Lecturer Florentin ŞERBAN, PhD
E-mail: florentin.serban@csie.ase.ro
Department of Applied Mathematics
The Bucharest Academy of Economic Studies
Doctoral School of Mathematics, University of Bucharest
Associate Professor Anca-Teodora ŞERBAN-OPRESCU, PhD
E-mail:teodora.oprescu@rei.ase.ro
Department of Modern Languages and Business Communication
The Bucharest University of Economic Studies
Associate Professor George-Laurenţiu ŞERBAN-OPRESCU, PhD
E-mail:george.serban@economie.ase.ro
Department of Economic Doctrines and Communication
The Bucharest University of Economic Studies

APPRAISAL OF SCIENTIFIC RESEARCH IN EUROPEAN COUNTRIES. AN ENTROPY-BASED ANALYSIS

Abstract. *In this paper, we put forward an efficient approach based on entropy measures to assess academic performance, by using the values of some relevant indicators provided for the 28 EU countries. The concept of entropy has been frequently utilized to evaluate the quantity and influence of information offered by the data under analysis. The methodology is based on the construction of a composite indicator, represented by the weighted sum of the indicators taken into account. The weights assigned to the indicators selected to perform the analysis stand as measure of significance for each factor involved in the construction of the composite indicator. The results obtained using the present methodology are very similar to official international rankings, confirming the reliability and accuracy of this approach. In terms of economic insight, we compute an efficiency index as the ratio of the outcome value obtained and the mean value for the corresponding 10 years' time span of Research and Development (R&D) expenditure in Higher Education sector for every country. The results obtained using this approach reveal that some countries with lower research score succeed to publish papers at much lower costs than developed countries in R&D sector.*

Keywords: *entropy, indicator, scientific research, comprehensive assessment.*

JEL Classification: C10, C63, I23

1. INTRODUCTION

Bearing in mind that research is a special and unconventional type of activity, assessing scientific research performance of units/individuals requires a complex methodological toolkit, necessary to account for a lot of factors that may be influenced by both research quantity and quality.

Increased interest in the topic has, in addition, been prompted by the emergence of a wide market where research has become a particular type of commodity and a profit-making business. Recently, one of the major concerns in this field deal with the development of powerful methods to assess the performance of the academic research segment in the context of huge complexity of available data. Furthermore, individual or unit academic performance evaluation, as well as University rankings have been frequently put under scrutiny for lack of consistency in relevant data sets and reliable methodology. Several attempts have been made to develop efficient measures to appraise performance of academic units. Among these, Bonaccorsi and Daraio, 2007, Daraio et al., 2011 and Daraio et al., 2014 contributed to methodology development, with a special focus on EU landscape. Nevertheless, quantitative evaluation of scientific research deploys a set of complex methodological tools, many of which have not raised deserved awareness.

Considering the above, entropy constitutes a useful tool to assess not only the amount, but also the consequence of information provided by certain criteria used to assemble a composite indicator. In the present study we use a wide-ranging method for assessing scientific research when the values of some relevant indicators are accessible for every country. The approach relies on methodology used in Liu and Cui, 2008, Ouyang et al., 2012, Badin et al., 2016 to assess scientific research for 28 EU countries. The methodology used in this paper has already been successfully applied in Badin et al., 2016, to assess academic performance in the case of Romanian universities. Based on this methodology, the analysis relies on the design of a composite indicator constructed as the weighted sum of indicators considered. As in Badin et al., 2016, preliminary steps include computing the weights of indicators; we then move to assess the discrepancies amid indicator values using the concept of entropy, which allows to assign relative weights to the indicators; when a large degree of difference is detected, the entropy is smaller and more exact information is thus offered; consequently, the weight of the corresponding indicator is larger; the weights of the indicators stand as measure of the degree of significance for each criterion implicated in defining the composite indicator; we use entropy method to obtain the weight vector based on a predefined decision matrix. Our investigation considers only publications indexed in Web of Science database - Articles and Proceedings Papers, corresponding to 28 EU countries for the 2005-2014-time span. As such, the application documents EU countries in terms of their scientific research output. Numerical results of our findings are presented and discussed.

From an economic perspective, the present analysis also features computing an efficiency index as the ratio of the outcome value obtained and the mean value for the corresponding 10 years' time span of Research and Development (R&D) expenditure in Higher Education sector for every country.

The paper continues as follows: Section 2 describes in more detail the concept of entropy. Section 3 presents the algorithm used to evaluate scientific research activity. In section 4 the efficiency index is computed as the ratio of the outcome value and the mean value for the corresponding 10 years' time span of R&D expenditure in Higher Education sector for every country. Data processing and numerical results are displayed for the 28 EU countries. The paper ends with a Conclusions section and directions for upcoming research.

2. ENTROPY MEASURES

Based on Boltzman's (1896) use of entropy in statistical physics, C.E. Shannon (1948) proposed entropy to measure the amount of information supplied by a probabilistic experiment or a random variable. The new theory had a surprising evolution and its application has quickly extended to other fields - physics, biology, economics, sociology, culture.

The success of the theory at the basis of information measurement and information utility is founded on findings of famous mathematicians, among whom R.W. Hartley, R.S. Fisher, N. Wiener, J. Von Neumann, A.I. Hincin, D.K. Fadeev, A.N. Kolmogorov, A. Renyi, C. Tsallis. The Romanian school of Mathematics has brought important contributions to this theory; O. Onicescu introduced and studied a new measure of the informational state of a system, Onicescu's information energy. S. Guiasu introduced the weighted entropy concept as a measure of information supplied by a probabilistic experiment for which elementary events are characterized by objective probabilities and by qualitative (objective or subjective) weights.

Uncertainty can be quantified using the concept of entropy and may be interpreted as randomness or fuzziness. In order to deal with randomness in decision-making problems in different fields, various techniques which use information measures or risk measures have been developed, see, for example, Preda et al., 2015, Dedu and Serban, 2015, Toma, 2014, Preda et al., 2014, Serban et al., 2013, Serban et al., 2011.

Definition. The Shannon entropy corresponding to the random variable $X = (x_1, x_2, ..., x_n)$ with state probability vector $(p_1, p_2, ..., p_n)$, $\sum_{i=1}^{n} p_i = 1$, is given by: $H(X) = -\sum_{i=1}^{n} p_i \ln p_i$.

The concept of entropy displays two interpretations. Thus, it can be regarded as a measure of uncertainty featured by the experiment or random variable or measure of information provided by the experiment or random variable under consideration. The difference between these two interpretations resides in our stance in relation to the moment of performing the experiment. If positioned before carrying out the experiment, then entropy measures the uncertainty concerning the results of the endeavor. When positioned after the completion of the experiment, then entropy measures the information supplied by the experiment.

Proposition. We have state system $X = (x_1, x_2, ..., x_n)$ with the state probability vector $(p_1, p_2, ..., p_n)$, $\sum_{i=1}^{n} p_i = 1$. Let H(*X*) be the Shannon entropy corresponding to the system *X*. Then the following properties hold:

1. $H(X) \geq 0$, with equality relationship if and only if there exists $i \in \{1,2,...,n\}$ such that $p_i = 1$ and $p_j = 0$, $\forall j \in \{1,2,...,i,-1,i+1,...,n\}$;

2. $H(X) \leq \ln n$, with equality if and only if $p_i = \frac{1}{n}$, $\forall\ i \in \{1,2,...,n\}$.

Generally, in systems theory, entropy is used to indicate the degree of internal disorder. When a system state is quantified by the use of a particular indicator, the observed significantly big differences in the values of the respective indicator stand for small value of the entropy measure; on the other hand, small differences correspond to a big value of the entropy measure.

If the indicator is of less importance for describing the system state, its weight is small. But, if it provides more information, its weight is large. Ideally, entropy reaches its peak value when indicator values used to assess the system status are equal.

Let us assume there are *m* evaluation objects and *n* evaluation indicators and the original data matrix $X = (x_{ij})_{m x n}$ is created. For an indicator *j*, the better the difference of the index value x_{ij}, the better the part played by the indicator in the overall appraisal.

We consider a state system $X = (x_1, x_2, ..., x_n)$ with the state probability vector $(p_1(x_1), p_2(x_2), ..., p_n(x_n))$. The Shannon entropy measure, expressed in this framework by $\mathrm{H}(X) = -\sum_{i=1}^{n} p_i(x_i) \ln p_i(x_i),$ has been used to evaluate the quantity of information available. Then, this measure has been normalized, in order to obtain values in the interval (0, 1), by using the measure $\mathrm{H}(X) = -\frac{1}{\ln n} \sum_{i=1}^{n} p_i(x_i) \ln p_i(x_i)$.

3. COMPREHENSIVE ASSESSMENT BASED ON ENTROPY

The present study aims to assess the scientific research output for the 28 EU countries, by using entropy measures. The data set used in this section and displayed in Table 1 takes into consideration the number of Proceedings papers and articles indexed in Web of Science database, corresponding to the 2005-2014-time span.

Table 1. ISI publications indexed in Web of Knowledge during 2005-2014

Country	Articles	Proceedings papers
Austria	114173	22592
Belgium	167053	28167
Bulgaria	21662	5782
Croatia	30114	7058
Cyprus	6980	2265
Czech Republic	88253	37185
Denmark	121067	14469
Estonia	13202	3365
Finland	102038	18198
France	622210	124592
Germany	877903	171702
Greece	96742	23392
Hungary	57022	12241
Ireland	81186	14785
Italy	503348	112258
Latvia	4780	5164

Lithuania	18280	7811
Luxembourg	5551	1275
Malta	1573	382
Netherlands	301183	42309
Poland	192491	47782
Portugal	92699	26019
Romania	58665	38407
Slovakia	28807	11598
Slovenia	32084	5840
Spain	438394	76296
Sweden	203496	27324
United Kingdom	808963	106518

The algorithm which will be used to determine the weights consists in performing the following steps:

Step 1.

a) Data standardizing. Original data matrix X = $(x_{ij})_{m\,x\,n}$ is first standardized by $y_{ij} = \frac{x_{ij} - \min x_{ij}}{\max x_{ij} - \min x_{ij}}$, j = 1, 2

b) Formula $z_{ij} = \frac{y_{ij} - \overline{y}_j}{S_j}$ is applied to continue the standardization, where $\overline{y}_j$ and S_j are the mean value respectively standard deviation of j^{th} index.

c) Entropy computation requires a natural logarithm, so the index value must be positive. We set $u_{ij} = z_{ij} + d$, where *d* is a number greater than $\left|\min z_{ij}\right|$. As a result, we obtain the standard matrix $U = (u_{ij})_{m\,x\,n}$. We have obtained $d = 0.81665$.

Table 2 displays the final results corresponding to Step 1.

Table 2. The results obtained by performing Step 1

Y1	Y2	Z1	Z2	U1	U2
0.12849	0.12964	-0.27625	-0.30358	0.54040	0.51307
0.18883	0.16218	-0.05981	-0.17479	0.75684	0.64186
0.02292	0.03152	-0.65491	-0.69191	0.16174	0.12474
0.03257	0.03897	-0.62031	-0.66243	0.19634	0.15422
0.00617	0.01099	-0.71500	-0.77316	0.10165	0.04349
0.09891	0.21482	-0.38235	0.03353	0.43430	0.85018
0.13636	0.08223	-0.24804	-0.49123	0.56861	0.32542
0.01327	0.01741	-0.68954	-0.74774	0.12711	0.06891
0.11464	0.10399	-0.32592	-0.40509	0.49073	0.41156
0.70822	0.72502	-0.32592	-0.40509	0.49073	0.41156
1.00000	1.00000	2.84976	3.14102	3.66641	3.95767
0.10860	0.13431	-0.34760	-0.28510	0.46905	0.53155
0.06327	0.06922	-0.51018	-0.54270	0.30647	0.27395
0.09085	0.08407	-0.41127	-0.48393	0.40538	0.33272
0.57259	0.65302	1.31667	1.76780	2.13332	2.58445
0.00366	0.02791	-0.72401	-0.70619	0.09264	0.11046
0.01906	0.04336	-0.66875	-0.64504	0.14790	0.17161
0.00454	0.00521	-0.72085	-0.79603	0.09580	0.02062
0.00000	0.00000	-0.73713	-0.81665	0.07952	0.00000
0.34189	0.24473	0.48919	0.15190	1.30584	0.96855
0.21786	0.27668	0.04431	0.27834	0.86096	1.09499
0.10399	0.14964	-0.36415	-0.22441	0.45250	0.59224
0.06515	0.22195	-0.50345	0.06176	0.31320	0.87841
0.03108	0.06547	-0.62566	-0.55755	0.19099	0.25910
0.03108	0.06547	-0.62566	-0.55755	0.19099	0.25910
0.49847	0.44311	1.05081	0.93704	1.86746	1.75369
0.23042	0.15726	0.08935	-0.19427	0.90600	0.62238
0.92133	0.61952	2.56758	1.63520	3.38423	2.45185

Step 2. Computation of the weights of indices.

a) The probabilities p_{ij} for the i^{th} sample in the j^{th} index are computed by using the formula $p_{ij} = \frac{u_{ij}}{\sum_{i=1}^{m} u_{ij}}$, $i = \overline{1,m}$; $j = \overline{1,n}$. Table 3 displays the final results corresponding to this stage.

Table 3. The results obtained by performing the first stage of Step 2

P1	P2	COUNTRY
0.02606	0.02514	**Austria**
0.03650	0.03145	**Belgium**
0.00780	0.00611	**Bulgaria**
0.00947	0.00756	**Croatia**
0.00490	0.00213	**Cyprus**
0.02094	0.04166	**Czech Republic**
0.02742	0.01595	**Denmark**
0.00613	0.00338	**Estonia**
0.02366	0.02017	**Finland**
0.02366	0.02017	**France**
0.17680	0.19392	**Germany**
0.02262	0.02605	**Greece**
0.01478	0.01342	**Hungary**
0.01955	0.01630	**Ireland**
0.10287	0.12664	**Italy**
0.00447	0.00541	**Latvia**
0.00713	0.00841	**Lithuania**
0.00462	0.00101	**Luxembourg**
0.00383	0.00000	**Malta**
0.06297	0.04746	**Netherlands**
0.04152	0.05365	**Poland**
0.02182	0.02902	**Portugal**
0.01510	0.04304	**Romania**
0.00921	0.01270	**Slovakia**
0.00921	0.01270	**Slovenia**

0.09005	0.08593	**Spain**
0.04369	0.03050	**Sweden**
0.16320	0.12014	**United Kingdom**

b) The computation of the entropy measure e_j corresponding to the j^{th} index is performed by using the formula $e_j = -\frac{1}{\ln m} \sum_{i=1}^{m} p_{ij} \ln p_{ij}$, $j = \overline{1,n}$.

We have obtained the results $e_1 = 0.83173$; $e_2 = 0.85087$.

c) The computation of the utility corresponding to the j^{th} index is performed by using the formula $d_j = 1 - e_j$.

We have obtained the results $d_1 = 0.16827$; $d_2 = 0.14913$.

d) The weight of the j^{th} index is standardized by using the transformation

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j}, \text{ for } j = \overline{1,n}.$$

We have obtained $w_1 = 0.53015$; $w_2 = 0.46985$

Step 3. Evaluation of samples. The value of the i^{th} sample in the j^{th} index is given by $f_{ij} = w_j \cdot y_{ij}$, for $j = \overline{1,n}$. Hence, the total value for the i th sample is given by

$f_i = \sum_{j=1}^{n} f_{ij}$, $i = \overline{1,m}$. Table 4 displays the final results corresponding to Step 3.

Table 4. The results obtained by performing Step 3

F1	F2	SCORE	COUNTRY
0.06812	0.06091	0.12903	**Austria**
0.10011	0.07620	0.17631	**Belgium**
0.01215	0.01481	0.02696	**Bulgaria**
0.01727	0.01831	0.03558	**Croatia**
0.00327	0.00516	0.00844	**Cyprus**
0.05244	0.10093	0.15337	**Czech Republic**

0.07229	0.03863	0.11092	**Denmark**
0.00704	0.00818	0.01522	**Estonia**
0.06078	0.04886	0.10964	**Finland**
0.37546	0.34065	0.71611	**France**
0.53015	0.46985	1.00000	**Germany**
0.05757	0.06311	0.12068	**Greece**
0.03354	0.03252	0.06607	**Hungary**
0.04816	0.03950	0.08766	**Ireland**
0.30356	0.30682	0.61038	**Italy**
0.00194	0.01311	0.01505	**Latvia**
0.01011	0.02037	0.03048	**Lithuania**
0.00241	0.00245	0.00486	**Luxembourg**
0.00000	0.00000	0.00000	**Malta**
0.18125	0.11499	0.29624	**Netherlands**
0.11550	0.13000	0.24549	**Poland**
0.05513	0.07031	0.12544	**Portugal**
0.03454	0.10428	0.13882	**Romania**
0.01648	0.03076	0.04724	**Slovakia**
0.01648	0.03076	0.04724	**Slovenia**
0.26426	0.20820	0.47246	**Spain**
0.12216	0.07389	0.19605	**Sweden**
0.48844	0.29108	0.77952	**United Kingdom**

Running the input data through the proposed algorithm leads to the ranking of countries in descending order, displayed in Table 5. The highest score belongs to the first ranked country.

Table 5. Research score of academic units

COUNTRY	OUTCOME
Germany	1.00000
United Kingdom	0.77952
France	0.71611
Italy	0.61038
Spain	0.47246
Netherlands	0.29624
Poland	0.24549
Sweden	0.19605

Belgium	0.17631
Czech Republic	0.15337
Romania	0.13882
Austria	0.12903
Portugal	0.12544
Denmark	0.11092
Finland	0.10964
Ireland	0.08766
Hungary	0.06607
Slovakia	0.04724
Slovenia	0.04724
Lithuania	0.03048
Croatia	0.03558
Bulgaria	0.02696
Estonia	0.01522
Latvia	0.01505
Greece	0.12068
Cyprus	0.00844
Luxembourg	0.00486
Malta	0.00000

4. ECONOMIC APPROACH: OUTCOME VERSUS INVESTMENT

Using data from Eurostat database, the efficiency index is computed as the ratio of the outcome value obtained in the previous section and an investment indicator, which represents the average value, for the corresponding 10 years' time span, of Research & Development Expenditure in Higher Education sector for every country, scaled by the mean investment value of the sample. The results obtained are presented in Table 6.

Table 6. Research score vs efficiency indices of academic units

COUNTRY	OUTCOME	INVESTMENT	EFFICIENCY INDEX
Austria	0.12903	0.9766	0.1321
Belgium	0.17631	0.8072	0.2184
Bulgaria	0.02696	0.0099	2.7143

Croatia	0.03558	0.0478	0.7442
Cyprus	0.00844	0.0187	0.4506
Czech Republic	0.15337	0.2627	0.5839
Denmark	0.11092	1.0235	0.1084
Estonia	0.01522	0.0476	0.3196
Finland	0.10964	0.6312	0.1737
France	0.71611	4.2357	0.1691
Germany	1.00000	5.9593	0.1678
Greece	0.12068	0.2654	0.4548
Hungary	0.06607	0.1025	0.6448
Ireland	0.08766	0.3200	0.2740
Italy	0.61038	2.7117	0.2251
Latvia	0.01505	0.0273	0.5521
Lithuania	0.03048	0.0699	0.4360
Luxembourg	0.00486	0.0324	0.1501
Malta	0.00000	0.0073	0.0000
Netherlands	0.29624	1.8896	0.1568
Poland	0.24549	0.4275	0.5743
Portugal	0.12544	0.4160	0.3015
Romania	0.13882	0.0638	2.1766
Slovakia	0.04724	0.0742	0.6363
Slovenia	0.04724	0.0431	1.0949
Spain	0.47246	1.7637	0.2679
Sweden	0.19605	1.5181	0.1291
United Kingdom	0.77952	4.2475	0.1835

These ranking displays very interesting results: countries with lower research score are more efficient with respect to expenditure in higher education. Countries with lower research score succeed in publishing papers at much lower costs than developed countries in the sector of research and development. At the same time, the analysis refers to a general set of data and the obtained results could be further confirmed by the use of a more specific set of data that would reveal in a more compelling manner the structure of investment in research and development in higher education and cost proportion specific for publication.

5. CONCLUSIONS AND FUTURE RESEARCH

As stated in the introduction, the present endeavor has been prompted by the increased scholar interest in research evaluation and methods to best make use of the density of available data. The ranking based on the research scores displayed in Table 5 is very similar to official international rankings, which validates the entropic approach as an efficient tool. This study confirms the findings in Badin et al., 2016, showing consistent results when different data sets are analyzed. In addition, the present study highlights a straightforward, reliable and practical instrument for assessing scientific research output using information measures. It should be noted that, when the ratio of the outcome value obtained using an entropy measure and the mean value for the corresponding 10 years' time span of Research & Development expenditure in Higher Education sector for every country is computed, the results indicate that countries with lower research score are more efficient with regard to expenditure in higher education.

This paper has taken into account two categories of publications of great significance for the research output. Although other indicators, such as country size or GDP, have not been explored in this study, they are to be added and observed in a future analysis, in order to complement and confirm the present results. Aside from adding to the data sets to be processed using information measures, future research also has in view to select what entropy measure would be liable to provide more dependable and well-rounded findings.

ACKNOWLEDGEMENT: "***This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-2905".***

REFERENCES

[1] **Badin, L., Serban, F., Serban-Oprescu, A.T., Dedu, S. (2016)**, ***Reflecting on Romanian Universities Ranking: An Entropy-based Approach to Evaluate Scientific Research,*** in: International Conference on Informatics in Economy, IE 2016: Education, Research & Business Technologies, 403-408;

[2] **Bonaccorsi A., Daraio C., Eds. (2007), *Universities and Strategic Knowledge Creation. Specialization and Performance in Europe*;** *Edward Elgar Publisher*, Cheltenham, UK;

[3] **Daraio C. et al. (2011),** ***The European University Landscape: A Micro Characterization Based on Evidence from the Aquameth Project*;** Research Policy, 40, 148-164;
[4] **Daraio, C., Bonaccorsi A., Simar L. (2014),** ***Rankings and University Performance: A Conditional Multidimensional Approach*,** DIAG Technical Report, no. 9;
[5] **Dedu, S., Serban, F. (2015),** ***Modeling Financial Data Using Risk Measures with Interval Analysis Approach***; Procedia Economics and Finance, 22, 610-617;
[6] **Liu, W., Cui, J. (2008),** ***Entropy Coefficient Method to Evaluate the Level of Sustainable Development of China's Sports;*** International Journal of Sports Science and Engineering, 2, 2, 2008;
[7] **Ouyang, D., Xu, B., Li, Y., Huang H. (2012)**, ***A Comprehensive Evaluation of Culture Force Based on Entropy Method: Cross-Sectional Data in Hunan Province***; Journal of Convergence Information Technology, 7, 23, 416-424;
[8] **Preda, V., Dedu, S., Sheraz, M. (2014),** ***New Measure Selection for Hunt and Devolder's Semi-Markov Regime Switching Interest Rate Model*;** Physica A, Volume 407, 350-359;
[9] **Preda, V., Dedu, S., Gheorghe, C. (2015),** ***New Classes of Lorenz Curves by Maximizing Tsallis Entropy under Mean and Gini Equality and Inequality Constraints*;** Physica A, Volume 436, 925-932;
[10] **Purcaru, I. *(*1988),** ***Informatie si corelatie***, *Enciclopedica Publishing;* Bucharest;
[11] **Shannon, C.E. (2001)**, ***A Mathematical Theory of Communication***; Computing and Communications Review, 5, 3–55;
[12] **Şerban, F., Ştefănescu, M.V., Ferrara, M. (2013),** ***Portfolio Optimization in the Framework Mean-variance-VaR;*** *Economic Computation and Economic Cybernetics Studies and Research*, 47, 1, 61-79;
[13] **Şerban, F., Ştefănescu, M.V., Dedu, S. (2011),** ***Building an Optimal Portfolio Using a Mean-VaR framework;*** In: Demiralp, M., Bojkovic, Z., Repanovici, A. (Eds.), Mathematical Methods and Techniques in Engineering and Environmental Science, 126-131;
[14] **Şerban, F., Ştefănescu, M.V., Dedu, S. (2011),** ***The Relationship Profitability-Risk for an Optimal Portfolio Building with Two Risky Assets and a Risk-free Asset;*** International Journal of Applied Mathematics and Informatics, 5, 4, 299-306;
[15] **Toma, A. (2014),** ***Model Selection Criteria Using Divergences;*** Entropy, 16, 5, 2686-2698;
[16] ec.europa.eu/eurostat.