Theoretical and Applied Economics
Volume XXVII (2020), No. 2(623), Summer, pp. 131-146

An approach to the index of economic freedom using the data mining technique in the economic environment context

Ionuț NICA
Bucharest University of Economic Studies, Romania
ionut.nica@csie.ase.ro

Abstract. *Economics, at the macroeconomic level, can be viewed as a complex dynamic system consisting of a multitude of subsystems linked together by direct and indirect connections of different types. This dynamic system forms an economic cybernetic network composed from these subsystems, which evolves according to the economic environment and its own objectives, which are conditioned by the achievement of a general common objective. One of the most important indices that could characterize the economic environment for all countries is the Economic Freedom Index and it is very important to be taken into account by companies that want to be active in international market. In this article we analyzed the Index of Economic Freedom, divided into four subcategories: the rule of law, the size of the government, the efficiency of the regulations and the opening of the market. The analysis was performed on a set of data extracted for the year 2019 corresponding to 174 countries. Also, with the R software solution, I will apply multidimensional data analysis techniques to perform the proposed research.*

Keywords: economic freedom, R, data mining, cybernetics approach, dynamic system, classification.

JEL Classification: N1, N4, A1, C1.

1. Introduction

The scientific knowledge in any field of human activity, implies, regardless of the nature and the specific of the concrete objectives pursued, a complex and rigorous quantitative analysis of the phenomena and processes that are the object of the research. It is visible to everyone, and more and more, that in the modern era almost every individual engaged in a human activity deals, in one way or another, directly or indirectly, with data and information, with collecting, their processing and interpretation.

Although data analysis methods and techniques are used in most areas of human activity, we can say that data analysis has the largest use in the economic and social field, and the efficiency of its use in this field has an extremely high efficiency.

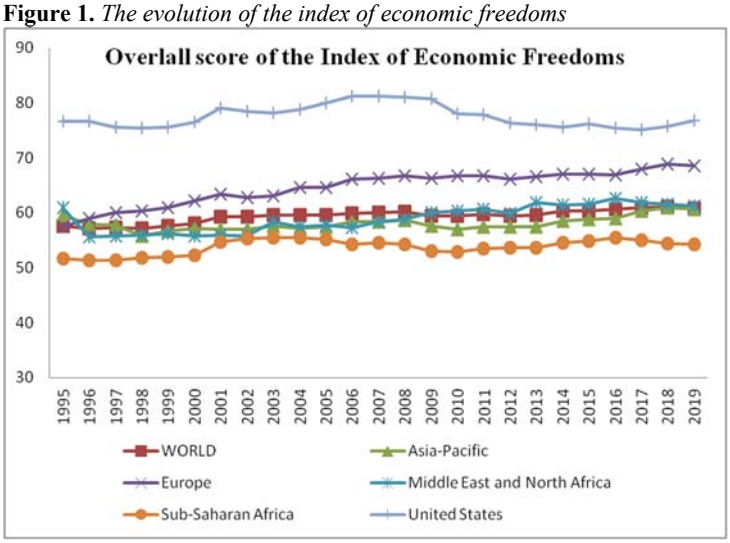
The Index of Economic Freedom is an annual index and ranking created in 1995 by The Heritage Foundation and The Wall Street Journal to measure the degree of economic freedom in the countries of the world. The creators of the index have adopted an approach similar to that of Adam Smith in The Wealth of Nations, according to which the basic institutions that protect the freedom of individuals to pursue their own economic interests lead to greater prosperity for society at large. (Ott, 2016)

The Index of Economic Freedom and it is very important to be taken into account by companies that want to be active in international market. If the value of this index is small, we can say that the system of governance is corrupt and there are a lot of factors that stand in the way of trade. If the value of the index is very high, companies can safely invest in the respective economy. (Musetescu, 2011)

The analysis was based on the corresponding data of 174 countries, extracted for the year 2019. The data source can be found at the following web page: https://www.heritage.org/index/download.

The 2019 ranking points to aspects of economic freedom between 0 and 100, with 0 meaning “maximum repression” and 100 meaning “total economic freedom”. The Economic Freedom Index is calculated as an arithmetic mean between its twelve components, divided into four categories, as we mentioned earlier, in total 13 quantitative variables. As qualitative variables, the names of the analyzed countries were used, as well as the regions in which they are found. They bear the names: “United States”, “Asia-Pacific”, “Europe”, “Middle East and North Africa” and “Sub-Saharan Africa”.

Figure 1. *The evolution of the index of economic freedoms*



Source: The data were extracted from the https://www.heritage.org and processed by the authors.

In the cybernetic approach of the economy at the macroeconomic level we must start, first of all, from the knowledge of the subsystems and the way in which they interact in the process of achieving their own and the general objectives. Thus, households, firms, commercial banks, different public institutions can be analyzed as separate (individual) systems at the microeconomic level, but they form at the macroeconomic level sectors (systems) that have emerging properties and objectives, thus deriving from the simultaneous operation of the multitude of systems of the same type at the microeconomic level. (Scarlat and Chirita, 2019)

Economic freedom is the essential condition of development, as demonstrated by classical and neoclassical economists (A. Smith, J.B.B., Mises, F.A. Hayek) in their research. Therefore, the macroeconomic system is influenced by this economic freedom, contributing to its well-being.

The components of the Heritage Foundation Economic Freedom are based at the four pillar: rule of law, limited government, regulatory efficiency and open markets. The first pillar is composed of two sub-components: property rights freedom and freedom from corruption. The second is based also to two sub-components: fiscal freedom and government spending. The next three sub-components, business freedom, labor freedom and monetary freedom constituted the base of the pillar three. Trade freedom, investment freedom and financial freedom are the sub-components of the last pillar. (Naanwaab, 2013)

According to the information of the heritage.org, the countries with the greatest economic freedom on the world map are Hong Kong, Singapore, New Zealand, Switzerland, Australia, Ireland, United Kingdom, Canada, United Arab Emirates and Taiwan. The country without ranked are Iraq, Libya and Liechtenstein (https://www.heritage.org/ index/ranking). It is easy to see that there is a direct link between economic freedom and the prosperity of a country. The Heritage Foundation report and the *Wall Street Journal* prove that poverty is not a fatality, that wealth is not a miracle and is not reserved for the privileged.

2. Theoretical background and mathematical model about data mining techniques

The main purpose of the analysis of the principal components is to reduce the size of the data set. Principal component analysis is a multidimensional analysis technique which, by constructing linear combinations of initial variables $\boldsymbol{x_i}$, allows the information contained in them to be rewritten in an identical number of other variables $\boldsymbol{z_i}$, called principal components, with variance distributed decreasing and uncorrelated between them.

Mathematically, if we consider a vector of n original variables $(x_1, x_2, x_3, \dots x_n)$, the analysis of the principal component consists in the identification of a variance maximizing linear transformation by applying it to a vector of p principal components $(z_1, z_2, z_3, \dots z_p)'$ $(p < n)$ which retains a satisfyingly large amount of initial variability. The principal components can be written: (Maer-Matei et al., 2019)

$$z_1 = \alpha_{11}x_1 + \alpha_{21}x_2 + \cdots + \alpha_{n1}x_n$$

$$z_2 = \alpha_{12}x_1 + \alpha_{22}x_2 + \cdots + \alpha_{n2}x_n$$

$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$

$$z_n = \alpha_{1n}x_1 + \alpha_{2n}x_2 + \cdots + \alpha_{nn}x_n$$

In the matrix form, this relationship can be transposed into:

$$\begin{pmatrix} z_1 \\ z_2 \\ . \\ . \\ . \\ . \\ z_n \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{21} \dots . \alpha_{n1} \\ \alpha_{12} & \alpha_{22} \dots . \alpha_{n} \\ & . \\ & . \\ & . \\ & . \\ \alpha_{1n} & \alpha_{2n} \dots . \alpha_{nn} \end{pmatrix} \begin{pmatrix} x \\ x_2 \\ . \\ . \\ . \\ . \\ x_n \end{pmatrix}$$

Where

$$z = \begin{pmatrix} z_1 \\ z_2 \\ . \\ . \\ . \\ . \\ z_n \end{pmatrix}, \text{x} = \begin{pmatrix} x \\ x_2 \\ . \\ . \\ . \\ . \\ x_n \end{pmatrix}, A' = \begin{pmatrix} \alpha_{11} & \alpha_{21} \dots . \alpha_{n1} \\ \alpha_{12} & \alpha_{22} \dots . \alpha_{n} \\ & . \\ & . \\ & . \\ & . \\ \alpha_{1n} & \alpha_{2n} \dots . \alpha_{nn} \end{pmatrix}$$

2.1 . The mathematical model of the principle components

The identification of the principle components starts from the formulation of the problem of maximizing their variance.

If the covariance matrix of initial variables is $\sum$, then variance of the principle components, then variance of the principle components z_i can be written:

$$var(z_i) = var\left(\propto^{(i)'} x\right) = \propto^{(i)'} \sum \propto^{(i)}$$

Where $\propto^{(i)} = \begin{pmatrix} a_{1i} \\ \cdots \\ a_{ni} \end{pmatrix}$ is the i-vector of the column of matrix A.

Therefore, for construction z_i, it is necessary to identify the vector $\propto (i)$ which ensures a maximum level of variability.

Although, traditionally, PCA(1) is applied to the covariance matrix of the initial variables, its application is also common to the matrix of the correlation coefficients between them. Opting for the latter is justified by the existence of variables with different units of measurement, which affect the covariance and variance, and the obtained components would be meaningless, but not the correlation coefficients. (Morrison, D.S., 1976)

The choice of the number of principal components needed to effectively synthesize the set of variables, under minimal information loss, can be made according to several criteria (Rencher, 2002):

- Criterion of the quantity of information – the choice of a small number of variables to ensure a good representation of the initial variables provided that the percentage of the total variance retained by them is large enough. The choice of the threshold over which the conserved variance is considered satisfactory is at the discretion of the one performing the analysis, an amount of 70-75% of the information being sometimes sufficient, at other times the loss of 25-30% of the variability being considered too high.
- Kaiser's criterion – keeping those principal components whose variances are greater than the average. In the case of using standardized data or applying the PCA on the correlation matrix, the level with which the eigenvalues will be compared is 1 (the average level of variance).
- Granulosity criterion (Screeplot) – this criterion involves analyzing the graph constructed based on the eigenvalues of the variance-covariance matrix Σ and identifying a turning point. For example, if up to the eigenvalue noted with λ3 the graph had a slow downward slope, but between λ3 and λ4 the descent is steep and followed by a relatively constant evolution between the following eigenvalues, the number of principal components selected will be equal to 3. The explanation for this choice is that the information gain brought from component 4 is very small, almost insignificant, compared to the variability retained by the first 3 components.

2.2 . Mathematical model of factor analysis

The factorial analysis aims to identify and evaluate those factors of unobservable nature that are the basis of the common evolution of the variables and that generate the manifestations of the correlations between them. These latent and directly measurable factors are intrinsically found in the data structure, and their existence is suggested by the structure of the relationships between the variables. It is a method often used in the socio-economic and behavioral sciences, as it allows to highlight certain preferences, traits, behavioral factors that underlie the evolution of the studied phenomena.

The factorial analysis model expresses each variable as a linear combination of common and specific factors. In addition to these two categories of factors, in the mathematical

writing of the model there is also a residual component (the errors) which overcomes accidental factors, measurement errors, being insignificant in explaining the variables.

The model can be written as follows:

$$x_1 - \mu_1 = a_{11}f_1 + a_{12}f_2 + \cdots + a_{1p}f_p + u_1 + \varepsilon_1$$

$$x_2 - \mu_2 = a_{21}f_1 + a_{22}f_2 + \cdots + a_{2p}f_p + u_2 + \varepsilon_2$$

……..

$$x_n - \mu_n = a_{n1}f_1 + a_{n2}f_2 + \cdots + a_{np}f_p + u_n + \varepsilon_n$$

Where:
μ_i is the mean of the variable i;
f_i is the commune factor;
u_i represent the uniqueness element;
ε_i represent the residual element;
$p < n$;
$the\ coefficient\ a_{ni}\ represent\ the\ intensity\ of\ the\ factor.$

Given the low importance of the residual component, given its random nature, many authors ignore the errors or consider them part of the uniqueness and thus eliminate this component from the model writing.

The assumptions of the factorial model are (Rencher, 2002):
- The average factors are 0, variance equal to 1 and the factors are uncorrelated it means that their covariance matrix is therefore the identity matrix.
- The unique factors are, in turn, of average 0, uncorrelated between them but of variance different from 1.
- The common factors and the unique factors are not correlated.

There are a multitude of criteria used in determining the number of common factors, their efficiency being the subject of a high number of studies over time.

A. The Bartlett test – this involves testing the hypothesis that all eigenvalues are equal. When the null hypothesis can no longer be rejected from a number of factors, that is the number of main components retained.

B. The minimum of partial averages (minimum partial average – MAP) – is a method that involves calculating the partial correlations between the variables, conditioned by the common factors. The squares of the elements on either side of the diagonal are summed, and the number of common factors will be the one for which the minimum of the squares of the partial correlations calculated after eliminating the influence of the common factors is recorded.

The contingency table is the primary input of this analysis. The main purpose of the correspondence analysis is to provide an intuitive graphical representation, most often bi- or three-dimensional, that clearly shows the correspondences between the categories of the variable in the column and that of the rows. Therefore, the objective is to reduce the number

of axes from min {(p-1),(q-1)} to a smaller number of dimensions allowing easy identification of associations between variables.

Greenacre (2007) formulates the problem of correspondence analysis as identifying a small subspace S that contains the centroid and against which the sum of distances (measured as hi-squared distances from centroid) is minimal.

3. Description of the variables used in the analysis

The Economic Freedom Index is calculated as an arithmetic mean between its twelve components, divided into four categories, as we mentioned earlier, in total 13 quantitative variables. As qualitative variables, the names of the analyzed countries were used, as well as the regions in which they are found. They bear the names: "United States", "Asia-Pacific", "Europe", "Middle East and North Africa" and "Sub-Saharan Africa". The quantitative variables have the following description:

A. The rule of the law: properties right freedom and freedom from corruption

The degree of legal protection of a country of private property rights and the degree of application of these laws. It is divided into the following sub-factors: physical property rights; intellectual property rights; the power of investor protection; the risk of expropriation; quality of land management. The degree of efficiency and fairness of the judicial system, especially as regards property laws. It is divided into the following sub-factors: judicial independence; the quality of the judicial process; the probability of obtaining favorable judicial decisions.

The integrity of the government analyzes how widespread forms of political corruption and practices such as bribery, extortion, nepotism, cronyism, patronage, embezzlement and graft are involved. It is divided into the following sub-factors: public confidence in politicians; bribes and unusual payments; transparency in the elaboration of government policies; absence of corruption; perception of corruption; government and public transparency.

B. Limited government: fiscal freedom and government spending

The fiscal pressure analyzes the marginal tax rates on personal and corporate income, respectively the general level of taxation (including direct and indirect taxes imposed by all levels of government) as a percentage of GDP. Its sub-factors are: the highest marginal tax rate on individual incomes; the highest marginal tax rate on corporate income; total fiscal pressure as a percentage of GDP. Government spending quantifies the pressure of government spending, including government consumption and all payments for social programs. The ideal level varies from country to country, but zero-level spending is used as a benchmark. The fiscal health analyzes how well a country manages its budget, quantifying debt and growing deficit. It is divided into the following sub-factors: average deficits as a percentage of GDP for the last three years (80% of the score); debts as a percentage of GDP (20% of the score).

C. Regulatory efficiency: business freedom, labor freedom and monetary freedom

Business freedom analyzes the cost, time and freedom to open, operate and close a business, taking into account factors such as electricity. It is divided into thirteen sub-factors: starting a business – procedures (number); starting a business – time (days); starting a business – cost (% of income per capita); starting a business – minimum capital (% of income per capita); obtaining a license – procedures (number); obtaining a license – time (days); obtaining a license – cost (% of income per capita); closing a business – time (years); closing a business – cost (% of ownership); closing a business – recovery rate (cents in dollars); obtaining electricity – procedures (number); obtaining electricity – time (days); obtaining electricity – cost (% of income per capita).

Labor freedom quantifies the intrusiveness of labor rights, such as the minimum wage, laws that inhibit redundancies, compensation requirements and measurable regulatory restrictions for employment and hours worked, plus the labor force participation rate as an indicative measure of employment opportunities in the labor market. It is divided into the following sub-factors: the ratio between the minimum wage and the average value added per worker; obstacle to hiring additional workers; the rigidity of the hours; difficulty dismissing non-performing employees; period of mandatory legal notice; compulsory payment for dismissal; the labor force participation rate.

Monetary freedom analyzes how stable prices are and how much microeconomics is involved. It is divided into the following sub-factors: the weighted average inflation rate for the last three years; price control.

D. Open market: trade freedom, investment freedom and financial freedom

Trade freedom quantifies the extent to which tariff and non-tariff barriers affect imports and exports of goods and services into and from the country. Its sub-factors are: the weighted average tariff rate; non-tariff barriers (NTB).

Investment freedom analyzes how free or restricted the investment capital flow of individuals and companies.

Financial freedom indicates bank efficiency, as well as how independent the financial sector is from the government. This aspect concerns five broad areas: the measure of government regulation of financial services; the degree of state intervention in banks and other financial firms through direct and indirect ownership; the influence of the government on the allocation of credit; the extent of financial and capital market development; opening to foreign competition.

4. Data mining analysis, results and discussions

First, we will analyze each numerical variable from the point of view of descriptive statistics, such as mean, median, standard deviation, coefficient of variation, quartiles, but also distribution, with the help of asymmetry and vault coefficients. It was interesting to note that based on the heterogeneous distribution and existence of high value outliers, we can conclude that there are countries where the factors that determine the integrity of the

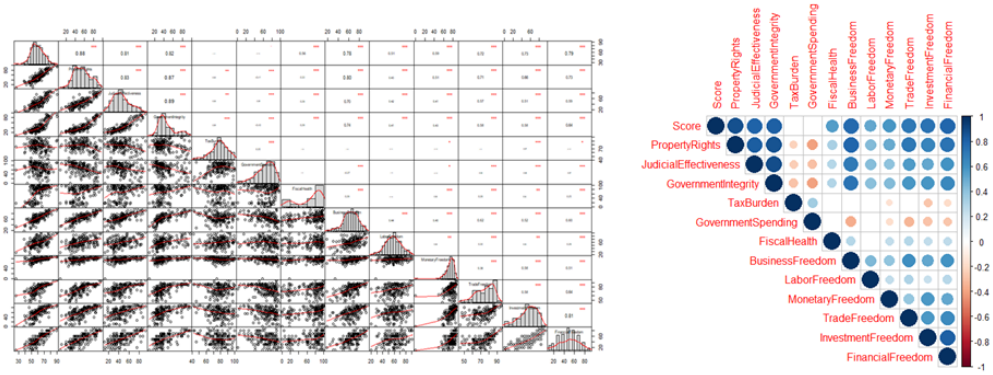
government are at a very high level compared to the rest of the countries, so these countries have a high level of unusually low corruption. We observe by the presence of some outliers of small values, so that the fiscal pressure in certain countries is a high one, a fact that is also reflected in a significant decrease of the Economic Freedom Index. The histogram of the Business Freedom variable confirms the previous analysis, and the boxplot shows low aberrant values, which confirms that some countries are dealing with restrictive policies regarding the business environment.

Of all the variables analyzed, the values of freedom in terms of work are closer to a normal distribution. However, according to the boxplot, there are countries that are well below the usual level in this chapter (low value outlays). The histogram of the variable Monetary Freedom shows that prices have a very high stability in most countries, but according to the boxplot, we also have a significant number of countries suffering from high instability in this regard.

4.1. Principal component analysis

The analysis of the principal components is a technique of multidimensional analysis of the data sets and that aims to summarize the information. This operation uses an orthogonal transformation to convert a set of observations of correlated variables into a set of values of linearly uncorrelated variables, called principal components. This transformation is performed in such a way that the first principal component has the largest possible variance, representing as much data variability as possible. The obtained vectors are non-correlated orthogonal bases. The first stage in the analysis of the principle components is the investigation of the correlation matrix. Thus, one can find information redundancies and apply the analysis of the principle components.

Figure 2. *The matrix of correlation (graphic representation in R)*



Source: Authors' own research results.

A positive correlation between two variables results in the fact that when the value of one variable increases, then the other variable will have a higher value. Negative correlation implies a decrease in the value of one variable as the other increases. The intensity of the connection is measured on a scale from 0 to 1 (-1, in case of negative correlation), 0 being a total lack of correlation between indicators, and 1 (respectively -1) a total correlation (identical indicators).

For example, since we have a strong positive correlation of 0.79 between the Score and Financial Freedom variables, we can say that when increasing by one point the financial freedom, the final score will increase by 0.79 points. Similarly, there is a negative correlation of -0.40 between government spending and government integrity. Thus, as government spending increases by one point, government integrity will decrease by 0.40 points.

Further we will determine if the coefficients determined are statistically significant by comparison with the significance threshold 0.05.

Figure 3. *The matrix of probabilities associated with the correlation coefficients*

P

	Score	PropertyRights	JudicialEffectiveness	GovernmentIntegrity	TaxBurden	GovernmentSpending	FiscalHealth
Score		0.0000	0.0000	0.0000	0.3257	0.0648	0.0000
PropertyRights	0.0000		0.0000	0.0000	0.0031	0.0000	0.0000
JudicialEffectiveness	0.0000	0.0000		0.0000	0.0011	0.0001	0.0001
GovernmentIntegrity	0.0000	0.0000	0.0000		0.0001	0.0000	0.0000
TaxBurden	0.3257	0.0031	0.0011	0.0001		0.0000	0.2005
GovernmentSpending	0.0648	0.0000	0.0001	0.0000	0.0000		0.5412
FiscalHealth	0.0000	0.0000	0.0001	0.0000	0.2005	0.5412	
BusinessFreedom	0.0000	0.0000	0.0000	0.0000	0.2685	0.0000	0.0001
LaborFreedom	0.0000	0.0000	0.0000	0.0000	0.1432	0.4174	0.1715
MonetaryFreedom	0.0000	0.0000	0.0000	0.0000	0.0450	0.0205	0.0010
TradeFreedom	0.0000	0.0000	0.0000	0.0000	0.4905	0.0000	0.0001
InvestmentFreedom	0.0000	0.0000	0.0000	0.0000	0.0003	0.0001	0.0012
FinancialFreedom	0.0000	0.0000	0.0000	0.0000	0.0175	0.0003	0.0007

	BusinessFreedom	LaborFreedom	MonetaryFreedom	TradeFreedom	InvestmentFreedom	FinancialFreedom
Score	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PropertyRights	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
JudicialEffectiveness	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GovernmentIntegrity	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
TaxBurden	0.2685	0.1432	0.0450	0.4905	0.0003	0.0175
GovernmentSpending	0.0000	0.4174	0.0205	0.0000	0.0001	0.0003
FiscalHealth	0.0001	0.1715	0.0010	0.0001	0.0012	0.0007
BusinessFreedom		0.0000	0.0000	0.0000	0.0000	0.0000
LaborFreedom	0.0000		0.0011	0.0000	0.0028	0.0002
MonetaryFreedom	0.0000	0.0011		0.0000	0.0000	0.0000
TradeFreedom	0.0000	0.0000	0.0000		0.0000	0.0000
InvestmentFreedom	0.0000	0.0028	0.0000	0.0000		0.0000
FinancialFreedom	0.0000	0.0002	0.0000	0.0000	0.0000	

Source: Authors' own research results.

It can be observed that a correlation between fiscal pressure and the Economic Freedom Index is not significant, since the associated P-value is 0.32 > 0.05. Of course, this is just an exception in the dataset. Thus, the use of principal component analysis is fully justified.

Own values aim to indicate how the information taken down decreases with each component. Thus, the first component takes the largest amount of information, the second less than the first and so on. The *princomp* function also highlights this aspect.

Figure 4. *Output after applying the principal function*

```
> acp <- princomp(dateNumerice, cor = TRUE)
> summary(acp)
Importance of components:
                          Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6    Comp.7
Standard deviation     2.6186727 1.1806256 1.0390414 0.94094578 0.84757806 0.76400897 0.69119731
Proportion of Variance 0.5274959 0.1072213 0.0830467 0.06810607 0.05526066 0.04490075 0.03675029
Cumulative Proportion  0.5274959 0.6347172 0.7177639 0.78586995 0.84113061 0.88603136 0.92278165
                          Comp.8     Comp.9    Comp.10    Comp.11    Comp.12       Comp.13
Standard deviation     0.5879691 0.52816452 0.41335714 0.340724301 0.303660683 2.486199e-03
Proportion of Variance 0.0265929 0.02145829 0.01314339 0.008930235 0.007093062 4.754760e-07
Cumulative Proportion  0.9493745 0.97083283 0.98397623 0.992906462 0.999999525 1.000000e+00
```

Source: Authors' own research results.

Because we worked with standardized variables, we will keep in the analysis only those components that have the dispersion value greater than 1. From the previous figure we can notice the dispersion of 2.61 of the first component, 1.18 of the second component and 1.03

of the second third component. It is noted in figure above that the first component takes 52% of the information, the first two components take 63%, and the first three 71%. This is yet another proof that the component elements are getting less and less information as their number grows. Thus, we will have three principal components, according to the coverage percentage criterion. Further, we will highlight the utility of our own vectors for calculating the scores of the first principal component.

The formula for scores of the first principal component in this analysis is as follows: $W_1 = -0.36x_1 - 0.35x_2 - 0.32x_3 - 0.33x_4 + 0.09x_5 + 0.15x_6 -$

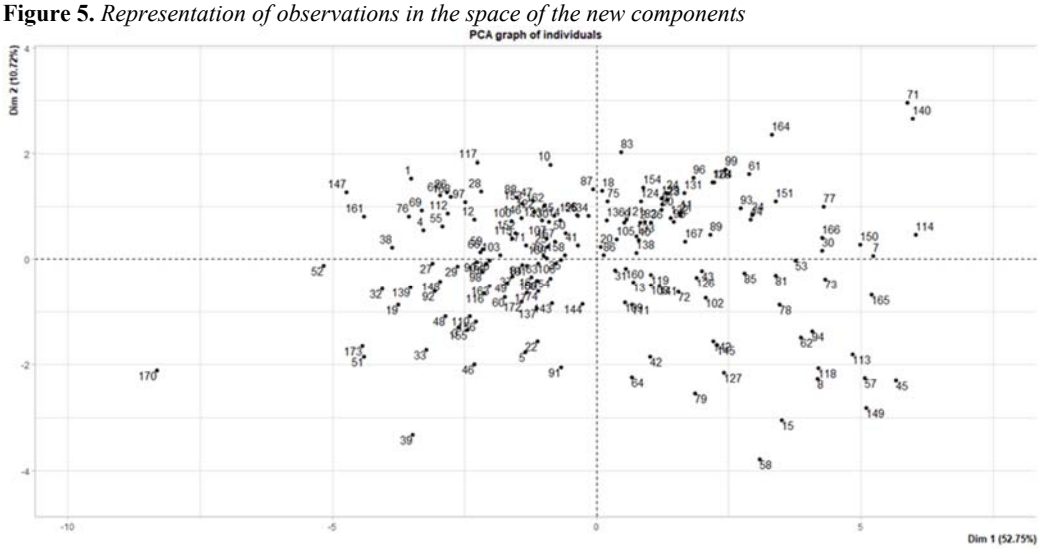
$-0.15x_7 - 0.31x_8 - 0.18x_9 - 0.23x_{10} - 0.29x_{11} - 0.29x_{12} - 0.31x_{13}$

The circle of correlations is used to graphically represent the links between the initial variables and the principal components kept in the analysis. The closer a variable is to the edge of the circle, the stronger the correlation with the respective component.

Thus, we can draw the following conclusions regarding the correlation of the principal components:

- Component 1: strong direct links to Score (0.95), Property Rights (0.93), Judicial Effectiveness (0.85) and Government Integrity (0.88).
- Component 2: positive correlations with Tax Burden (0.75) and Government Spending (0.66).
- Component 3: positive correlation with Fiscal Health (0.57) and negative with Labor Freedom (-0.45).

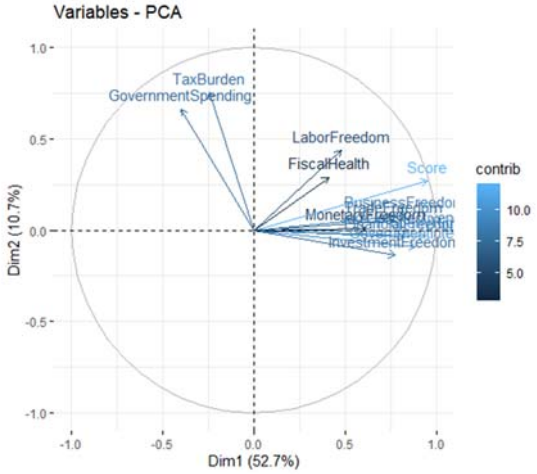
Figure 5. *Representation of observations in the space of the new components*



Source: The graphics was processed by the authors.

The reduction of the dimensionality of the data allowed a visualization of the 174 countries in a two-dimensional graph and the identification of the similarities and differences existing between them from the point of view of the variables that compose the Index of Economic Freedom.

Figure 6. *Biplot – the contribution of the original variables to the main components and the amount of information taken from them*



Source: The graphics was processed by the authors.

The figure above shows that the first main component takes 52.7% and the second 10.7% of the information. Thus, the first two main components take up 63.4% of the information, and together with the third, 71% of the information. Thus, the remaining 29% of the information contained in the other 10 components will be lost.

4.2. Factor analysis

Factor analysis is a multivariate analysis technique used to describe the variability between observed and correlated variables through a small number of uncorrelated variables, called factors.

Using the Bartlett sphericity test and calculating the KMO index is a crucial phase in factor analysis.

It has as a basic principle the comparison between the correlation matrix and the unit matrix in order to see if the identification of common factors has statistical justification.

- *The null hypothesis* H_0: The variables are orthogonal, so the creation of common factors is not justified.
- *Alternative Hypothesis* H_1: The analysis finds at least one common factor, so we can construct common factors in a justified way.

P-value will have to be below the significance threshold of 0.05

```
$chisq
[1] 5886.149

$p.value
[1] 0

$df
[1] 78
```

We obtain the values chisq = 5886.149, P-value = 0 and a number of degrees of freedom of 78. P-value is below the threshold of 0.05, being even 0, so an excellent value, for which we can go further with the analysis.

The KMO index shows the proportion of the variance of the variables that can be caused by latent factors, identifying variables that are poorly correlated with the others to decide if it makes sense to identify common factors.

For the highest utility of the factor analysis, the value of the KMO index should be as close to 1. The degrees of factorability following the application of the KMO index are the following: 0.90-1.00 – very good workability; 0.80-1.89 – good workability; 0.70-0.79 – average factoriality; 0.60-0.69 – mediocre factoriality; 0.50-0.59 – poor factorability; 0.00-0.49 – no factors.

```
> teste$KMO
[1] 0.20149
```

The value obtained for the KMO index is 0.20, resulting from the fact that there is no factorability in the present case.

Although the Bartlett test showed good value, the KMO index did not indicate a satisfactory value in this regard. Thus, we cannot continue the factorial analysis.

4.3. Correspondence analysis

After researching the correlation matrix, we can see that we have a positive and strong correlation of 0.83 between the Property Rights and Judicial Effectiveness variables, so we will choose these two variables in the correspondence analysis.

Correspondence analysis is a multivariate statistical technique, similar to the analysis of the main components, but applied on categorical variables and not continuous, as in the case of the PCA. Like the PCA, it offers a way of presenting a data set in the form of a two-dimensional graph.

Correspondence analysis is based on the concept of inertia, namely the measurement of the variable in the contingency table using the Chisq test.

The Property Rights variable was named PR, and the Judicial Effectiveness variable was called JE, both having three classes: Low, Moderate, High.

After grouping the data, we were able to obtain the contingency table associated with the two variables in the following table.

Table 1. *Contingency table associated with the variables PR and JE*

PR\JE	Low	Moderate	High
Low	101	1	20
Moderate	15	20	0
High	5	0	11

Source: Authors' own research results.

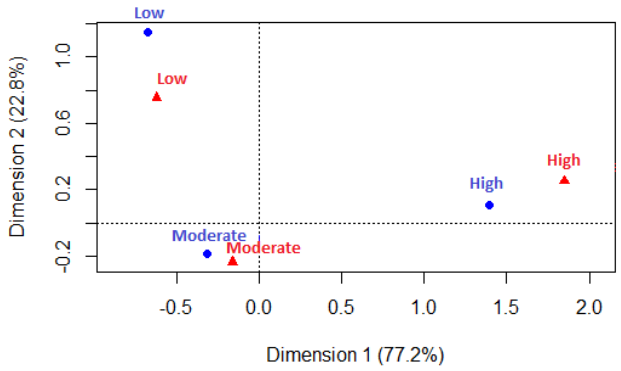
In order to determine the association between the two categorical variables, we will apply the Chisq test, after which the P-value must be lower than the significance threshold 0.05.

```
        Pearson's Chi-squared test

data:  table(dateac[, 4], dateac[, 5])
X-squared = 113, df = 4, p-value <2e-16
```

Thus, we determined the value of $X^2 = 113$, the number of degrees of freedom being 4, and the P-value is of the order $2 * 10^{-16}$, a value well below the threshold of 0.05. Therefore, the link between the two categorical variables is statistically significant and we can estimate the influence of legal effectiveness on property rights.

Figure 7. *Graph of correspondence analysis for PR and JE variables*



Source: Authors' own research results.

We can conclude that, indeed, legal efficiency and property rights are strongly correlated with each other, since their equivalents occupy similar positions on the graph of correspondence analysis.

5. Conclusions

The macroeconomic conditions of a country have a major impact on the strategies of the companies to be present in the respective economy. If economic growth and currency appreciation are an attraction for foreign companies, strong inflation and currency depreciation are challenges that marketing specialists must face.

The expected result of our research in this article was a clear correlation between how the index of economic freedom is composed and the importance of using the techniques of multidimensional data analysis.

International experts, participating in the drafting of the Heritage Foundation Report, argue that, in general, global economic freedom has expanded. At the same time, they emphasize that the states in the top 25% of the top demonstrate higher rates of economic growth. Also, the incomes of the inhabitants of the states with greater economic freedom, are usually about 10 times higher than those in the states with less economic freedom.

Based on the analyzes performed on the components of the index analyzed, I have also noticed that there are countries where the factors that determine the integrity of the government are at a very high level compared to the rest of the countries, so these countries

have a lower level of corruption. At the same time, I noticed that the fiscal pressure in certain countries is high, which is also reflected in a significant decrease in the index of economic freedom.

Another observation was made regarding government spending that in some countries does not have a positive economic justification. According to the correlation matrix, we noticed that there are strong links between the following variables: Score – Property Rights, Property Rights – Government Integrity, Judicial Effectiveness – Government Integrity, Government Integrity – Business Freedom, Business Freedom – Financial Freedom, Monetary Freedom – Investment Freedom, Trade Freedom – Financial Freedom and Investment Freedom – Financial Freedom.

Therefore, we can say that the economic environment of a country is influenced by the behavior of the index of economic freedom so that in our future research we will approach this subject from the point of view of economic cybernetics and the approach based on the system dynamics. It is important to represent the cyber system of the components of the economic freedom index that placed in a dynamic system will simulate behavior with an impact on the economic environment.

Note

(1) PCA – Principle component analysis.

References

Angelopoulou, E., Balfoussia, H. and Gibson, H., 2013. Building a Financial Conditions Index for the Euro Area and Selected Euro Area Countries. What does it tell us about the crisis?, ECB Working Paper Series, No. 1541.

Beacj, W.W. and Kane, T., 2008. Methodology: measuring the 10 economic freedoms. In Kim, R., Feulner, E.J. and O'Grady, M.A. (eds.), *Index of Economic Freedom,* Chapter 4, The Heritage Foundation and Dow Jones & Company, Inc., Washington DC.

Bjørnskov, C., 2016. Economic freedom and economic crises, *European Journal of Political Economy* <https://www.sciencedirect.com/science/article/pii/S0176268016301082>

Chirita, N. and Nica, I., 2019. *Cibernetica firmei. Aplicații și studii de caz.* Economica.

Czinkota, M. and Ronkainen, I., 2004. *International Marketing,* Seventh Edition, Thomson South-Western.

Erokhin, V., Endovitsky, D., Bobryshev, A., Kulagina, N. and Ivolga, A., 2019. Management Accounting Change as a Sustainable Economic Development Strategy during Pre-Recession and Recession Periods: Evidence from Russia, *Sustainability*, 11, 3139; doi:10.3390/su11113139.

Everitt, B.S., Landau, S., Leese, M. and Stahl, D., 2011. *Cluster Analysis,* Wiley.

García-Sánchez, E., Guerrero-Villegas, J. and Aguilera-Caracuel, J., 2019. How Do Technological Skills Improve Reverse Logistics? The Moderating Role of Top Management Support in Information Technology Use and Innovativeness, *Sustainability*, *11*(1), p. 58 < https://doi.org/10.3390/su11010058>

Gorecki, J., 2020. Simulation-Based Positioning of Circular Economy Manager's Skills in Construction Projects, *Symmetry*, *12*(1), p. 50 <https://doi.org/10.3390/sym12010050>

Greenacre, M.J., 2007. *Correspondence Analysis in Practice,* Second Edition, Taylor & Francis Group;

James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. *An Introduction to Statistical Learning with Applications in R,* Springer, New York, NY, USA.

Kollu, T. and von Rosen, D., 2005. *Advanced Multivariate Statistics with Matrices,* Springer.

Maer-Materi, M.M., Mocanu, C., Zamfir, A.M. and Georgescu, T., 2019. Skill needs for Early Career Researchers – A text Mining Approach, *Sustainability*.

Menardi, G. and Torelli, N., 2014. Training and assessing classification rules with imbalanced data. *Data Min. Knowl. Disc*., 28, pp. 92-122.

Morrison, D.F., 1976. *Multivariate Statistical Methods,* The Worton School.

Musetescu, A., 2012. The importance of the Economic Freedom Index in characterizing the economic environment of a country <http://www.revistadestatistica.ro/suplimente/2012/1/srrs1_2012a45.pdf>

Musetescu, A., 2011. *International marketing,* Pro Universitaria.

Naanwaab, C., 2013. Determinants of productivity in Africa: The role of economic freedom, *Journal of Development and Agricultural Economics*, 5(10), pp. 390-402.

Ott, J., 2016., Measuring Economic Freedom: Better Without Size of Government. *Soc. Indic Res* 135, pp. 479 -498 (2018) <https://doi.org/10.1007/s11205-016-1508-x>

Rencher, A., 2002. *Methods of Multivariate Analysis,* Second Edition, Wiley.

Scarlat, E. and Chirita, N., 2019. *Cibernetica sistemelor economice,* Third Edition, Economica.

Yelland, P.M., 2010. An Introduction to Correspondence Analysis, *The Mathematical Journal*, No. 12, dx.doi.org/doi:10.3888/tmj.12|4.