Evaluation of the socio-economic situation of European Union countries, taking into account accuracy of statistical data

Małgorzata Stec[1], Małgorzata Wosiek[2]

Abstract

The aim of this paper is to evaluate the socio-economic situation of EU countries in 2016, taking into consideration accuracy of statistical data. The study used ten variables defining the socio-economic situation of EU countries. Linear ordering of EU countries was made using the zeroed unitarisation method. An assessment of the impact of uncertainty in the measurement of diagnostic variables on the value of a synthetic measure was also carried out. For this purpose, a procedure using the Monte Carlo method was proposed. The results indicate that the accuracy of statistical data may influence the results of the linear ordering of EU countries.

Keywords: European Union countries, synthetic measure, Monte Carlo method

***JEL Classification*:**C15, O11, O57

DOI: 10.14659/SEMF.2018.01.50

1 Introduction

Credible and high quality statistical data should provide a basis for the objective empirical research, as this decides about the research's final results. Person who uses statistical data, when taking them from official sources, considers them to be accurate. However, the way of gathering statistical information will not be error-free. It must be highlighted that a reduction of a number of errors in data published in yearbooks is achieved by applying complicated analyses and corrective calculations, which are employed by institutions gathering such data (e.g. Eurostat and statistical offices of each country), but obtained statistical data cannot be deemed accurate. In the economic literature, the analysis concerning influence of errors, which stem from inaccuracy of statistical data, is ignored, which may result in drawing incorrect conclusions in regard to the researched topic.The aim of this paper is to evaluate a socio-economic situation of EU countries, including the accuracy of statistical data. 10 variables determining socio-economic situation of EU countries in year 2016 were employed. The linear ordering of EU countries was done by the zero edunitarisation method. Moreover, an influence of uncertainty of measurement of diagnostic variables on the values of synthetic

[1] Corresponding author: University of Rzeszów, Faculty of Economics, Department of Quantitative Methods and Economic Informatics, M. Ćwiklińskiej 2 St., 35-601 Rzeszów, Poland, e-mail: malgorzata.a.stec@gmail.com.

[2] University of Rzeszów, Faculty of Economics, Department of Microeconomics, M. Ćwiklińskiej 2 St., 35-601 Rzeszów, Poland, e-mail: mwosiek@ur.edu.pl.

measure was investigated. For this, a procedure using the Monte Carlo method was proposed.

2 Theoretical basis for statistical data accuracy

The quality of statistical data is defined by three characteristics (Domański and Pruska, 2000; Kordos, 1988):

- usefulness of data in regard to users' needs (this postulate is fulfilled when a user solving specific topics, plans and realizes special research),
- validity (results are less useful with passing of time),
- accuracy (expressed by similarity of statistical information to real values, that is, the value which would be obtained if for all units of a group under investigation, the data was gathered and processed without errors).

Measurement error is a dissimilarity between determined value (measured) and real value. Question arises, this being the case when the published data can differ from the real data, whether it is worth using them for research or not? For such question, the answer can only be positive, however, it is important to ensure that such statistical data is as accurate as possible.

In technical science the problem regarding evaluation of the results' accuracy was settled in 1995 in the following document Guide to the Expression of Uncertainty in Measurement (2008). It was determined that the components of a measured result are: measured value and a bracket of uncertainty around this value. The formal definition of the term 'uncertainty of measurement' is as follows: uncertainty (of measurement) parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measured. The parameter a uncertainty of measurement may be, for example, a standard deviation called standard measurement uncertainty (or a specified multiple of it), or the half-width of an interval, having a stated coverage probability (Balazs, 2008).As Diettrich (1991) notes, all measurements are subject to error because no quantity can be known exactly, hence, any measurement has a probability of lying within a certain range. No measurement is perfect. The idea of an uncertainty in measurements is nevertheless something that has to be accepted as far as possible allowed for.

The accuracy of statistical data can be, as in technical science, identified with an error or uncertainty of measurement. In case of statistical data, the real value of measured quantity is very often unknown. In such situation, the Uncertainty of Measurement theory is employed.

In the Fig. 1, an interpretation of uncertainty of statistical parameter is presented. Read from a yearbook or taken from a data base, a value is deemed as a nominal value of a variable *X*. It is marked with X_n symbol for a multi-dimensional analysis, where *n* denotes a number of the following variable taken from a set of diagnostic variables. The essence of the presented approach is an assumption that there is no certainty whether this nominal value is the real value. It is presumed, however, that with the assumed distribution of probability the real value can be found in a range indicated by this distribution. Both distribution of probability and range of uncertainty of diagnostic variables were calculated on the basis of researcher's knowledge regarding methods of collecting this kind of data.

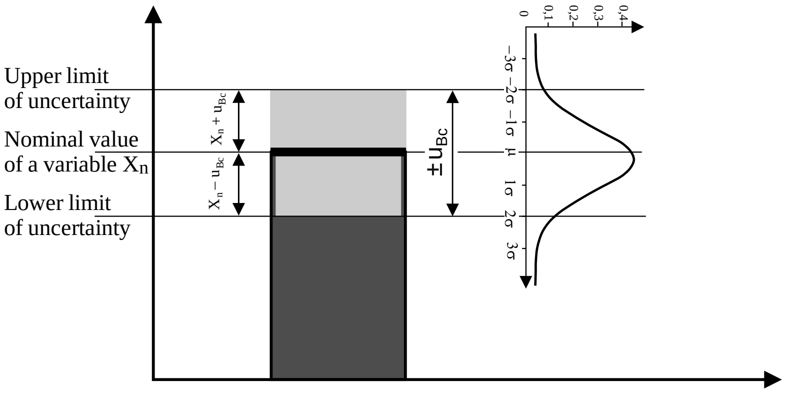

Fig. 1. The essence of setting the uncertainty of statistical value.

In yearbooks, no information regarding the calculated values of uncertainty affecting a specific statistical value is published. Due to this situation, the authoress performed such calculations on a basis of the available knowledge regarding methods of obtaining data used to determine a specific value.

Nominal value for each $X_n$variable employed to calculate a synthetic measure was taken from the Eurostat database. Calculated relative value of uncertainty $u_{Bc}$was converted to an upper and a lower value of a range created around the nominal value, in which, with regard to an assumed probability (depending on an adopted distribution), the real value is present.

$$X_R = X_n \pm u_{Bc} \tag{1}$$

where: X_R - range where the real value of variable is present,

X_n - nominal value of variable,

u_{Bc} - calculated value of variable's uncertainty.

Depending on the way of obtaining data, a value of uncertainty will differ. The most accurate data is collected by official state registers, which are legally responsible for its updates. Such registers include various records, for instance, census, legal entities, institutions, etc. However, state registers are not fully up-to-date which means that they are susceptible to uncertainty concerning real values included in data. One example can be errors caused by a lack of regular updates, which results in data being different from the real values. For instance: people going abroad permanently or for a longer period of time not always report this fact for a record. Similar situations also occur for other registers.

Assuming that the subsequent variables, which are a basis for developing a synthetic measure, are affected by uncertainty (their real values are unknown, only their estimations), it is necessary to analyse whether these ranges are not too vast to "blur" the difference between the subjects of research. Fig. 2 in a graphical way presents the essence of comparison of the values of the variables of two objects (countries), for which the ranges of uncertainty "overlap". Such case will occur if for a variable being a stimulant, the upper limit of uncertainty for a subject that has a lower position in a rank has bigger value, than the lower limit of uncertainty of a subject that is higher in a rank (Stec, 2017).Overlapping of ranges of uncertainty may take place if values of variables for two or more objects (countries) barely differ and calculated uncertainties are relatively big. With little diversification of subjects' values of variables the following situation may arise, in which a few objects can be characterised by similar values of a specific variable, which hinders the interpretation of real differences between these objects. Analogous situation relates to synthetic measures, too.

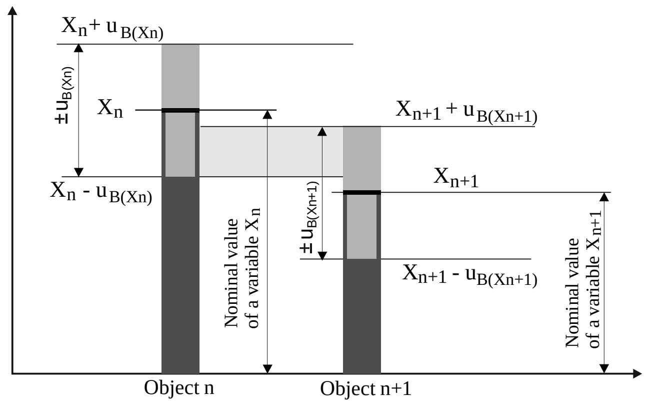

Fig. 2. A case of overlapping ranges of uncertainty between two subjects.

3 Methods applied

In this paper, the zeroed unitarisation method was used in order to calculate a value of synthetic measure for all countries in terms of socio-economic situation in year 2016. (Kukuła, 2014).

A normalization of the variable values was conducted using the following formulas:

$$z_{ij} = \frac{x_{ij} - \min_i\{x_{ij}\}}{R_j} \quad \text{for stimulating factors} \qquad (2)$$

$$z_{ij} = \frac{\max_i\{x_{ij}\} - x_{ij}}{R_j} \qquad \text{for non-stimulating factors} \qquad (3)$$

where: z_{ij} - normalized value of j-th variable for the i-th object, x_{ij},- value of j-thvariable for the i-th object, R_j- range for the j-th variable.

The synthetic measure was calculated as an arithmetic mean of the normalized value of variables:

$$MS_i = \sum_{j=1}^{m} z_{ij} \qquad (4)$$

where: MS_i - synthetic measure in i-th object, m- number of variables.

Due to the fact that the employed method to calculate the synthetic measure leads to change of the measuring scale, the calculation of the uncertainty of the synthetic measure by analytic method would provide false results. That is why, the Monte Carlo[3] method was employed and calculations were performed in the *R* application (Walesiak and Gatnar, 2009).

In order to calculate the value of uncertainty of the synthetic measure, it was concluded that for a sample big enough (the calculations were performed on a set of data counting 1000 for each object), standard deviation can be considered as a measure of distribution identified with a range of uncertainty of the synthetic measure. The following algorithm of procedure was chosen:

- for each diagnostic variable, 1000 values were drawn for every object (28 countries) which fulfilled the following conditions:
  - value of each drawn variable was comprised in the assumed range of uncertainty created around the nominal value for this variable,
  - drawn values for each variable had normal distribution,
- from the drawn variables, sets of data were created (1000 sets for each object),
- drawn sets of data underwent normalization,
- on the basis of the normalized set of data, the synthetic measures were calculated (1000 values of partial measures),
- from 1000 set synthetic, partial measures standard deviation was calculated, which constituted the measure of uncertainty of synthetic measure.

The above mentioned procedure allowed for calculation of nominal values of synthetic measures for each object (country) and their uncertainty.

[3] The Monte Carlo method solves a numerical problem by performing calculations on random variables, it is a tool for solving quantity problems, when analytical methods based on formulas, estimators, etc., fail.(Kopczewska et al., 2016; Liu, 2008; Niemiro, 2013).

4 Diagnostic variables employed in the research

The evaluation of a socio-economic situation for 28 EU countries was done with an employment of 10 diagnostic variables (Table 1)[4]. Also, an influence of the uncertainty of diagnostic variables on the results, regarding ordering of objects in terms of values of proposed variables was assessed. Table 1 shows a compilation of investigated values of uncertainty and a number of cases of overlapping (collisions) ranges of uncertainty for each diagnostic variables caused by too small difference between their nominal values in relation to the calculated uncertainty.

Table 1. Diagnostic variables determining socio-economic situation of all EU countries in year 2016.

	Diagnostic variables	Calculated value of uncertainty	Number of collisions
X1.	Crude rate of natural change of population per 1000 persons (S)	0.01%	1
X2.	Infant deaths rate per 1000 live births (D)	0.01%	10
X3.	Employed persons per 1000 population (S)	0.1%	5
X4.	Unemployment rate (based on LFS) in % (D)	0.5%	3
X5.	Gross value added by kinds of activity services in% (S)	1.0%	19
X6.	Exports of goods and services in % of GDP (S)	1.0%	5
X7.	Investment rate in % (S)	1.0%	16
X8.	Research and development expenditure (% of GDP) (S)	1.0%	7
X9.	At-risk-of-poverty rate in % (D)	1.0%	11
X10.	Students of higher education institutions per 10 thous. population (S)	0.01%	1

Depending on the way of obtaining statistical data, different values of uncertainty were adopted. In case of three diagnostic variables, the uncertainty was assumed at the level of 0,01%, for five variables 1%. For two cases, uncertainty was assumed as follows - at the level of 0,1% and 0,5%.In case of a diagnostic variable X5, there were 19 colliding situations; that is those in which the ranges of uncertainty of these variables partially overlap. It should be

[4]Examples of similar research can be found in: Barro, 1991;Del Campo et al., 2008;Ertur and Koch,2006;Grzebyk and Stec, 2015; Qizilbash, 2001; Rakauskienė and Kozlovskij, 2014; Stec et al., 2014.

notes that with 28 objects (countries), the maximum number of collisions (overlapping of neighbouring objects) totals 27.

5 Empirical results

Table 2 presents the values of synthetic measures for EU countries calculated by the zeroedunitarisation method with the values of uncertainty range for this measure. The nominal values of the respective synthetic measures are subject to analysis, which were put in order in the traditional way. Ranges of uncertainty constitute additional information enabling to verify the created rank of objects. Standard deviation is an uncertainty measure for each country's synthetic measure. To make the obtained result more trustworthy, standard deviation σ_i was multiplied by coefficient 1,96. This determined value, added to the nominal value of the MS_imeasure, created the upper limit of the uncertainty range and when subtracted - the lower limit of uncertainty ($MS - 2\sigma$; $MS + 2\sigma$).

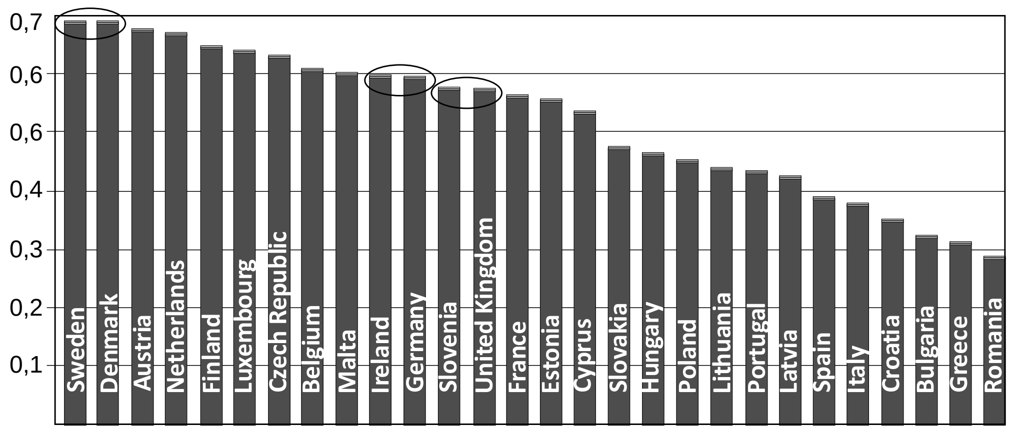

Fig. 3. Rank of EU countries in the area of a socio-economic situation with reference to the uncertainty of diagnostic variables in year 2016.

The uncertainty of the synthetic measure expressed in a relative form ranges from 0,13% for Luxembourg to 0,55% for Greece.The value of the synthetic measure MS_i is a basis for ordering the EU countries in terms of a socio-economic situation. In 2016, as regards the socio-economic situation, the leading positions in the rank of EU countries were taken by: Sweden, Denmark, Austria and Netherlands. Last positions were taken by: Romania, Greece, Bulgaria and Croatia. The comparison of upper and lower limit values of neighbouring uncertainty ranges in the rank of objects allowed for verification whether the differences between nominal values of synthetic measures of these objects are not so small enough that there are no grounds to differentiate their positions (Fig.3).

Table 2. Value of synthetic measure for EU countries calculated by the zeroedunitarisation method with the values of uncertainty ranges for this measure*.

No	Countries	Value of synthetic measure	2σ	$MS - 2\sigma$	$MS + 2\sigma$
1	Sweden	0.6877	0.0017	0.6860	0.6894
2	Denmark	0.6861	0.0019	0.6843	0.6880
3	Austria	0.6730	0.0018	0.6712	0.6748
4	Netherlands	0.6660	0.0019	0.6641	0.6679
5	Finland	0.6448	0.0017	0.6431	0.6465
6	Luxembourg	0.6363	0.0008	0.6355	0.6371
7	Czech Republic	0.6290	0.0016	0.6274	0.6306
8	Belgium	0.6066	0.0019	0.6047	0.6085
9	Malta	0.5991	0.0019	0.5972	0.6010
10	Ireland	0.5945	0.0011	0.5934	0.5956
11	Germany	0.5930	0.0017	0.5913	0.5947
12	Slovenia	0.5730	0.0016	0.5714	0.5746
13	United Kingdom	0.5719	0.0018	0.5701	0.5738
14	France	0.5599	0.0019	0.5580	0.5618
15	Estonia	0.5548	0.0016	0.5532	0.5564
16	Cyprus	0.5338	0.0020	0.5318	0.5358
17	Slovakia	0.4726	0.0016	0.4711	0.4742
18	Hungary	0.4619	0.0015	0.4604	0.4634
19	Poland	0.4494	0.0015	0.4479	0.4509
20	Lithuania	0.4358	0.0014	0.4344	0.4373
21	Portugal	0.4324	0.0016	0.4308	0.4340
22	Latvia	0.4226	0.0016	0.4210	0.4242
23	Spain	0.3870	0.0016	0.3854	0.3886
24	Italy	0.3762	0.0016	0.3746	0.3778
25	Croatia	0.3479	0.0015	0.3464	0.3494
26	Bulgaria	0.3203	0.0015	0.3189	0.3218
27	Greece	0.3100	0.0017	0.3083	0.3117
28	Romania	0.2856	0.0015	0.2841	0.2872

*by underlining we mean overlapping ranges

The analysis of the obtained results allows for confirmation that the differences of synthetic measures are too small to unequivocally acknowledge the positions of the following countries: Sweden and Denmark, Ireland and Germany, Slovenia and United Kingdom. There is, therefore, a risk of making a mistake that a country classified lower in ranks has, after all, a higher value of synthetic measure than a country classified higher in ranks. It was therefore concluded that there are no grounds for diversification of the positions of these counters.

Conclusions

The following conclusions can be drawn based on the carried out research:

- The evaluation of a socio-economic situation for 28 EU countries in 2016 was done with use of 10 diagnostic variables. The zeroed unitarisationmethod was employed for the empirical research.
- The results confirm the diversity of EU countries in terms of a socio-economic situation. In 2016 Sweden, Denmark, Austria and the Netherlands were the leaders. The lowest level in this context represent the following countries: Romania, Greece, Bulgaria and Croatia.
- Taking into consideration the uncertainty of the values of synthetic measures, it may influence the final conclusions drawn from the research.
- In case of comparison research, in which the results of linear order of the objects have a considerable meaning, it seems purposeful to include the influence of uncertainty of the values of diagnostic variables on the calculation of the synthetic measure.

References

Balazs, A. (2008). International Vocabulary of Metrology-Basic and General Concepts and Associated Terms. *Chemistry International, 25*.

Barro, R. J. (1991). Economic growth in a cross section of countries. *The quarterly journal of economics, 106*(2), 407-443.

Del Campo, C., Monteiro, C. M. & Soares, J. O. (2008). The European regional policy and the socio-economic diversity of European regions: A multivariate analysis. *European Journal of Operational Research, 187*(2), 600-612.

Dietrich, C. F. (1991). *Uncertainty, calibration and probability: the statistics of scientific and industrial measurement*. CRC Press, 3.

Domański, C. & Pruska, K. (2000). *Nieklasyczne metody statystyczne*. Polskie Wydawnictwo Ekonomiczne, 65.

Ertur, C. & Koch, W. (2006). Regional disparities in the European Union and the enlargement process: an exploratory spatial data analysis, 1995–2000. *Annals of Regional Science*, *40*(4), 723–765.

Grzebyk, M. & Stec, M. (2015). Sustainable development in EU countries: concept and rating of levels of development. *Sustainable Development*, *23*(2), 110-123.

JCGM/WG 1 2008 Working Group. (2008). Evaluation of measurement data–guide to the expression of uncertainty in measurement. In:*Tech Rep JCGM 100: 2008 (BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP and OIML*.

Kopczewska, K., Kopczewski, T. & Wójcik, P., (2016). *Metody ilościowe w R: aplikacje ekonomiczne i finansowe*. CeDeWu, 221.

Kordos, J. (1988). *Jakość danych statystycznych*. Państwowe Wydawnictwo Ekonomiczne.

Kukuła, K. (2014). Budowa rankingu województw ze względu na wyposażenie techniczne rolnictwa w Polsce. *WiadomościStatystyczne*, *7*, 62-76.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.

Niemiro, W. (2013). *Symulacje stochastyczne i metody Monte Carlo*. University of Warsaw, Warszawa.

Qizilbash, M. (2001). Sustainable development: Concepts and rankings. *Journal of Development Studies*, *37*(3), 134-161.

Rakauskienė, O. G. & Kozlovskij, V. (2014). Overview of EU economic cohesion process performance for new member states. *Procedia-Social and BehavioralSciences*, *110*, 284-292.

Stec, M. (2017).*Taksonomiczna analiza poziomu rozwoju społeczno-gospodarczego województw Polski. Studium przypadku-województwo podkarpackie*. Wydawnictwo Uniwersytetu Rzeszowskiego, 87-99.

Stec, M., Filip, P., Grzebyk, M. & Pierścieniak, A. (2014). Socio-economic development in the EU member states—concept and classification. *Eng. Econ*, *25*(5), 504-512.

Walesiak, M. & Gatnar, E. (Eds.). (2009). *Statystyczna analiza danych z wykorzystaniem programu R*. Wydawnictwo Naukowe PWN.