Sterea Ellada and Dytiki Ellada in Greece; Sicily, Calabria, Puglia, Molise, Campania and Sardegna in the South of the South of Italy; Severozapaden in Bulgaria; Sud-Vest Oltenia and Sud-Est in Romania; Extremadura and Andalusia in southern Spain. On the contrary, the best positions, and therefore the lowest values in the SEVI, are reached by regions of Denmark, Sweden, the Netherlands, all of Ireland, and by the regions where the capitals

are located, such as Helsinki-Uusimaa in Finland, Praha in Czech Republic, Wien in Austria, Île de France in France, Berlin in Germany.

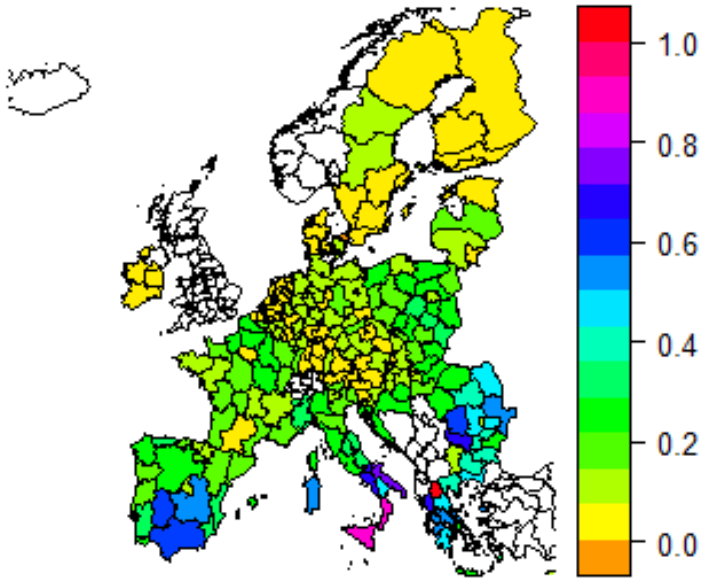

Figure 2. Socio Economic Vulnerability Index (SEVI) in the EU27 regions, 2017 (n=233, NUTS2).

If we compare the results of the SEVI with the allocation mechanism of the Structural Funds (GDP per capita), some remarkable findings can be obtained with implications for Cohesion Policy. More specifically, we could identify the regions that would be harmed in terms of allocation of Structural Founds if the traditional criterion was applied, which consists of taking as reference a single indicator of economic activity (GDP pc) instead of a set of indicators that complement the GDP and accurately reflect the socio-economic situation of the EU regions in order to face the challenges of Cohesion Policy. Table 2 displays the results for the SEVI and the GDP per capita and ranks the regions in these two concepts. The GDP per capita ranking orders the regions from lowest to highest values. Table also indicates how each region would be rated according to the GDP per capita criteria (namely, group 1 equal to or above 100% of the EU average; the group 2 equal to or above 75% and below 100% of the EU average; and the group 3,

regions with a GDP per capita below 75% of the EU average). The last column calculates the difference between the SEVI and GDP per capita rankings which allows us to identify the regions with downward changes and upwards changes.

For example, Lombardia in Italy records a difference in rankings equal to -128. According to its GDP per capita, it would be considered a region of group 1 and, consequently, it would be less benefited in the allocation of Structural Funds, but it occupies in the SEVI ranking position 31 of 233. That is, it is in the quintile of the most vulnerable regions that should be subject to greater budgetary effort under the Cohesion Policy. In this situation there are 16 Italian regions. The opposite situation is when the difference in rankings is positive, which would indicate that the region is less vulnerable than its relative position in GDP per capita reflects. As noted, this situation is recorded in regions of countries in ancient Eastern Europe that traditionally record low levels of GDP, but in comparative terms perform well in another social and economic indicators.

These differences in the map of priority regions could be a source of debate on the introduction of new game rules in terms of community regional policy. Within the framework of "GDP and beyond", the classification of European regions for Cohesion Policy proposes should be carried out taking into account a battery of socio-economic indicators that will provide a more reliable view of the situation of citizens.

Table 2. Comparison of EU27 regions rankings, 2017: downwards and upwards changes

Region, NUTS2	Rank(1)	SEVI	Rank(2)	GDPpc(a)	Group	SEVI-GDPpc(1-2)
ITC3 - Liguria	31	0.0403	159	108.5	1	-128
ITF1 - Abruzzo	35	0.0384	100	84	2	-65
ITC1 - Piemonte	36	0.0380	146	104	1	-110
ITI2 - Umbria	41	0.0358	101	84	2	-60
ITI3 - Marche	42	0.0357	123	92.5	2	-81
ITC2 - Valle d'Aosta/Vallée d'Aoste	46	0.0347	183	120	1	-137
ITC4 - Lombardia	47	0.0338	197	129.5	1	-150
FRF2 - Champagne-Ardenne	49	0.0331	109	87	2	-60
ITI1 - Toscana	57	0.0319	150	105	1	-93
ES53 - Illes Balears	58	0.0317	135	97.5	1	-77
EL30 - Attiki	59	0.0312	124	92.5	1	-65
HU23 - Dél-Dunántúl	69	0.0287	9	44.5	3	60
HU32 - Észak-Alföld	71	0.0283	6	43	3	65
ITI4 - Lazio	75	0.0278	168	112	1	-93
ITH4 - Friuli-Venezia Giulia	78	0.0268	155	106	1	-77
PL84 - Podlaskie	80	0.0265	18	49.5	3	62
ITH3 - Veneto	85	0.0248	170	113	1	-85

ITH1 – Prov. Autonoma di Bolzano/Bozen	90	0.0227	214	146.5	1	-124
ITH5 - Emilia-Romagna	94	0.0222	186	121	1	-92
ES51 - Cataluña	98	0.0218	164	111	1	-66
HU33 - Dél-Alföld	107	0.0201	14	48.5	3	93
ITH2 - Provincia Autonoma di Trento	116	0.0183	189	123	1	-73
RO32 - Bucuresti - Ilfov	118	0.0181	212	142.5	1	-94
PL63 - Pomorskie	121	0.0178	56	67.5	3	65
HU21 - Közép-Dunántúl	126	0.0171	45	63.5	3	81
SI03 - Vzhodna Slovenija	127	0.0170	60	70	3	67
DEA1 - Düsseldorf	136	0.0161	200	130.5	1	-64
PL91 - Warszawski stoleczny	138	0.0160	215	151.5	1	-77
CZ05 - Severovýchod	139	0.0159	72	74.5	3	67
BE10 - Région de Bruxelles-Capitale	141	0.0157	230	200	1	-89
LT02 - Vidurio ir vakaru Lietuvos reg.	145	0.0146	49	64	3	96
BE34 - Prov. Luxembourg (BE)	151	0.0137	73	74.5	3	78
BE35 - Prov. Namur	152	0.0136	89	81	2	63
CZ03 - Jihozápad	158	0.0125	82	78	2	76
HU12 - Pest	162	0.0121	24	54	3	138
CZ02 - Střední Cechy	171	0.0107	94	83	2	77
EE00 - Eesti	174	0.0106	83	78.5	2	91
NL13 - Drenthe	180	0.0103	112	88	2	68
CZ06 - Jihovýchod	186	0.0101	91	82	2	95
NL12 - Friesland (NL)	188	0.0098	115	88.5	2	73
IE04 - Northern and Western	189	0.0096	98	83.5	2	91
FI1D - Pohjois- ja Itä-Suomi	191	0.0096	119	91	2	72
FI19 - Länsi-Suomi	198	0.0087	134	97.5	2	64
SE22 - Sydsverige	211	0.0077	147	105	1	64
NL23 - Flevoland	212	0.0077	127	96.5	2	85
SE12 - Östra Mellansverige	221	0.0065	153	105.5	1	68
SE33 - Övre Norrland	227	0.0058	166	111.5	1	61

Note. SEVI: Socio economic vulnerability index; (a) Gross domestic product (GDP) at current market prices, purchasing power standard (PPS, EU27 from 2019), per inhabitant in percentage of the EU27 (from 2019) average. Adapted by Eurostat, Regional economic accounts (nama_10r_2gdp).

Appendix A. Definitions and sources of the individual indicators

Indicator	Definition	Source	Geographic level	Date of data used
1	Intramural R&D expenditure Business enterprise sector (percentage of gross domestic product)	Eurostat, Statistics on research and development (rd_e_gerdreg)	NUTS1 for some regions in BE, LT, NL, PL. NUTS2 for all the other countries.	Average 2015-2016, except: BE 2014-2015; AT, DE, EL, IE 2015; NL 2014; FR 2013.

2	Intramural R&D expenditure Government sector+Higher education sector (percentage of gross domestic product)	Eurostat, Statistics on research and development (rd_e_gerdreg)	NUTS1 for some regions in BE, LT, NL, PL. NUTS2 for all the other countries.	Average 2015-2016, except: BE 2014-2015; AT, EL, IE, PL, SE 2015; NL 2014; FR 2013.
3	Mean population exposure to PM2.5 Micrograms per cubic metre	OECD, Environment Database - Exposure to PM2.5	NUTS2, own elaboration	Average 2016-2017
4	Percentage of households with internet access at home	Eurostat, ICT usage in households and by individuals (isoc_r_iacc_h)	NUTS1 for some regions in DE, EL, PL. NUTS2 for all the other countries.	Average 2017-2018
5	Percentage of individuals who used the Internet for interaction with public authorities (last 12 months)	Eurostat, ICT usage in households and by individuals (isoc_r_gov_i)	NUTS1 for some regions in DE, EL, PL. NUTS2 for all the other countries.	Average 2017-2018
6	Regional gross domestic product (GDP) purchasing power standard per inhabitant adjusted by the country Gini index of disposable household income [GDP per capita*(1-Gini index)].	Eurostat, Regional economic accounts (nama_10r_2gdp) and Income and living conditions (ilc_di12).	NUTS2	Average 2016-2017
7	Youth unemployment rate % from 15 to 24 years (Female+male)	Eurostat, Regional labour market statistics (lfst_r_lfu3pers)	NUTS1 for some regions in AT, DE, FI, HU, LT, PL, PT. NUTS2 for all the other countries.	Average 2016-2017
8	Male unemployment rate % from 20 to 64 years (male)	Eurostat, Regional labour market statistics (lfst_r_lfu3pers)	NUTS2	Average 2016-2017
9	Female unemployment rate % from 20 to 64 years (female)	Eurostat, Regional labour market statistics (lfst_r_lfu3pers)	NUTS2	Average 2016-2017
10	Percentaje of elederly people in population (75 years or over).	Eurostat, Population change - Demographic balance and crude rates at regional level (demo_r_gind3)	NUTS2	Average 2016-2017
11	Early leavers from education and training denotes the percentage of the population aged 18 to 24 having attained at most lower secondary education and not being involved in further education or training.	Eurostat, Educational attainment level and transition from education to work (based on EU-LFS) (edat_lfse_04)	NUTS1 for some regions in AT, DE, FI, FR, IT, PL. NUTS2 for all the other countries.	Average 2016-2017

12	Individuals aged 25-64 who successfully completed terciary education (levels 5-8 ISCED 2011) over the population with the same age (In %)	Eurostat, Educational attainment level and transition from education to work (based on EU-LFS) (edat_lfse_04)	NUTS2	Average 2016-2017

References

Advisory Committee on Official Statistics (2009). *Good practice guidelines for the development and reporting of indicators*. Wellington: Statistics New Zealand.

Commission of the European Communities (2009). Communication from the Commission to the Council and the European Parliament. *GDP and beyond. Measuring progress in a changing world*. COM(2009) 433 final. Brussels.

Craven, P., & Wahba, G. (1971). Smoothing noisy data with Spline functions. *Numerische Mathematik, 31*, 377-403.

European Commission (2018). Proposal for a regulation of the European Parliament and of the Council on the European Regional Development Fund and on the Cohesion Fund. 2018/0197(COD). COM(2018)372 final. Strasbourg.

Friedman, J.H. (1991). Multivariate Adaptive Regresion Splines. *The Annals of Statistics, 19*(1), 1-141.

Greenwell, B.M. Boehmke, B.C., & McCarthy, A. J., (2018). A Simple and Efective Model-Based Variable Importance Measure, arXiv:1805.04755v1 [stat.ML].

Maggino, F. (2017a). Dealing with Syntheses in a System of Indicators (Chapter 5). In F. Maggino (ed.) Complexity in Society: From Indicators Construction to their Synthesis (pp. 115-138). Cham, Switzerland: Springer International Publishing. DOI: 10.1007/978-3-319-60595-1

Maggino, F. (2017b). Developing Indicators and Managingthe Complexity (Chapter 4). In F. Maggino (ed.) Complexity in Society: From Indicators Construction to their Synthesis (pp. 87-114). Cham, Switzerland: Springer International Publishing. DOI: 10.1007/978-3-319-60595-1

Mazziotta, M., & Pareto, A. (2019). Use and Misuse of PCA for Measuring Well-Being. *Social Indicators Research, 142*(2), 451–476. DOI:10.1007/s11205-018-1933-0

O'Donnell, G., Deaton, A., Durand, M., Halpern, D., & Layard, R. (2014). Wellbeing and Policy. London: Legatum Institute.

Sánchez, A., & Ruiz-Martos, M. (2018). Europe 2020 Strategy and Citizens' Life Satisfaction. *Journal of Happiness Studies, 19*(8), 2315–2338. DOI: 10.1007/s10902-017-9928-0

Stiglitz, J., Sen, A., & Fitoussi, J. (2011). Mismeasuring Our Lives: Why GDP doesn't Add Up. New York: The New Press.

Van den Bergh, J. (2009). The GDP paradox. *Journal of Economic Psychology, 30*(2), 117–135.