WEIGHTING METHOD FOR DEVELOPING COMPOSITE INDICES. APPLICATION FOR MEASURING SECTORAL SPECIALIZATION

Ana Maria SAVA
Bucharest University of Economic Studies, Romania

E-mail: anamaria.sava89@yahoo.com

Abstract
When building a composite index, one might desire attributing different weights to factors whose influence it aggregates. Deciding on what weight to allocate to each factor may prove to be a difficult task if there is no possibility of finding an independent variable for the construct one tries to quantify. The present paper proposes a twist in using Principal Component Analysis as means for determining the weights of multiple factors based on which an index may be created. One example where finding an independent variable is not an option might be: developing an index for measuring sectoral specialization. Although over the years several instruments for measuring this construct have been developed, there is still no unanimous and universally accepted way of quantifying sectoral specialization and this paper designs a new index for measuring it by applying the weighting method advanced herein.

Keywords: *Weighting method; determining weights; composite index; PCA; no independent variable; sectoral specialization; Arts, entertainment and recreation*

1. Introduction

The goal of the present paper is to advance an instrument that enables building composite heterogeneity indices that takes into account the compound weighted influence of multiple factors considered to be relevant for assessing a certain phenomena. Following a previous approach in building such an index (Sava, 2016), this paper proposes a weighting method for designing indices in the absence of an independent variable. For exemplification reasons, the paper will present this method's functioning mechanism by developing a new sectoral specialization index (focusing on the local recreation industry). The paper will guide all the way from identifying the key factors whose influence will be taken into account in the index and determining the weight each factor should be attributed, to actually computing the index and discussing the output.

Over the years, economists have developed numerous instruments for measuring sectoral specialization. Palan (2010) divides them into two categories: specialization indices and heterogeneity indices. The first category – specialization indices – measure a country's absolute level of specialization, while the second – heterogeneity indices – measure the deviation of a country's industrial structure as compared to the average structure of a reference group of countries. Each of the two approaches presents both strengths and weaknesses.

Specialization indices give as output measures that can be interpreted per se, thus allowing focused application (measuring specialization only for the item of interest) and enabling straight-forward analysis of time series. However, their greatest disadvantage is that in their computation the development of other structures is not taken into consideration. On the contrary, heterogeneity indices use as benchmark exactly the average economic structure of the elements considered in the analysis. The major downside of these indices is that employing biased samples can generate wrongful results.

Acknowledging both advantages and disadvantages of the two categories of indices, and given the topic of interest for the present paper, the attention will be focused on the second category – heterogeneity indices. Examples of this type of indices are numerous: the Organization for Economic Co-operation and Development uses *Hannah-Kay Index* to highlight the sectoral composition measured on 20 industrial aggregates (OECD, 2013). European Central Bank quantifies sectoral specialization by means of *Krugman Index* that portrays the structure of a country's economy as compared to the EU structure (ECB, 2004). National Bank of Slovakia uses, next to Krugman index, the *Concentration Index* that shows for a given country a specific industry's contribution to the EU total and the *Lilien Indicator* that measures the speed of structural changes in employment (Čutková and Donoval, 2004).

Other heterogeneity indices mentioned by Palan (2010) and used in practice are: the *Index of Inequality in Productive Structure* that is similar to Krugman index but which grants large deviations an increased weight, *Relative Gini Index* extensively used both in concentration and specialization analyses, or the *Theil Index* that represents in fact a variation of the *Shannon Index* which establishes the employment level of a country in relation to that of the countries considered as reference group. Ioncică et al. (2010) proposed an index that allows calculating the degree of specialization for services sectors. The index proposed by them, called *Tertiary Specialization Index*, takes into account the share of services in GDP, employment and exports. As built, it can be applied to determine how specialized is the whole tertiary sector of a country, or one of its key service industries.

Among the aforementioned instruments, only the Tertiary Specialization Index is a composite index, but it does not imply assigning different weights to factors whose influence it aggregates. In this consisted also the main limitation of the prior variant of the specialization index that will be further presented (Sava, 2016), and for which purpose the weighting method that will be further presented was designed.

2. Description of the weighting method used

First and foremost, it is worth mentioning that the weights assigned through the method presented in this chapter are directly linked to the dataset used in the analysis, as is the computation of the heterogeneity index. Therefore, prior to explaining the weighting method, the coordinates of the index must be fixed.

Because the matter at hand implies assigning weights to factors for creating a composite index (therefore, not being able to rely on an independent variable to decide on factors' relevance for the measured construct), the accuracy of the results provided by the index depends primarily on the choice of factors. These have to be relevant for the studied concept and objectively chosen. The optimal choice of the indicators to be aggregated in the index implies that they can be measured on an ordinal scale, although it is not compulsory for this scale to have a fixed point of origin or an upper/lower limit.

Usual indicators for measuring sectoral specialization are: the sector's contribution to GDP, the share of employment of the sector in total employment, the size of the industry, the degree of concentration of companies in the sector, the share of exports, the government spending allocated to the sector, the value of private investments in the sector, or the expenditure for research and development activities in the sector. From this perspective, the index proposed herein will not deviate from the norm, as the following four factors were envisaged, each of them considered to have a positive influence on the degree of specialization or development of an economic sector:

- Gross production of the sector as share of GDP ($\frac{GP_S}{GDP}$);
- Employment in the sector as share of total employment ($\frac{E_S}{E}$);
- Sectoral government spending as share of total government spending ($\frac{G_S}{G}$);
- Household expenditure for products/services provided within the sector as share of the average shopping basket ($\frac{C_S}{C}$).

For exemplification reasons, the sector for which the sectoral specialization index proposed in this paper will be computed is one often neglected from similar studies – the Arts, entertainment and recreation sector.

All coordinates being established, data collection stage follows. For the current application, data was collected for 31 countries and for a timeframe of 11 years (starting from 2004 and ending with 2014, the year of the last available data). Data processing for running the analysis involves the calculation of normalized values for each of the four indicators used by applying Formula 1 for the individual samples of 31 countries, by treating each year separately.

$$X_t^n = \frac{X_t - \min(X_t)}{\max(X_t) - \min(X_t)} \quad (1)$$

where X_t^n is the normalized value of the factor ($\frac{GP_S}{GDP}$, $\frac{E_S}{E}$, $\frac{G_S}{G}$ or $\frac{C_S}{C}$);
X_t is the factor value at time *t*;
$\min(X_t)$ is the minimum value of the factor at time *t* within the sample;
$\max(X_t)$ is the maximum value of the factor at time *t* within the sample.

Only at this stage, the actual method of determining weights may be applied. The method consists of running a Principal Component Analysis and *restricting the number of components to be extracted to one*. In order to grasp which approach towards defining the weights is better, the PCA was run in two different manners:

- First approach: running the analysis on all data from all 11 years combined as to benefit from the robustness given by a large sample (SPSS output is presented in Table 1). In this case, the loads of each factor in the definition of the singular component represent the weights assigned to the factors.
- Second approach: grouping data by years and running the same analysis 11 times, once for each year and then averaging the results, by use of arithmetic mean, for each factor (SPSS output is displayed in Table 2). In this case, the weights assigned to the factors are represented by the average of the results obtained for the 11 analyses.

Table 1. First approach: PCA Output run on aggregated normalized data

Component Matrix[a]

	Component
	1
GP_s / GDP	,658
E_s / E	,800
G_s / G	,370
C_s / C	,753

Extraction Method: Principal Component Analysis.
a. 1 components extracted.
Source: Author's work

Table 2. Second approach: PCA Output run on breakdown-by-year normalized data

Component Matrix[a]

	Component											*Average*
	1											
Year	*2014*	*2013*	*2012*	*2011*	*2010*	*2009*	*2008*	*2007*	*2006*	*2005*	*2004*	
GP_s/GDP	,771	,727	,772	,773	,757	,700	,672	,475	,593	,592	,692	,684
E_s/E	,823	,836	,783	,754	,712	,747	,774	,867	,853	,883	,848	,807
G_s/G	,387	,388	,159	,141	,053	,085	,190	,583	,613	,539	,700	,349
C_s/C	,654	,628	,747	,768	,791	,795	,819	,824	,777	,797	,781	,762

Extraction Method: Principal Component Analysis.
a. 1 components extracted.
Source: Author's work

First observation is that all factors appear to be having a positive load in the definition of the component, therefore confirming the initial hypothesis – that each of them is considered to have a positive influence on the degree of specialization or development of the considered economic sector.

Moreover, both approaches towards defining the weights display very similar results, thus arguing for the robustness of the analysis. The biggest difference is recorded in the weight assigned to the third factor (sectoral government spending as share of total government spending) where the second method of calculation has displayed a result by 6% lower than the first method, while assigning the first factor (gross production of the sector as share of GDP) a load with 4% higher.

Table 3. Rescaling results as to obtain final weights

	Weight prior to rescaling	Weight after rescaling
GP_s / GDP	,684	,263
E_s / E	,807	,310
G_s / G	,349	,134
C_s / C	,762	,293
Sum	*2,602*	*1,000*

Source: Author's work

As the second method is considered more reliable (the input of elements in the sample is not multiplied artificially as each year is treated as a separate entity), the results obtained by the second approach will be used as weights in the calculation of the sectoral specialization index. Furthermore, in order to facilitate comprehension of the set of weights obtained, results were rescaled as to sum up to 1, and Table 3 shows the final set of weights used in computing the index.

3. Computation of the specialization index

Once the weights are set, the sectoral specialization index proposed herein can be calculated as a weighted average of the four factors (Formula 2). It ought to be noted that the specialization index is computed using the raw data, as opposed to the Principal Component Analysis that run on normalized data.

$$Sp = \frac{\sum x_i \times p_i}{\sum p_i} \tag{2}$$

where Sp is the level of sectoral specialization;

x_i is the raw value (prior to normalization) of the factor ($\frac{GP_S}{GDP}$, $\frac{E_S}{E}$, $\frac{G_S}{G}$ or $\frac{C_S}{C}$);

p_i is the weight attributed to the factor ($\frac{GP_S}{GDP}$, $\frac{E_S}{E}$, $\frac{G_S}{G}$ or $\frac{C_S}{C}$).

In order for the results to be comparable, the last step in the computation of the index is to normalization the output, again by addressing each year separately. Therefore, the index can take values between 0 and 1, where proximity to 0 implies lack of specialization and proximity to 1 means a very high specialization degree.

Table 4. Sectoral specialization index computed for the Arts, entertainment and recreation sector (period 2004-2014)

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Austria	0,365	0,383	0,327	0,324	0,286	0,278	0,310	0,300	0,274	0,251	0,224
Belgium	0,151	0,173	0,124	0,080	0,062	0,028	0,086	0,083	0,085	0,049	0,049
Bulgaria	0,050	0,203	0,062	0,068	0,079	0,000	0,117	0,098	0,086	0,066	0,108
Croatia	0,605	0,553	0,518	0,488	0,319	0,390	0,393	0,296	0,320	0,319	0,262
Cyprus	0,474	0,541	0,479	0,496	0,413	0,408	0,476	0,477	0,472	0,431	0,367
Czech Republic	0,449	0,462	0,436	0,388	0,339	0,325	0,326	0,299	0,267	0,227	0,226
Denmark	0,518	0,530	0,470	0,430	0,373	0,381	0,379	0,369	0,349	0,325	0,316
Estonia	0,876	0,958	1,000	0,907	0,593	0,474	0,514	0,460	0,412	0,488	0,460
Finland	0,428	0,451	0,389	0,374	0,323	0,342	0,361	0,353	0,334	0,289	0,263
France	0,361	0,389	0,338	0,307	0,255	0,274	0,313	0,308	0,285	0,263	0,254
Germany	0,287	0,301	0,228	0,219	0,182	0,180	0,213	0,216	0,206	0,184	0,173
Greece	0,070	0,111	0,000	0,000	0,000	0,003	0,000	0,000	0,000	0,000	0,021
Hungary	0,412	0,442	0,367	0,324	0,282	0,249	0,333	0,312	0,293	0,272	0,270
Iceland	0,624	0,616	0,538	0,571	0,412	0,381	0,378	0,416	0,416	0,387	0,393
Ireland	0,309	0,334	0,341	0,322	0,309	0,286	0,302	0,311	0,296	0,253	0,222
Italy	0,164	0,157	0,073	0,057	0,063	0,068	0,139	0,107	0,106	0,079	0,069
Latvia	1,000	1,000	0,858	1,000	0,719	0,585	0,526	0,544	0,533	0,538	0,527
Lithuania	0,375	0,316	0,309	0,254	0,138	0,174	0,191	0,183	0,201	0,191	0,238
Luxembourg	0,112	0,228	0,069	0,089	0,069	0,035	0,067	0,071	0,045	0,037	0,041
Malta	0,400	0,520	0,540	0,866	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Netherlands	0,486	0,519	0,472	0,453	0,403	0,403	0,417	0,409	0,385	0,344	0,316
Norway	0,332	0,334	0,252	0,276	0,234	0,252	0,263	0,252	0,215	0,177	0,167
Poland	0,128	0,193	0,180	0,158	0,190	0,191	0,256	0,244	0,213	0,143	0,143
Portugal	0,164	0,202	0,125	0,068	0,061	0,027	0,081	0,067	0,053	0,009	0,000
Romania	0,000	0,000	0,007	0,004	0,041	0,052	0,162	0,164	0,149	0,120	0,178
Slovakia	0,257	0,310	0,246	0,222	0,158	0,210	0,262	0,287	0,282	0,254	0,268
Slovenia	0,487	0,528	0,494	0,446	0,365	0,330	0,403	0,331	0,311	0,238	0,259
Spain	0,519	0,554	0,530	0,507	0,429	0,386	0,419	0,393	0,317	0,288	0,301
Sweden	0,516	0,511	0,493	0,469	0,399	0,385	0,400	0,397	0,399	0,379	0,339
Switzerland	0,293	0,306	0,247	0,192	0,202	0,180	0,203	0,200	0,171	0,133	0,132
United Kingdom	0,463	0,499	0,462	0,433	0,374	0,337	0,339	0,330	0,326	0,283	0,281

Notes:
1) The horizontal bars are proportional with the values obtained for the index as compared to the entire sample of 31 countries taking into account the whole period of 11 years.
2) Background shades of green mark the differences in the evolution of the index for each country: a darker shade corresponds to higher values, while a lighter shade corresponds to lower values.

Source: Author's work, computed based on data retrieved from Eurostat (n.d.), Knoema (n.d.a), Knoema (n.d.b), INSEE (n.d.), INSSE (n.d.), NSI (n.d.), Statistics Iceland (n.d.), Statistics Norway (n.d.)

The analysis allows for observing a country's evolution across the years in terms of sectoral specialization. The trend can be regarded from two perspectives: evolution of scores or evolution in ranking. Tables 4 and 5 show the results obtained (scores and ranking) after computation of the sectoral specialization index, as defined herein. Both ways of looking at results present advantages in using them, but also require caution in interpreting the data.

Table 5. Countries' ranking considering the specialization index computed for the Arts, entertainment and recreation sector (period 2004-2014)

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Austria	17	17	18	16	16	16	17	16	18	17	19
Belgium	26	28	26	26	28	28	28	28	28	28	28
Bulgaria	30	25	29	28	25	31	27	27	27	27	26
Croatia	4	5	6	7	14	6	9	18	11	9	14
Cyprus	10	6	9	6	5	4	4	3	3	4	5
Czech Republic	12	13	13	13	12	14	15	17	19	19	18
Denmark	6	7	11	12	10	10	10	9	8	8	8
Estonia	2	2	1	2	3	3	3	4	5	3	3
Finland	13	14	14	14	13	11	12	10	9	10	13
France	18	16	17	18	18	17	16	15	16	14	16
Germany	22	23	23	22	22	23	22	22	22	21	22
Greece	29	30	31	31	31	30	31	31	31	31	30
Hungary	14	15	15	15	17	19	14	13	15	13	11
Iceland	3	3	4	4	6	9	11	5	4	5	4
Ireland	20	18	16	17	15	15	18	14	14	16	20
Italy	24	29	27	29	27	25	26	26	26	26	27
Latvia	1	1	2	1	2	2	2	2	2	2	2
Lithuania	16	20	19	20	24	24	24	24	23	20	17
Luxembourg	28	24	28	25	26	27	30	29	30	29	29
Malta	15	9	3	3	1	1	1	1	1	1	1
Netherlands	9	10	10	9	7	5	6	6	7	7	7
Norway	19	19	20	19	19	18	19	20	20	22	23
Poland	27	27	24	24	21	21	21	21	21	23	24
Portugal	25	26	25	27	29	29	29	30	29	30	31
Romania	31	31	30	30	30	26	25	25	25	25	21
Slovakia	23	21	22	21	23	20	20	19	17	15	12
Slovenia	8	8	7	10	11	13	7	11	13	18	15
Spain	5	4	5	5	4	7	5	8	12	11	9
Sweden	7	11	8	8	8	8	8	7	6	6	6
Switzerland	21	22	21	23	20	22	23	23	24	24	25
United Kingdom	11	12	12	11	9	12	13	12	10	12	10

Note: Background colors highlight the position in the ranking, in relation to the entire sample of 31 countries taking into account the whole period of 11 years. Green shades correspond to upper positions of the ranking, while red shades mark lower positions.
Source: Author's work, computed based on data from Table 4

Each year, a country's scores are calculated relative to the other countries' individual performances; therefore a positive evolution of scores does not necessarily imply an increase in specialization (it might be just due to a decrease in other countries' performances). But, in the case where the competitive context remains broadly unchanged, such an approach offers a more contoured overview of the evolution. By focusing the attention on the ranking evolution, one might be tricked into thinking a country registers a striking increase in specialization, but if it is the case of outrunning countries with very close specialization levels, then an increase of less than 1% may generate such an outcome. However, outranking a long established country that consistently displayed a high degree of specialization may constitute a notable performance and looking at the ranking can become a good indicative.

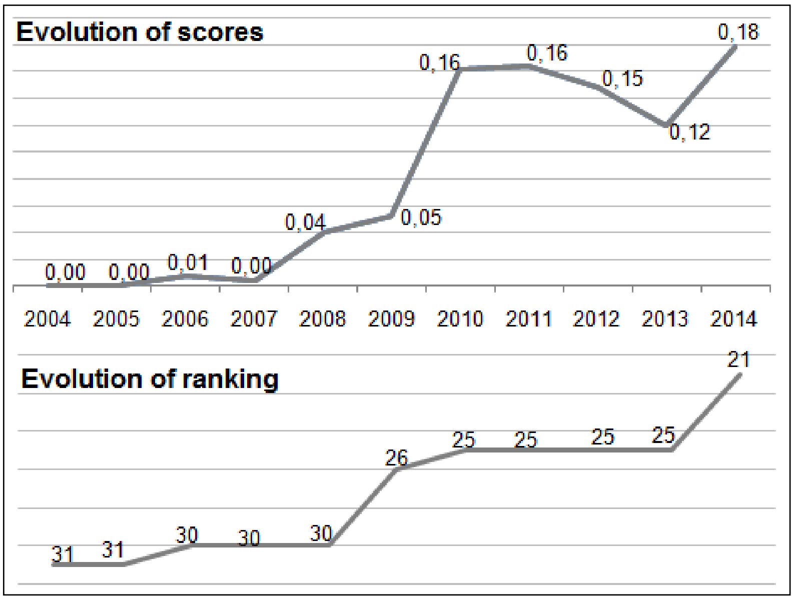

Figure 1. Romanian Arts, entertainment and recreation sector's specialization evolution (period 2004-2014)

Source: Author's work

For illustration purposes, a zoom-in on the results is depicted in Figure 1, where is presented Romania's evolution regarding the specialization of the Arts, entertainment and recreation sector. Both indicators' evolution converges to the same overall conclusion – that over time Romania's position has strengthened in European context by means of an increase in relative specialization. However, if one were to look only at the evolution of scores, the 2011-2013 scores might be misleading, as although the score has dropped, Romania maintained its position in the ranking due to the concomitant decrease in other countries' performance.

4. Concluding remarks

Although initially designed as a method for reducing large sets of factors into a more manageable number of components, the Principal Component Analysis is used herein as means of reducing a rather small number of factors into just one single component with the aim of using the component loads as weights for developing a composite index.

By following the steps described, this method may be applied to developing other indices (it is not bound to working only for specialization measurements). Furthermore, although in the present paper the specialization index was used for quantifying local recreation specialization, its application may be generalized to other economic sectors.

The main limitation of the study lies in the scarcity of data collected for the analysis. Although data was gathered generally from a single source (Eurostat), some additional figures were retrieved from other various online sources (data provided by national statistics institutes, online databases) and further data processing techniques as to obtain homogeneity were then applied. Even though the time period considered in the analysis is quite extended, there were several gaps in the data that had to be filled in by estimations. Moreover, sample size is rather small and all elements are concentrated in just one geographic region.

Because it is a heterogeneity index (implying that a country's results are obtained as a result of the structure of the reference group of countries), interpreting output ought to be carefully considered because both score evolution and ranking evolution can be misleading and may cause drawing biased conclusions. Therefore, competitive context should always be a concern in interpreting results.

References

1. Čutková, Z. and Donoval, M. **Sectoral specialization in the SR,** BIATEC, Vol. 10, No. 12, 2004, 5–7
2. European Central Bank, **Sectoral specialization in the EU,** A macroeconomic perspective, Occasional Paper Series, Vol. 19, 2004
3. Eurostat (n.d.) **Eurostat Database** [online], Accessed April 10, 2016, from http://ec.europa.eu/eurostat/data/database
4. Ioncica, M., Draghici, M., Petrescu, C. and Ioncica, D. **Services specialization (a possible index) and its connection with competitiveness: the case of Romania,** The Service Industries Journal, Vol. 30, No. 12, 2010, pp. 2023-2044
5. Knoema (n.d.a) **Croatia - Recreation and culture: share of household expenditure (COICOP 09) (%)** [online], Accessed April 10, 2016, from https://knoema.com /cpc_ecnacoi/candidate-countries-and-potential-candidates-annual-national-accounts-breakdown-of-final-consumption?tsId=1000090
6. Knoema (n.d.b) **Final consumption expenditure of households** [online], Accessed April 10, 2016, from https://knoema.com/SNA_TABLE5_2015/final-consumption-expenditure-of-households?country=1000300-switzerland
7. National Institute of Statistics and Economic Studies of France – INSEE (n.d.) **Quarterly payroll employment by sector - Main groups** [online], Accessed April 10, 2016, from http://www.bdm.insee.fr/bdm2/choixCriteres?codeGroupe=1185
8. National Institute of Statistics of Romania – INSSE (n.d.) **Tempo Online** [online], Accessed April 10, 2016, from http://statistici.insse.ro/shop/?lang=en
9. OECD **OECD science, technology and industry scoreboard 2013. Innovation for growth,** 2013 [pdf]. Accessed April 10, 2016, from http://www.oecd.org/sti/scoreboard-2013.pdf
10. Palan, N. **Measurement of Specialization – The Choice of Indices,** FIW - Working Paper, No. 62, December 2010
11. Republic of Bulgaria National Statistical Institute – NSI (n.d.) **Sector General Government** [online], Accessed April 10, 2016, from http://www.nsi.bg/en/content/5267/sector-general-government
12. Sava, A.M. **Correlating Local Recreation Specialization to Prosperity: Study on European Union Countries,** Entrepreneurship, Business and Economics, Vol. 2, 2016, pp. 319-330
13. Statistics Iceland (n.d.) **Economy** [online], Accessed April 10, 2016, from http://px.hagstofa.is/pxen/pxweb/en/Efnahagur/Efnahagur__fjaropinber__fjarmal_opinber__fjarmal_opinber/?rxid=a5ce19c0-4600-4c6c-9dce-8f726e07869a
14. Statistics Norway (n.d.) **Survey of consumer expenditure** [online], Accessed April 10, 2016, from https://www.ssb.no/statistikkbanken/SelectVarVal/Define.asp?MainTable=UtgHusType&KortNavnWeb=fbu&PLanguage=1&checked=true