A Comparative Analysis of Business and Economics Researchers in the Visegrad Group of Countries, Austria and Romania Based on the Data Obtained from SciVal and Scopus

Imre Dobos[1] | *Budapest University of Technology and Economics, Budapest, Hugary*
Péter Sasvári[2] | *University of Public Service, Budapest, Hungary*

Received 3.2.2023, **Accepted** (reviewed) 28.4.2023, **Published** 15.9.2023

Abstract

Research background: The aim of the paper is to compare the performance of economic researchers in Austria, Romania and the Visegrad 4 (Czech Republic, Hungary, Poland, and Slovakia) using performance indicators of researchers from the Scopus and SciVal databases. In the comparison of countries, Austria is included as a benchmark country, while the other five countries represent the countries of the former Eastern bloc. In the study, the definition of an economic researcher is based on indicators that can be obtained from databases. The study focuses first on the statistical properties of the indicators and then groups' researchers from countries using these indicators.

Purpose of the article: Paper pursued two goals. First, by presenting the relationships between the data obtained from the Scopus/SciVal databases, to present the most important key indicators, then to group the researchers with the help of the analyzed indicators, and to compare the publication performance of the chosen countries. A researcher is considered to be an economic researcher in the study whose at least thirty percent of the published articles in the SCImago database are in the subject areas of Business, Management, and Accounting and Economics, Econometrics, and Finance.

Methods: Three methods were used to perform the study. First, principal component analysis, multicollinearity analysis with variance inflation factor (VIF), and partial correlation analysis were performed using the correlation matrix. Second, using the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) ranking

[1] Department of Economics, Faculty of Economic and Social Sciences, Budapest University of Technology and Economics, Műegyetem rkp. 3, 1111 Budapest, Hungary. Corresponding author: e-mail: dobos.imre@gtk.bme.hu. ORCID: <https://orcid.org/0000-0001-6248-2920>.

[2] Faculty of Public Governance and International Studies, University of Public Service, Ludovika tér 2, 1083 Budapest, Hungary. E-mail: sasvari.peter@uni-nke.hu. ORCID: <https://orcid.org/0000-0002-4031-4843>.

procedure, researchers from each country were ranked using indicators. Finally, the distribution of ninths and tenths of ranked researchers was analyzed for each country. Three data sets were used for the analysis. A representative sample proportional to the population of a country, followed by the principle known in team sports that each country nominates the same number of athletes, and finally a dataset of all selected researchers.

Findings & value added: The first most important result can be stated that the stochastic linear relationships that can be described with the three data sets are very similar, the causal relationships are also the same. Based on the principal component analysis, the indicators can be divided into two groups: the component consisting of raw data and the component consisting of reference-based variables. In this case, too, the three datasets resulted in the same groups of variables. Of the eight indicators, two proved to be collinear: all references and the Hirsch index of all publications. A comparison of researchers from countries showed that economic researchers in Austria perform best, and researchers from other countries only follow in each dataset. The results are similar; it is difficult to rank between countries.

Keywords	***DOI***	***JEL code***
Science metrics, economics, management, multivariate statistics	*https://doi.org/10.54694/stat.2023.5*	*A11, C10, C12, I20*

INTRODUCTION

International competitiveness has become a top priority for countries in today's increasingly integrated political establishment, which is based on international relationships. Macilwain (2010) found that activities related to science, technology and innovation have a direct impact on social and economic well-being and also promote sustainable development. As for the Central European countries examined in the present study, the fullest possible transition to a knowledge-based economy is the key to their competitiveness as well as their positioning among the member states of the European Union. It can be achieved by investing in the modernization and development of higher education as well as research and development, in which respect the Visegrad Group of countries lag behind their western neighbours. According to a study by Bőgel et al. (2020), higher education in these countries is not moving in the right direction, and investment in R&D activities is low at both state and company levels. The transfer of new technologies, the ability to adapt to the changed environment is intended to avoid the middle-income trap in the studied countries, which ultimately threatens with the possibility of finding themselves in a stuck position if the Central European region is unable to renew and keep pace with the developed Western states. The biggest challenge for them in this regard is growth, in other words, their ability to achieve high productivity and create high value-added products, services and innovation in their own region. In their work, Nölke and Vliegenthart (2009) describe the Central European region as a region of dependent market economies where a strong hierarchy can be observed between the headquarters of multinational corporations and local corporations.

The present study examines six Central European countries. Four of these states also form a smaller entity within the European Union, which is based on their common political, cultural and historical ties (these states are the Czech Republic, Poland, Hungary and Slovakia). These member states are collectively referred to as the Visegrad Group of countries. What these states have in common is that they became communist satellite states of the Soviet Union after World War II. A kind of regime change and democratic transition as well as the creation of a market economy took place in these countries in the 1990s. By now these countries have become members of the North Atlantic Treaty Organization (NATO) and the European Union (EU), thus giving a significant contribution to the Europeanisation of their internal systems. Two other countries are added to the analysis as reference points, one of which is Austria as the leading state in the Central European region, with significant links to the Visegrad Group of countries. In addition to Austria, Romania is also included in the study as another reference country, which, as it happened in the case of the Visegrad Group of countries, came under the control of the Soviet Union, but the democratic transition did not take place there as peacefully as in the other four countries mentioned above. With regard to the EU, Austria and Romania are also good choices since Austria gained admission in 1995, much earlier than Romania, which joined the union twelve years later in 2007. The six studied countries differ in terms of the number of population, as Poland and Romania can be classified as larger countries, Austria, Hungary and the Czech Republic are medium-sized states, and Slovakia can be considered as a smaller country.

The paper aims to first analyze the statistical properties of indicators and data sets, then to analyze the performance of the selected countries in the discipline of Economics and Management based on the publication performance of researchers included in the data sets. The purpose of research is to compare the selected Central European countries based on three different data sets, one being a representative sample by the population of the countries, the second being a data set including the top 50 Economic researchers in each country, and the third data set consisting of every researcher satisfying the disciplinary requirements. These disciplinary – publication – requirements have been set in order to define who the Economic researcher in these countries is. In the paper, we first analyzed the relations between the indicators studied. We carried out the analyses for the three data sets simultaneously. These analyses included correlation analysis, principal component analysis, multicollinearity analysis, the linear regression estimation of the collinear indicators, and partial correlation analysis to determine the cause-and-effect relations. Then, the countries were ranked, Austria serving as a benchmark country.

The main research questions are how to define and who are the Economic researchers in the selected Central European countries. Secondly, what kind of correlations can be observed between the indicators describing the publication performance of the selected researchers? Thirdly, what are the principal components and what is cause-and-effect relation between these indicators, in order to define the best publication strategy leading to the highest publication performance of these researchers. Last but not least, what is the ranking of these countries based on the publication performance of their Economic researchers?

After the introduction, our study continues with the examination of the scientific performance of the selected countries, followed by the definition of the group of economic researchers and the issues surfaced during the compilation of the data set. In the next chapter, statistical analyses are carried out to explore the logical system of indicators characterizing the performance of the analyzed researchers. In following chapter, the TOPSIS ranking technique is used to determine the ranking of each country, our results are discussed then. Finally, conclusions are drawn from the research results.

1 EXAMINING THE SCIENTIFIC PERFORMANCE OF THE EAST CENTRAL EUROPEAN COUNTRIES: A SHORT LITERATURE REVIEW

The measurement of scientific performance can be performed at four levels: at the individual level of the researcher, at the level of scientific journals, at the level of research institutions (including universities and research centres), and at the level of countries (Gevers, 2014). Different bibliometric and scientometric performance indicators have been defined for each level, focusing primarily on the quality of scientific activity. Quality indicators are usually organized around an internationally accepted database and regulatory system, being used as reference points.

A ranking was found by Szuflita-Zurawska and Basinska (2021) in their study carried out among the Visegrad Group of countries. Based on their results, Poland has the highest productivity in terms of publication numbers, while the Czech Republic leads in terms of the number of publications per researcher in internationally indexed journals, and Hungary dominates in the number of ERC grants as well as publications written in international cooperation in general. A similar conclusion was reached by Dobos et al. (2021), whose results show that the Central European countries do not belong to the international scientific elite, a more conscious planning can be observed in only two of them, the Czech Republic and Poland, thus giving the leading role in the region to these two countries.

Few remarks should be taken regarding the selected Central European countries. On the one hand, it is important to mention that the studied countries are well comparable to one another because none of them belongs to the Anglo-Saxon countries so they may encounter language barriers in scientific publication (Jurajda et al., 2017). In addition, with regard to the post-Soviet states, a parallel can be drawn in their development in the field of social sciences and, more importantly, in the discipline of economics, which is a discipline that was neglected and pushed into the background during the Soviet period. In these countries, research in social sciences and economics could begin only in the 1990s and was significantly underfunded in comparison to natural and technical sciences (Vanecek, 2008).

Grančay et al. (2017) analyze the academic requirements defined regarding the economists of the Central European region between 2000 and 2015. They found that there was a dynamically increasing scholarly output of economists being a 317% increase regarding the Web of Science indexed publications and the impact of their publications achieve reputation interpreted in the increasing number of citations risen by 228%. They also found basic changes in the requirements by the implementation of newly defined science policies after the political transformation of these states. In their study and in Pajič's paper (2014) also, the most important pillars of national research evaluation measures are listed. Based on their results, national publication strategies can be drawn. While some of the countries – especially in the case of the Czech Republic – the local journals gained international indexing, Hungary has long been focusing on international collaborations and publishing in worldclass journals. Their findings show that from the absolute scholarly output, Poland and the Czech Republic are the leading countries of the region. On the other hand, Hungary thanks its traditionally strong scientometric evaluation system, publishes to a global audience extensively. Grančay and his co-authors' results point out that however these countries tend to become competitive actors in the international science community, they publish in their local and regional journals to a great extent, meaning that "only communication channels are internationalized, but not the communication itself". Economists from the region share similar risk to social scientists, that in order to publish to international journals, they tend to focus more on topics interesting for a wider global audience. This means in the long term that local or country-specific topics are abandoned (Löhkivi et al., 2012; Purkayastha at al., 2019).

2 WHO CAN BE CONSIDERED AN ECONOMIC RESEARCHER?

Due to the diversity and complexity of social and economic challenges, scientific research increasingly requires an inter- or multidisciplinary approach (Abramo et al., 2012). There is a clear consensus that there are no longer any field constraints or barriers in today's scientific work and that interdisciplinary research has become widely spread (Porter et al., 2009). Carrying out research in increasingly larger research groups is a general trend and the participants of these research groups tend to be professionals active in different disciplines. It occurs precisely because of these processes that the accurate definition and separation of certain disciplines are becoming increasingly difficult. Economics has long been separated from the social sciences and including several subfields. Such subfields include Finance, Economics and Organizational science or operations research (Truc et al., 2020). According to the classification of the Economic and Social Research Council (2021), Economics has two major disciplines, which are Economics and Management and Organizational science, respectively.

In their article (2021), Dobos et al. made an attempt to determine who could be considered an Economics Researcher. In their study, according to a preliminary definition, Economics Researchers were considered as such if a significant proportion of their publications appeared in an economic journal. Accordingly, their analysis was conducted in line with the subject areas listed in the SCImago Journal Ranking (SJR). In the SJR, journals are grouped at two levels, which are subject areas and subject categories. In their study, Economics Researchers were selected on the basis of subject areas, which are the following:

- Business, Management, and Accounting,
- Decision Sciences, and
- Economics, Econometrics, and Finance.

An additional filter condition during the compilation of the sample was that only those having already had publications in one of the listed disciplines were taken into account. This is also the group of researchers with an author ID in the Scopus citation database.

Then, in the second stage of the selection those having already published in at least three ranked journals in the subject areas mentioned in the field of Economics or in the field of social sciences were taken into account. This second selection step became necessary because if a researcher published only in the field of Decision Sciences, he could have gotten into the data set as a mathematician involved in operations research without having published in the other two areas of economic sciences. Due to the strong relations between Economics and social sciences, certain subject categories listed within the subject area of the social sciences are considered in our analysis. These include Development, Human Factors and Ergonomics, and Transportation.

Based on all these considerations, in this paper, those researchers are considered to be Economic researchers who published at least 30 percent of their papers in the subject areas of Business, Management, and Accounting, and Economics, Econometrics, and Finance. It is important to note, however, that a journal can be indexed in several subject areas and subject categories, as well as it can cover a diverse research profile. However, the discipline classification of the SJR can provide a good indication of the focus of a given paper. When setting the lower limit at 30 percent, a rule of thumb was applied so most researchers were included in our data set.

3 COMPILATION OF THE DATA SET

We started the data analysis by selecting the researchers of the monitored countries who satisfy the requirements mentioned in the last section. For the data compilation, we used the Scopus citation database and the SciVal research intelligence program based on the Scopus database. Following the disciplinary requirements detailed above, we narrowed the data set alongside the two mentioned sets of subject areas in the SJR:

- Business, Management and Accounting, and
- Economics, Econometrics and Finance.

We then determined the parameters to form the basis of our ranking. Apart from the subject areas, the total number of papers published in these two areas was chosen as a filter variable. While doing so, we relied on Scopus, which assigns to each researcher the subject areas where the researcher has already published. This function of the database is based on the scientific classification of journals in the SJR ranking. As SciVal can only highlight 150 researchers per country, we used this maximum number as a basis. Thus, 900 researchers active in the field of Economics from the six selected countries were added to the initial database. However, we were compelled to face the fact that this solution did not prove to be completely reliable, either.

We had to keep on narrowing down the initial database of 900 researchers as there were several problems with identifying whether the researchers were actually employed in a given country. The initial database we had compiled from Scopus also listed those researchers among the 150 professionals in a given country who were ever (even temporarily) employed in that country, so when their affiliation was geographically identified, that country was included in their publications. In order to eliminate this problem, we had to examine the profiles of the researchers in the database one by one and find their actual institutions.

In summary, the final dataset includes researchers per country who, at the time of the data compilation, were employed in that country according to Scopus and had already had publications in the two chosen subject areas. Therefore, after the narrowing phase, our initial data set of 900 researchers was narrowed down to 658 in the distribution shown in Table 1. The database of Dobos et al.'s paper (2021) was also selected on this principle, therefore the representativeness remained questionable, so the results can be up for a debate. The table also contains two additional datasets. One dataset, which contains 278 researchers, is a representative sample. The basis of representativeness was the number of populations per country. The other dataset contains 300 data items, which includes the same number of researchers from each country, selected according to a similar principle as seen in team competitions. There, too, each national team takes part with an equal number of competitors, regardless of the size of the country.

Table 1 Distribution of researchers in the dataset by countries

Country	Number of researchers in the datasets by countries					
	278 persons		300 persons		658 persons	
	person	%	person	%	person	%
Austria	27	9.712	50	16.667	106	16.109
Czech Republic	32	11.511	50	16.667	110	16.717
Poland	114	41.007	50	16.667	114	17.325
Hungary	29	10.432	50	16.667	82	12.462
Romania	60	21.583	50	16.667	133	20.213
Slovakia	16	5.755	50	16.667	113	17.173
Total	**278**	**100.000**	**300**	**100.000**	**658**	**100.000**

Source: Own editing based on the Scopus database

The researchers selected for the 278-person and 300-person datasets were identified by using a ranking technique per country, i.e., the best professionals from each country were included in the datasets. Among the available methods, researchers from six countries were ranked using the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method. An essential feature of TOPSIS method is that the available data set is normalised in terms of the variables. There are several options for normalisation, including Euclidean distance, transforming data to [0,1] interval. After normalisation, the already normalised data are weighted by the method, which can be done with two approaches: subjective and objective. In the case of subjective weighting, the weights of the aspects are predetermined, while in the case of objective weighting, they start from the statistical properties of the data set and weighting is built on them. Two methods are known for the latter approach. One is the entropy-based method (Zou et al., 2006), while the other is the standard deviation-based Criteria Through Intercriteria Correlation (CRITIC) method (Diakoulaki et al., 1995). TOPSIS performs further calculations on the weighted normalised data matrix. For each aspect, the method determines the ideal and negative ideal, i.e., the preferred and rejected values.

In the next step, for each observation – in the present study, for each researcher – we determine the distance from the ideal and negative ideal points. A quotient is then formed, which is between the values of 0 and 1 and the distance from the ideal point is proportional to the sum of the distances measured from the two preferred points. This value is 1 if observation (researcher) is preferred in everything, and 0 if observation is the least good in everything. The geometric approach to this is to examine the distance from two privileged points in the normalised space of the variables, which is based on the triangular inequality well-known in geometry.

Based on Scopus and SciVal, we measured performance through eight variables freely available on the researchers' datasheets. The variables also included publication, citation, and co-author indicators. The variables used for the analysis are as follows (with abbreviations in brackets):

- number of total publications (DOC),
- total number of citations (TOT-CIT),
- the Hirsch index (H-I),
- number of co-authors (C-A),
- number of publications between 2010 and 2019 (SO),
- number of citations to publications published between 2010 and 2019 (CIT),
- the five-year Hirsch index between 2015 and 2019 (H5-I), and
- the Field-Weighted Citation Impact (FWCI).

The first four of the variables include achievements over the entire research career, while the last four considers the activity of the last ten years between 2010 and 2019 before the date of data collection (June 25, 2020). Of the variables, the FWCI certainly needs further explanation, while the others, including the Hirsch index, are well-known. The FWCI basically shows how referential the author's publications are. If the value of FWCI is greater than one, more citations are expected from the publication compared to other publications in similar subject areas. The calculation algorithm of the FWCI index can be found in Elsevier (2019) and Purkayastha el al. (2019).

4 STATISTICAL ANALYSIS OF THE DATA SET

During the examination of the three datasets (278 persons; 300 persons; 658 persons), we performed six analyses on eight variables to examine the relationship between the variables. We first mapped out the stochastic relationship between the variables by analysing the correlation matrix. Then, by using principal component analysis, we reduced the number of variables. In the third analysis, we analysed the multicollinearity between the variables by using the variance inflation factor (VIF). In the fourth stage of our analysis, collinear variables filtered with the help of VIF were estimated using linear regression. The fifth analysis explored the causal relationship between the variables using partial correlation. Statistical analyses were performed in parallel on the three datasets in order to compare their results. Our results show that for the parallel analysis, some comparability of the obtained results can also be performed, because the results obtained for the three data sets show only minor differences.

4.1 Correlation analysis

Table 2 summarizes the results of the correlation calculation. It stands out that the correlation between the selected variables is very high except for the FWCI index, while the FWCI shows a very weak linear relationship with the other six variables. By their nature, H-I and CIT show a weakly moderate relationship with FWCI, as both are citation-linked variables. There is a strong and very strong linear relationship between the other seven variables.

Another interesting feature of the correlations is that H-I shows a relatively strong correlation with all of the variables. The correlation matrix suggests that the variables can be divided into two groups. It can be observed that there is a strong correlation between the number of publications and the publications of the last 10 years, between all citations and publications of the last 10 years, and between the two Hirsch indices, as shown by the results marked in grey. The analysis also points out that the correlation coefficients of the three datasets differ insignificantly from each other, i.e. they reinforce each other's effect and direction. Each of the correlation coefficients is significant at level of .000 in Table 2. Due to lack of space, we have therefore not indicated the significance levels with the usual stars in SPSS.

Table 2 Correlations between the variables

Variables	Number of items	DOC	TOT-CIT	H-I	SO	CIT	H5-I	FWCI
C-A	278	−0.580	−0.379	−0.494	−0.466	−0.512	−0.476	−0.330
	300	−0.613	−0.409	−0.536	−0.494	−0.542	−0.519	−0.381
	658	−0.417	−0.254	−0.313	−0.339	−0.337	−0.315	−0.222
DOC	278		0.543	0.690	0.629	0.491	0.373	0.042
	300		0.582	0.741	0.680	0.526	0.474	0.147
	658		0.555	0.696	0.653	0.500	0.423	0.150
TOT-CIT	278			0.876	0.389	0.804	0.409	0.370
	300			0.855	0.428	0.796	0.448	0.427
	658			0.838	0.410	0.792	0.438	0.407
H-I	278				0.519	0.838	0.585	0.451
	300				0.572	0.823	0.660	0.549
	658				0.533	0.798	0.621	0.527
SO	278					0.433	0.522	0.145
	300					0.462	0.555	0.199
	658					0.492	0.582	0.238
CIT	278						0.707	0.612
	300						0.736	0.680
	658						0.725	0.651
H5-I	278							0.584
	300							0.638
	658							0.633

Source: Own editing based on the Scopus database

4.2 Principal component analysis

Table 3 shows the components of the variables. In the principal component analysis of the eight variables, we obtained two components for all three datasets that accounted for more than 70 percent of the variance. The fit of the model according to the Kaiser-Meyer-Olkin test was between 0.788 and 0.790, which values represent a mean model according to the accepted categorization.

In all three principal component analyses, we obtained two components each, which explain almost the same variance as rotation. It is also interesting that in each analysis the same variables were included in each component, but the variable H-I can be assigned to both components in each case. Since the majority of correlations are high, i.e. greater than 0.4 in absolute value, we can expect high collinearity between them, so examining multicollinearity was the next step to be taken.

4.3 Examination of multicollinearity with VIF index

There is no uniform rule in the literature as to which VIF values above which variables can be considered collinear. Although there are some empirically tested VIF thresholds that range from 2.5 to 10, in the case

Table 3 Variable components

Size of dataset	278 persons		300 persons		658 persons	
Variance explained	73.510 %		75.906 %		71.307 %	
KMO test	.789		.790		.788	
Variable	Component		Component		Component	
	1	2	1	2	1	2
Variance (%)	37.889	35.621	38.080	37.826	39.031	32.276
DOC	**0.925**	0.076	**0.922**	0.158	0.203	**0.888**
SO	**0.787**	0.139	**0.827**	0.153	0.272	**0.761**
H-I	**0.679**	**0.617**	**0.659**	**0.653**	**0.704**	**0.575**
C-A	**–0.632**	–0.329	**–0.644**	–0.356	–0.113	**–0.610**
FWCI	–0.084	**0.910**	–0.026	**0.926**	**0.891**	–0.063
CIT	0.478	**0.809**	0.441	**0.831**	**0.848**	0.394
H5-I	0.367	**0.720**	0.416	**0.719**	**0.751**	0.342
TOT-CIT	**0.571**	**0.577**	**0.542**	**0.599**	**0.649**	**0.485**

Note: Methods used: principal component analysis and Varimax rotation with Kaiser normalisation. Values in bold indicate values in the matrix greater than 0.5 to help assign components to variables.
Source: Own editing based on the Scopus database

of filtering redundancy out, there is no set of theoretical or logical rules by which they can be reliably determined. For this reason, accepting the recommendations of several studies (Lafi et al., 1992; Liao et al., 2012; O'Brien, 2007), 5 was chosen as a threshold. A similar analysis was performed in paper Dobos et al. (2021).

In the initial step, we used the inverse of the correlation matrix from principal component analysis, because the diagonal of the inverse matrix contains the VIF values with the inclusion of the other remaining variables. In the next step, the operation of calculating the diagonal of the inverse matrix was repeated, but only after the variable with the highest VIF value was dropped from the analysis. These steps were carried out until the VIF values fell below the predefined limit, in our case below 5.

The evolution of VIF values and the sequential screening of the variables are summarized in Table 4. It is worth noting here that there is no deterministic algorithm for filtering out collinear variables. As a first step, it is recommended in the literature to filter out the variable with the highest VIF value but any variable above the threshold is also suitable for the first step. In the next step, there are again two options: either the element with the highest VIF value is selected again or the variable with the largest decrease in the value of VIF. In our case, the first option was chosen – i.e. the element with the highest VIF value. The decision is justified by the fact that in the first step the VIF value of the variable H-I was the highest. This was followed by the CIT variable, while the collinear values for the other six variables were not very high.

The examination of the initial VIF values immediately revealed that the total number of publications, the number of co-authors, the publications of the last 10 years, the Hirsch index of the last 5 years, and the initial value of the FWCI index are less than 5, that is the threshold, meaning that these variables could not be included in the collinear variables to be eliminated due to the stepwise decrease of the VIF value. As a result, it can be concluded that the Hirsch index and the citations to the publications of the last 10 years had a linear dependence on the other variables. If we consider the content of the variables, the latter become evident. It is worth emphasizing, however, that in this case, too, the results of the three datasets move together.

Table 4 The evolution of VIF values

Variables	Number of items	1. phase	2. phase	3. phase
DOC	**278**	3.547	2.659	2.651
	300	4.418	3.034	3.033
	658	3.138	2.354	2.354
TOT-CIT	**278**	6.200	3.820	1.722
	300	6.048	3.792	1.865
	658	5.183	3.470	1.714
H-I	**278**	8.773		
	300	9.418		
	658	6.523		
C-A	**278**	1.876	1.853	1.816
	300	2.001	1.961	1.906
	658	1.287	1.273	1.261
SO	**278**	2.023	2.022	2.005
	300	2.195	2.195	2.186
	658	2.227	2.220	2.220
FWCI	**278**	2.255	2.149	1.937
	300	2.854	2.491	2.214
	658	2.359	2.178	1.937
CIT	**278**	6.674	6.497	
	300	6.576	6.572	
	658	5.864	5.859	
H5-I	**278**	3.117	3.010	2.227
	300	3.591	3.374	2.610
	658	3.351	3.187	2.515

Source: Own editing based on the Scopus database

4.4 Linear regression estimation of collinear variables

In the linear regression estimation of the collinear variables, the filtered two variables are estimated with the remaining six variables, the parameters and significance levels of which are illustrated in Table 5. Linear regression was performed by stepwise regression using the SPSS statistical analysis program instead of the usual enter method. Because we had two dependent variables for all of the three datasets, we performed six stepwise linear regressions. As it can be seen in the table, the R2 values of the six linear models are all very high, being above 0.8. The parameters of the variables and the constant are significant in each model, indicating the goodness of the models.

In estimating the Hirsch index, publications from the last 10 years did not participate in any of the model types. In addition, the number of co-authors was not included in the 278-person sample. In the citations of the last 10 years, however, neither the total number of publications nor the number of publications in the last 10 years participated in any of the regressions. The latter phenomenon can be attributed to the fact that the citation variable can be explained by variables derived from other citations.

Table 5 The parameters and significance levels of the equations

H–I	Const.	DOC	TOT–CIT	C–A	SO	FWCI	H5–I	R^2
278 persons	2.862	0.061	0.003	–	–	0.674	0.331	**0.882**
Significance	*.000*	*.000*	*.000*	–	–	*.000*	*.000*	–
300 persons	1.589	0.084	0.003	0.010	–	1.239	0.373	**0.894**
Significance	*.000*	*.000*	*.000*	*.017*	–	*.000*	*.000*	–
658 persons	1.904	0.072	0.004	0.005	–	1.012	0.342	**0.846**
Significance	*.000*	*.000*	*.000*	*.008*	–	*.000*	*.000*	–
CIT	**Const.**	**DOC**	**TOT–CIT**	**C–A**	**SO**	**FWCI**	**H5–I**	R^2
278 persons	–141.125	–	0.205	–0.712	–	61.312	44.802	**0.845**
Significance	*.000*	–	*.000*	*.006*	–	*.000*	*.000*	–
300 persons	–129.974	–	0.193	–0.699	–	79.459	43.090	**0.847**
Significance	*.000*	–	*.000*	*.005*	–	*.000*	*.000*	–
658 persons	–100.354	–	0.206	–0.262	–	62.023	41.211	**0.829**
Significance	*.000*	–	*.000*	*.009*	–	*.000*	*.000*	–

Source: Own editing based on the Scopus database

4.5 Analysis of partial correlations: cause-and-effect

Partial correlation is suitable for filtering out the effect of other variables when determining the correlation between two variables in a linear model. This can also be interpreted by mapping out the causal relationship between the two variables. Table 6 shows the partial correlations used to describe the causal relationships.

There are three types of causality, which are summarized in Pearl (2009):

- temporal priority,
- relationship, and
- non-spurious relationship.

In our case, temporal causality can be applied, for which we can build two causal chains. The first chain consists of the variables **C-A → DOC → TOT-CIT → H-I**, which means that the author must first build a network of co-authors and collaborate with other authors, if it is done successfully, the communication will be completed. The paper, which provides meaningful and significant results, is given a citation, and then the Hirsch index is calculated from the publications published by the author and the citations given to them. Acting similarly, we can form a logical chain about the last 10 years with the following causal sequence: **SO → CIT → H5-I → FWCI**. Of course, the two chains show temporality but it is worth noting that temporality does not necessarily mean an existing linear relationship, that is, correlation.

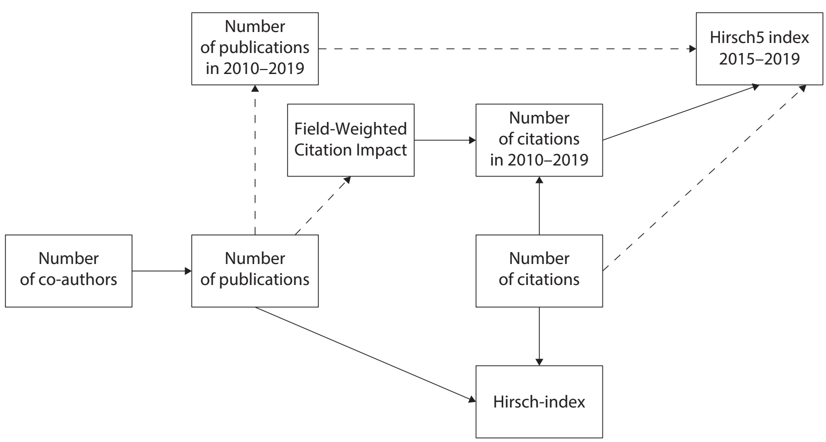
When exploring causal relationships, partial correlation values above 0.3 in absolute value are considered. There are five values between 0.40 and 0.62, while four more values are found between 0.30 and 0.40. If a value is entered in the two ranges for at least two datasets, it is assigned to the higher value. In Table 6, the examined partial correlations are indicated by colour. In this case, too, our results obtained in the three data sets prove to be very similar.

Table 6 Partial correlations

Variables	Number of items	DOC	TOT-CIT	H-I	SO	CIT	H5-I	FWCI
C-A	278	–0.417	0.054	0.111	–0.070	–0.157	–0.053	–0.199
	300	–0.426	0.055	0.141	–0.059	–0.169	–0.036	–0.204
	658	–0.281	0.021	0.104	–0.030	–0.097	–0.025	–0.094
DOC	278		–0.132	0.500	0.337	–0.036	–0.082	–0.389
	300		–0.121	0.560	0.334	–0.030	–0.034	–0.438
	658		–0.068	0.500	0.424	–0.017	–0.067	–0.310
TOT-CIT	278			0.620	0.052	0.473	–0.396	–0.149
	300			0.611	0.046	0.549	–0.432	–0.212
	658			0.575	0.020	0.565	–0.372	–0.190
H-I	278				0.009	0.163	0.185	0.217
	300				0.001	0.025	0.246	0.357
	658				–0.057	0.029	0.221	0.277
SO	278					–0.093	0.363	–0.093
	300					–0.063	0.330	–0.107
	658					0.015	0.377	–0.109
CIT	278						0.464	0.267
	300						0.455	0.302
	658						0.441	0.312
H5-I	278							0.214
	300							0.207
	658							0.252

Source: Own editing based on the Scopus database

Figure 1 Causal relationships between the variables



Source: Own editing based on the Scopus database

Figure 1 shows the causal relationships between the variables. Relationships between 0.40 and 0.62 are indicated in continuous and correlations between 0.30 and 0.40 in dashed lines. The figure shows that the block related to citations – including all citations, citations from the last ten years, the H-index, H5-I, and FWCI index – depends on the total number of publications, the number of publications in the last ten years, and on the number of co-authors. This suggests that the number of publications shows a strong correlation with the evolution of citations, while the number of co-authors is positively related to publication indices, i.e., to the total number of publications.

Based on the results, we can conclude that according to the causal system to be drawn, an increase in the number of co-authors increases the number of publications for a given author, and then the number of publications can increase the number of citations and thus the Hirsch indices.

5 RANKING OF RESEARCHERS USING TOPSIS RANKING TECHNIQUE

The TOPSIS method has already been used for compiling the data set. In the normalisation phase, we used the transformation of the variables to the interval [0,1], while the entropy-based method was used to determine the weights. This means that the rankings were not based on the rankings within each country, but on all the researchers in the data set.

5.1 The population-based sample including 278 researchers

First of all, it is worth examining the representative data set in proportion of its population with the TOPSIS method. The total number of researchers was 278, who were divided into ninths after the ranking. Thus, only 30 researchers were placed in the last ninth and 31 in the others. The means and standard deviations for each country are summarized in Table 7.

Table 7 shows that researchers in Austria have the highest average efficiency according to TOPSIS, followed by researchers from the Czech Republic and Slovakia, Hungary, Poland and finally Romania with the same values. It can be observed in the table that the average of the positions also shows this order. The relative standard deviations for the Slovak and Romanian researchers show that these countries have the largest deviation from their national average, while the most balanced data are seen in the case of the Polish professionals.

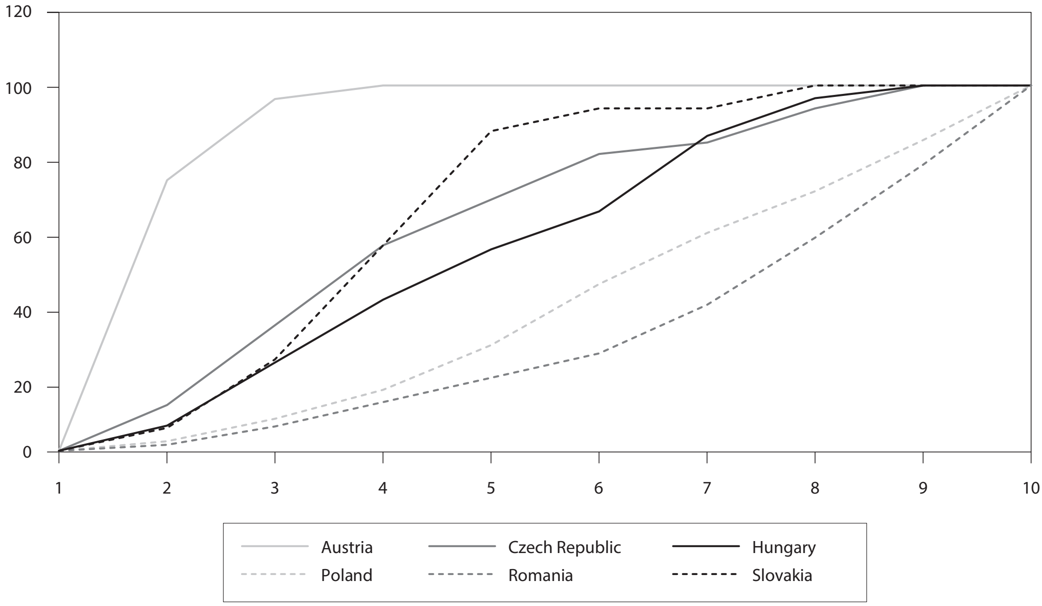
Data in Figure 2 is divided into ninths, which illustrates the extent to which researchers from each country are ahead in the rankings if their proportion is expressed in percentage. Thus, we examine how the distribution changes with the accumulation of each ninth by country. It can be seen from the figure that Austrian researchers give the best performance in total, followed by researchers from the Czech Republic and Poland. Hungary is found in the fourth place, ahead of Slovakia and Romania.

Table 7 The TOPSIS and ranking average, deviation and relative deviation of the 278-person data set

Countries	TOPSIS			Position		
	Average	Deviation	Relative deviation	Average	Deviation	Relative deviation
Austria	0.56	0.04	0.079	20.65	17.54	0.849
Czech Republic	0.48	0.03	0.066	100.16	61.93	0.618
Poland	0.45	0.03	0.059	165.18	70.33	0.426
Hungary	0.47	0.03	0.061	117.93	59.93	0.508
Romania	0.45	0.12	0.273	188.58	187.38	0.994
Slovakia	0.48	0.07	0.147	88.06	130.81	1.485

Source: Own editing based on the Scopus database

Figure 2 The ninths of the 278-person data set by countries



Source: Own editing based on the Scopus database

5.2 Same number of researchers per country, 300 researchers

After the population-based data set, the database containing the same number of researchers per country – a total of 300 people – is examined using the TOPSIS method. After sequencing, the researchers were divided into tenths so that each tenth would contain the same number of researchers, 30-30 persons. The means and standard deviations for each country are shown in Table 8.

The table shows that Austrian researchers have the highest average efficiencies according to TOPSIS, followed by their Czech counterparts. They are followed by Polish and Hungarian researchers, while Slovak and Romanian experts close the list with almost the same values. The average of the rankings also follows this order, while in terms of relative standard deviations, the Slovak and Romanian researchers

Table 8 The TOPSIS and ranking average, deviation and relative deviation of the 300-person data set

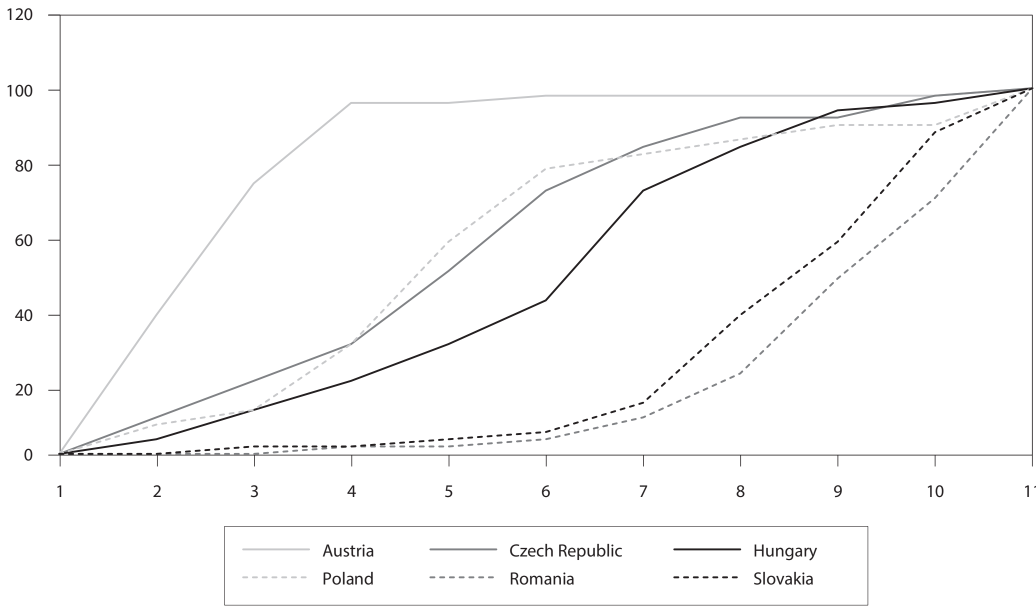
Countries	TOPSIS			Position		
	Average	Deviation	Relative deviation	Average	Deviation	Relative deviation
Austria	0.37	0.08	0.204	45.32	45.12	0.996
Czech Republic	0.31	0.04	0.126	123.60	66.13	0.535
Hungary	0.30	0.03	0.088	149.90	65.44	0.437
Poland	0.30	0.03	0.110	125.46	72.30	0.576
Romania	0.27	0.01	0.044	237.02	47.52	0.200
Slovakia	0.28	0.01	0.039	221.70	47.56	0.215

Source: Own editing based on the Scopus database

show the largest difference from their national average again. In this case, too, the Polish professionals have the most balanced data.

Figure 3 shows a similar picture to Figure 2. It can be observed that the Austrian researchers provide the best overall performance in this dataset as well. They are followed by the Czech Republic and Poland, while the Hungarian researchers are in the fourth place, followed by the Slovak and Romanian professionals.

Figure 3 The tenths of the 300-person data set by countries



Source: Own editing based on the Scopus database

5.3 Database of all available researchers, including 658 persons

Finally, the third data set was also examined using the TOPSIS method. This database consists of the profiles of all available researchers, including a total of 658 persons. Just as seen above, the researchers were divided into tenths. Table 9 shows the means and standard deviations for each country.

Table 9 The TOPSIS and ranking average, deviation and relative deviation of the 658-person data set

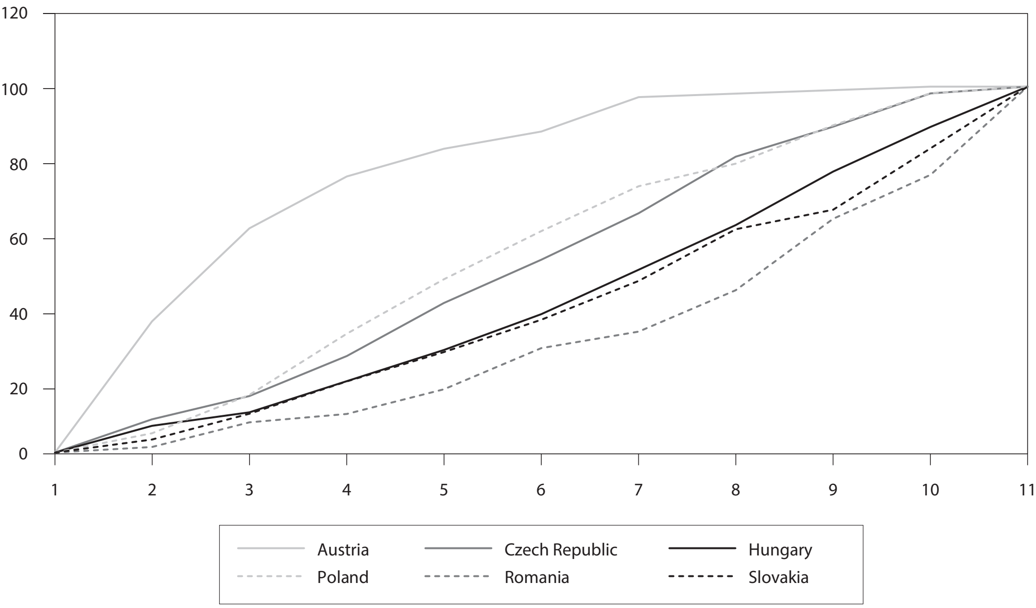
Countries	TOPSIS			Position		
	Average	Deviation	Relative deviation	Average	Deviation	Relative deviation
Austria	0.53	0.07	0.138	137.59	123.76	0.899
Czech Republic	0.47	0.05	0.096	310.95	161.81	0.520
Hungary	0.46	0.04	0.089	177.10	375.87	2.122
Poland	0.47	0.04	0.086	299.15	158.11	0.529
Romania	0.45	0.04	0.083	440.78	169.73	0.385
Slovakia	0.46	0.03	0.076	393.58	183.64	0.467

Source: Own editing based on the Scopus database

As in the previous two cases, researchers from Austria have the highest average efficiency according to TOPSIS, followed by the Czech Republic and Poland. Hungary and Slovakia closely follow each-other, and finally Romania closes the list. The average of the rankings also follows this order.

Figure 4 shows a similar result to those of Figures 2 and 3. It can be seen that the Austrian researchers provide the best performance in this data set as well. They are followed by the Czech Republic and Poland, while Hungary is ranked in the fourth place, ahead of Slovakia and Romania.

Figure 4 The tenths of the 658-person data set by countries



Source: Own editing based on the Scopus database

6 DISCUSSION

The analyses were two-ways in the paper. On the one hand, we analyzed the statistical properties of three compiled data sets to evaluate whether they differ substantially. Results show that the statistical properties are similar in all three data sets. This result shows that the correlations are similar in the data sets studied between the variables. Furthermore, the correlations are average and strong correlations. A weaker correlation can be found only in the case of the co-author and FWCI indicators. These results point out that the three data sets are correlated.

The relationship between the variables was then examined by principal component analysis. With principal component analysis, we obtained those two components explaining nearly three-quarters of the variance for all three datasets. This only confirms that the eight variables are linearly highly related. Therefore, it is worth examining the multicollinearity between the variables.

Multicollinearity was tested using the variance inflation factor (VIF). We obtained the result that the same two variables in each database, including the Hirsch index and the number of references received in the last ten years between 2010 to 2019, depend linearly on the other variables. This suggests that all the variables being duplicated in view of the two altered periods studied are eliminated from the analysis. This also means that the results of the analyses do not change significantly even after leaving these two variables. The two variables, as dependent variables were then linearly estimated with the remaining six variables, as independent variables in search of the answer to how we could recover the two variables from the databases with the other variables.

With partial correlation coefficients regarding the temporal causality, we then tested the causal relationships between our statistical variables (criteria). We obtained the result using the partial correlation that the number of co-authors is considered to be the most important input variable, while the output variable of the causal network is the five-year Hirsch index. Of course, the causality study also confirmed that Hirsch indices depend primarily on the number of publications and citations.

The other question stated at the beginning of the research was to compare and rank the countries based on the publication performance of the Economic researchers. From this aspect, we obtained the ranking of countries. This ranking shows the outstanding publication performance of Austrian Economic researchers, followed by the Czech and Polish researchers. Hungarian researchers are ranked in the middle based on all three databases, while Romania is considered to be the least successful compared to other countries. Hungarian researchers are generally stronger than their Slovak counterparts. These results suggest there we can observe still the economic and political effects long-lasting in the selected Central European countries, as we see leading position of Austrian researchers.

CONCLUSIONS

Higher education institutions in Central European countries are not at the forefront of the international scientific vanguard, they can rarely be found in the mainstream, and in terms of courses in Economics having worldwide popularity, they tend to be in the last third. Their fallback within the European Union member states can be explained by the publication performance, which is at a distance from the average publication performance of the Western European countries. To achieve a higher number of publications in internationally indexed journals by the authors of these countries, a change of attitude and culture within the academic community seems to be essential. However, an examination of the otherwise rather redundant Scopus database shows that the elite committed to economics research in Central Europe has a well-elaborated publishing strategy: they focus on increasing the number of publications, publishing them in the form of co-authorship, while intending to expand the volume of their citations on the basis of these factors, which will also have a positive effect on the changes in their Hirsch index. Based on their publication performance, typical groups cannot be described in terms of leading Economics Researchers in Central Europe, examining the dissemination of the results, they rather form a relatively homogeneous community. In the rankings of researchers made on the basis of quantitative results, the Polish and Czech economists are at the forefront, their Hungarian colleagues are found in the second line, while their Slovak and Romanian counterparts are observed to be lagging behind.

In the course of the future development of the survey methodology, it is definitely worth taking the population of each country into consideration, as it seems likely that the number of higher education institutions (academic research facilities) is related to the population (GDP volume, number of university and college students). In the future, the qualitative weighting of the quantitative indicators extracted from the Scopus database can also be performed, due to which it is worth considering the classification of a publication or citation (for example Q1/Q4 classifications in the SJR list).

ACKNOWLEDGMENTS

The authors thank Anna Urbanovics for her help in preparing the paper.

References

ABRAMO, G., D'ANGELO C.A., DI COSTA F. (2012). Identifying Interdisciplinarity through the Disciplinary Classification of Coauthors of Scientific Publications [online]. *Journal of the American Society for Information Science and Technology*; 63(11): 2206–2222. <https://doi.org/10.1002/asi.22647>.

BŐGEL, GY., MÁTYÁS L., ODOR L. (2020). Keeping Pace with the 21st Century [online]. *CEU Working paper.* [cit. 19.4.2020]. <http://www.personal.ceu.hu/staff/matyas/Bogel_Matyas_Odor_Visegrad4.pdf>.

DIAKOULAKI, D., MAVROTAS, G., PAPAYANNAKIS, L. (1995). Determining objective weights in multiple criteria problems: the critic method [online]. *Computers & Operations Research*, 22(7): 763–770. <https://doi.org/10.1016/0305-0548(94)00059-H>.

DOBOS, I., MICHALKÓ, G., SASVÁRI, P. (2021). The publication performance of Hungarian economics and management researchers: A comparison with Visegrád 4 countries and Romania [online]. *Regional Statistics*, 11(2): 165–182. <https://doi.org/10.15196/RS110207>.

ELSEVIER. (2019). *Research Metrics Guidebook* [online]. [cit. 19.4.2020] <https://www.elsevier.com/research-intelligence/resource-library/research-metrics-guidebook>.

GEVERS, M. (2014). Scientific performance indicators: a critical appraisal and a country-by-country analysis [online]. In: BLOCKMANS, M., ENGWALL, L., WEAIREPP, D. *Bibliometrics: Use and Abuse in the Review of Research Performance*, Portland Press, 43–53. [cit. 19.4.2020]. <https://portlandpress.com/pages/vol_87_bibliometrics_use_and_abuse_in_the_review_of_research_performance>.

GRANČAY, M., VVEINHARDT, J., ŠUMILO, Ē. (2017). Publish or perish: how Central and Eastern European economists have dealt with the everincreasing academic publishing requirements 2000–2015 [online]. *Scientometrics*, 111(3): 1813–1837. <https://doi.org/10.1007/s11192-017-2332-z>.

JURAJDA, Š., KOZUBEK, S., MÜNICH, D., ŠKODA S. (2017). Scientific Publication Performance in Post-Communist Countries: Still Lagging Far Behind [online]. *Scientometrics*, 112(1): 315–328. <https://doi.org/10.1007/s11192-017-2389-8>.

LAFI, S., KANEENE, J. (1992). An explanation of the use of principal components analysis to detect and correct for multicollinearity [online]. *Preventive Veterinary Medicine*, 13(4): 261–275. <https://doi.org/10.1016/0167-5877(92)90041-D>.

LIAO, D., VALLIANT, R. (2012). Variance inflation factors in the analysis of complex survey data [online]. *Survey Methodology*, 38(1): 53–62. [cit. 19.4.2020]. <https://www.researchgate.net/publication/279622236_Variance_inflation_factors_in_the_analysis_of_complex_survey_data>.

MACILWAIN, C. (2010). What science is really worth. [online]. *Nature News*, 456: 682–684. <https://doi.org/10.1038/465682a>.

NÖLKE, A., VLIEGENTHART, A. (2009). Enlarging the Varieties of Capitalism: the Emergence of Dependent Market Economies in East Central Europe [online]. *World Politics*, 61(4): 670–702. <https://doi.org/10.1017/S0043887109990098>.

O'BRIEN, R. (2007). A caution regarding rules of thumb for variance inflation factors [online]. *Quality & Quantity*, 41(5): 673–690. <https://doi.org/10.1007/s11135-006-9018-6>.

PAJIĆ, D. (2014). Globalization of the social sciences in Eastern Europe: genuine breakthrough or a slippery slope of the research evaluation practice? [online]. *Scientometrics*, 102(3): 2131–2150. <https://doi.org/10.1007/s11192-014-1510-5>.

PEARL, J. (2009). *Causality.* Cambridge University Press.

PORTER, A.L., RAFOLS, I. (2009). Is Science Becoming More In-terdisciplinary? Measuring and Mapping Six Research Fields over Time [online]. *Scientometrics*, 81(3): 719–45. <https://doi.org/10.1007/s11192-008-2197-2>.

PURKAYASTHA, A., PALMARO, E., FALK-KRZESINSKI, H., BAAS, J. (2019). Comparison of two article-level, field-independent citation metrics: Field-Weighted Citation Impact (FWCI) and Relative Citation Ratio (RCR) [online]. *Journal of Informetrics*, 13(2): 635–642. <https://doi.org/10.1016/j.joi.2019.03.012>.

SZUFLITA-ŻURAWSKA, M., BASIŃSKA, B. A. (2021). Visegrád Countries' Scientific Productivity in the European Context: a 10-year Perspective Using Web of Science and Scopus [online]. *Learned Publishing*, leap.1370. <https://doi.org/10.1002/leap.1370>.

TRUC, A., SANTERRE, O., GINGRAS, Y., CLAVEAU F. (2020). The Interdisciplinarity of Economics [online]. *SSRN Electronic Journal.* <https://doi.org/10.2139/ssrn.3669335>.

VANEČEK, J. (2008). Bibliometric analysis of the Czech research publications from 1994 to 2005 [online]. *Scientometrics*, 77(2): 345–360. <https://doi.org/10.1007/s11192-007-1986-3>.