ANNEX 8: STATISTICAL ANALYSIS

1. Introduction

The present study has attempted to define and to provide a sufficiently thorough and encompassing quantitative analysis of the relationship between corruption and organized crime, based on the available data.

Utilising measures of corruption and organized crime, the countries have been grouped. Then, an analysis has been made of the clusters of countries obtained from the survey so that comparisons can be made not only between the individual EU countries, but also between the clusters of countries themselves.

In this way, it is possible to analyse the development of corruption and organized crime in the European Union and, which would aid in the selection of the most effective approach of counteracting them.

The indicators that have been used in the analysis include:

- corruption indices,
- indices that measure organized crime,
- the effectiveness of government,
- the gray economy,
- as well as various macro-economic indicators.

Some of these are composite indices, while the rest have been obtained from surveys of experts or from populations surveys (e.g. Eurobarometer). The period for which data has been collected comprises the years 2004 – 2009.

The table below lists the main indicators that were tested: corruption indicators for specific institutions (police, customs, judiciary, administrative and political) and organised crime indicators for various illegal markets and activities (drugs, sex-trafficking, car theft, and money laundering).[171]

When selecting indicators for organised crime, the project team tried to identify indexes that are derived from objectively registered crimes likely to have been carried out by organised criminal groups, and other illegal markets data. For instance, in the case of usage of drugs by the population of a given country, instead of police records, representative surveys of drug usage prevalence were used. These studies are collected on a regular basis by the EMCDDA (prevalence data for Cannabis, Cocaine, Amphetamines, Ecstasy and Heroin). In the case of theft of motor vehicles, police records were used (collected by Eurostat), as they are considered more reliable due to the high percent of victims reporting these crimes to comply with the insurance requirements and to avoid administrative sanctions. This type of crimes is characterised by a sharp decline in amateur thefts, as criminal organisation and infrastructure have become a necessity.

TABLE 23. LIST OF TESTED INDICATORS

Type of indicator	Name of indicator	Source
Corruption – general indicators:	Control of Corruption Index (2007)	IBRD 2000-2007
	Extra Payments Bribes (2006)	CATO/GCR 2000-2006
	Corruption in National Institutions (2007)	Eurobarometer 2002 & 2005
Corruption – specific indicators:	Police Corruption (Experience & Perceptions)	Eurobarometer 2005 & 2007
Police corruption	Police Corruption experience & perception of asking bribes	Eurobarometer 2005 & 2007
	Police Corruption perceptions & experience of offering bribes	Eurobarometer 2005 & 2007
Customs corruption	Irregular Payments in Import Export permits (2005)	GCR 2001-2006
Judicial corruption	Irregular payments in judicial decisions (2006)	GCR 2002-2006
	Judicial independence	GCR 2001-2008
Admin. corruption	Irregular payments in Public Contracts (2006)	GCR 2001-2006
	Corruption in Local Institutions (2007)	Eurobarometer 2002 & 2005
Political corruption	Corruption in National Institutions (2007)	Eurobarometer 2002 & 2005
	Favouritism in decisions of government officials (2008)	GCR 2001-2008
Organised crime: general indicator	Organized crime (2008)	GCR 2001-2008
Organised crime: specific indicators:		
Drugs	Prevalence use amongst adults of Cannabis, Cocaine, Amphetamines, Ecstasy, Heroin	EMCDDA (national surveys)
Sex trafficking	Trafficking of people Convictions and investigations 2007	UNODC 2003-2007
Car theft	Police recorded thefts per 100,000 population	Eurostat 1999-2006
Money laundering	Pervasiveness of Money Laundering through banks (2005)	GCR 2002-2005
	Pervasiveness of Money Laundering through Non-bank Channels 2004	GCR 2002-2004

In the present analysis, two main goals have been set:

- **The first goal** was to group all 27 countries of the European Union in clusters with similar characteristics, based on the indicators of corruption and organised crime. This has been achieved by means of cluster analysis and neural networks.

[171] *Two additional indicators were considered but dismissed: illegal migrants – data by Frontex on illegal migrant arrest at borders – the data was highly volatile; Murder data – no recent conviction statistics were available to determine the level of organised crime related murders.*

- **The second goal** has been to examine the strength and direction of the relations between the selected indicators through correlation analysis, and also to model the relations thus obtained. In order to model the relationships, regression analysis was used.

The analysis had the following tasks:
1. To examine the relations amongst the indicators that measure corruption, organized crime, the effectiveness of the police and of government;
2. To standardize and transform the data;
3. To group the countries of the study with the aid of cluster analysis and neural networks, utilising indices of corruption, organized crime, the gray economy, and selected macro-economic indicators;
4. To determine the relationship between corruption and organized crime through regression and correlation analysis;
5. To draw conclusions and offer recommendations.

The scope of the data that was collected: included all European Union Member States, and the indicators used covered the period 2004 – 2009.

2. Description of methods used

2.1. Cluster analysis

Cluster analysis is a set of techniques that allows groupings of cases to be formed on the basis of one or more variables. Cluster analysis may be used where the number of groups is initially unknown, as well as after this number has been hypothesized or established.

There are three different procedures that can be used to cluster data: **hierarchical cluster analysis**, *k*-means cluster, and two-step cluster. If a small data set is available and the goal is to easily examine solutions with increasing numbers of clusters (as is the case), hierarchical clustering may be used.

Hierarchical clustering is one of the most straightforward methods of cluster analysis. Agglomerative hierarchical clustering starts with each case (in this study, each country) being a cluster. At the next step, two or more countries that have the smallest value for the distance measure (or largest value if one uses similarities) are joined into a single cluster. At the second step, either a third case is added to the cluster that already contains two cases or two other cases are merged into a new cluster. At every step, either individual cases are added to existing clusters, two individual cases are combined, or two existing clusters are combined.

The results of the application of the clustering technique are best described using a *dendrogram*. The branches illustrate when the cluster method joins subgroups containing the objects. The *length* of the branches indicates the distance between the subgroups when they are joined. *Distance* is a measure of how far apart two objects are, while *similarity* measures how similar two objects are.

There is no right or wrong answer as to how many clusters can be formed. The choice on the number of clusters depends on the aims of the research project in question. To find a good cluster solution, a study must consider the characteristics of the clusters at successive steps and decide when an interpretable solution, or a solution that has a reasonable number of fairly homogeneous clusters, has been arrived at.

2.2. Neural Networks

Neural networks are an alternative approach to cluster analyses which adds further explanatory value to the results.

2.2.1. Self Organizing Kohonen Maps

Kohonen maps are among the most popular kinds of neural networks. They are intended to identify clusters of similar data, and to determine their proximity as well. They work on the principle of 'unsupervised learning', realizing a process of clustering. Only input data is sent to the network, and it does not have any preliminary given output information.

The algorithm involved in Kohonen maps is a variation of multi-dimensional vectors clustering. With the help of this algorithm a mapping from a higher dimensional input space (determined by the number of indicators) to a lower dimensional (it is usually two-dimensional, but it is also possible to be one-dimensional) with preserving the topological resemblance, is achieved. This means that all vectors, which are adjacent to the topological map, are also adjacent in the input space. It should be noted that the opposite is not always true.

Beginning with these randomly situated centres of clusters, the algorithm gradually improves their position in such a way as to catch the input data clustering (the objects in the input space are represented as dots). As a result of the iterative procedure of learning the map self-organizes in such a way that the elements, corresponding to the centres and situated near one another in the input space, are also situated close to the topological map (the output layer). The algorithm is known as 'the winner takes all'.

After the Kohonen network is trained, the so called 'Unified Distance Matrix' (U-Matrix) is used for the recognition of clusters. In this way the distance (usually Euclid) from each neuron to its neighbours on the topological map is calculated. This distance determines in what colour the neuron is represented on the map. Small distances signify resemblance of the neuron-neighbours, while greater distances stand for greater difference. The colouring is done analogically to altitude maps – small values are coloured in green, and high ones- in brown. In this way, clusters on the map should form areas in green colours, and around them beige-brown-red areas should be situated- the boundaries of the clusters. Another colouring option is to use black and white. In this option white corresponds to small distances, and black corresponds to large ones. In this way clusters are coloured in white, and their boundaries – in black.

It is also possible that maps of variables are produced, used for describing the input vectors. In this way it can be identified in which region of the map the corresponding variable has low values, and in which region -high ones. This makes it possible, 'portraits' of the clusters to be made, that is, their description to be made up. The received combination of cards represents an original 'atlas', describing the situation of the variables and clusters in the combination of data.

2.3. Regression and Correlation Analyses

Regression and correlation analyses are statistical techniques used extensively to examine causal relationships between variables.

2.3.1. Linear Regression Analysis

In linear regression, the model used to describe the relationship between a single dependent variable and a single independent variable is:

$$Y_i = \beta_o + \beta_1 X_i + \varepsilon_i,$$

where, β_o and β_1 are referred to as the model parameters, Y_i is the value of the dependent variable, X_i is the value of the independent variable, *e* in this equation means "error" and refers to the fact that we don't expect any regression equation to perfectly predict Y_i.

The regression coefficient β_1 represents the change in the predicted value of dependent variable for each one-unit increase in independent variable. This means that if independent variable is changed by one unit, dependent variable will increase or decrease by β_1 units, on average.

The significance levels show how likely a result is due to chance. The most common level, used to mean something is good enough to be believed, is 0.95. This means that the finding has a 95% chance of being true. Instead it will show you 0.05, meaning that the finding has 5% chance of not being true, which is the converse of a 95% chance of being true.

To measure how strong the correlation is between the two variables, we can determine the amount of the total variation in ***Y*** that is associated with the regression model. This ratio is sometimes called **the coefficient of determination (R square)**. This coefficient describes what proportion of the variation in the dependent variable is associated with the variation of an independent variable. The value of the coefficient of determination ranges from 0 to 1.

2.3.2. Correlation Analysis

Another useful regression statistic that measures the strength of the correlation between to variables is the Pearson's correlation coefficient. This statistic is often represented by the symbol ***r*** and is determined by taking the square-root of the ***coefficient of determination***. The value of the

correlation coefficient ranges from 1 to -1. A value of 0 indicates that there is absolutely no relationship between the ***X*** and ***Y*** variables. The strength of the relationship between the ***X*** and ***Y*** variables increases as the value of r approaches 1 or -1. Positive correlation coefficients indicate that an increase in the value of the ***X*** variable results in an increase in the value of the ***Y*** variable. Negative correlation coefficients indicate that an increase in the value of the ***X*** variable results in a decrease in the value of the ***Y*** variable.

Correlation and regression analysis are related in the sense that both deal with relationships among variables. For simple linear regression, the sample correlation coefficient is the square root of the coefficient of determination, with the sign of the correlation coefficient being the same as the sign of β_1 – the regression coefficient.

3. Data analysis

3.1. Standardization and transformation of the data

In order to apply correctly cluster analysis two steps should be considered as a precondition:

- A ser of appropriate variables has to be chosen. This step is essential, because any change of the set of variables involved affects the results of the analysis.
- Values of the selected variables have to be transformed. This has to be done especially when there are variables presented on different scales (as is the case). There are different types of transformation. In this particular case a "mini-max" transformation is used. The values are standardized to a min of 0 and max of 10.

Data from 125 indicators have been used in the study, 6 of which on a scale from -2,5 to 2,5, 17 are from 1 to 5, 13 are from 1 to 7, 11 are from 0 to 10, 14 are from 0 to 100, 51 are from 0 to 100 %, and 8 are in absolute values. There is only one indicator on a scale from 0 to 5 and 4 indicators expressed in a currency. Depending on these measuring scales, all the variables are transformed within limits of from 0 to 10 by the mini-max transformation.

Should the data take absolute values, it is necessary first to norm them. An example of such data is the police data on recorded traffic of people and drugs, thefts of motor vehicles, convictions for murders. In order to norm the data, their absolute values are divided by the population number, and then they are transformed.

3.2. Grouping of the countries in the study using Cluster Analysis and Neural Networks

There are different approaches to classify the countries. The most frequently used are cluster analysis and neural networks. It should be stressed that the task of classification has usually more than one solution.

TABLE 24. SUMMARY OF THE REGRESSION RESULTS FOR THE INDICATOR CORRELATIONS TESTED

№	Year	Variable 1	Variable 2	R correlation
1	2007	Corruption in National Institutions_2007	Reliability of police services: Q: Police services in your country _2007	0.747
2	2007	Corruption in National Institutions_2007	Corruption in the Police_2007	0.659
3	2007	Corruption in National Institutions_2007	Corruption in Local Institutions_2007	0.974
4	2007	Corruption in National Institutions_2007	Favouritism in decisions of government officials_2007	0.834
5	2007	Corruption in National Institutions_2007	Drugs data UNODC	0.105
6	2003	Control of Corruption Index _2003	Shadow Economy Estimate as percentage of GDP_2003	0.647
7	2007	Control of Corruption Index _2007	Reliability of police services: Q: Police services in your country _2007	0.916
8	2007	Control of Corruption Index _2007	Corruption in the Police_2007	0.752
9	2005	Control of Corruption Index _2005	Irregular payments in Import/Export permits_2005	0.830
10	2006	Control of Corruption Index _2006	Irregular payments in judicial decisions_2006	0.913
11	2006	Control of Corruption Index _2006	Irregular payments in Public Contracts_2006	0.901
12	2007	Control of Corruption Index _2007	Corruption in Local Institutions_2007	0.851
13	2007	Control of Corruption Index _2007	Corruption in National Institutions_2007	0.837
14	2007	Control of Corruption Index _2007	Favouritism in decisions of government officials_2007	0.947
15	2007	Control of Corruption Index _2007	Drugsdata_UNODC	0.292
16	2006	Control of Corruption Index _2006	Trafficking of people Convictions and investiagations_2006	0.029
17	2006	Control of Corruption Index _2006	N_Crimes recorded by the police Theft of a motor vehicle_2006	0.529
18	2005	Control of Corruption Index _2005	Pervasiveness of Money Laundering through Banks_2005	0.885
19	2004	Control of Corruption Index _2004	Pervasiveness of Money Laundering through Nonbank Channels_2004	0.919
20	2003	Extra payments/bribes _2003	Shadow Economy Estimate as percentage of GDP_2003	0.472
21	2006	Extra payments/bribes _2006	Reliability of police services: Q: Police services in your country _2006	0.866
22	2005	Extra payments/bribes _2005	Irregular payments in Import/Export permits_2005	0.932
23	2006	Extra payments/bribes _2006	Irregular payments in judicial decisions_2006	0.921
24	2006	Extra payments/bribes _2006	Irregular payments in Public Contracts_2006	0.921
25	2006	Extra payments/bribes _2006	Favouritism in decisions of government officials_2006	0.908
26	2006	Extra payments/bribes _2006	Drugs data UNODC	0.319
27	2006	Extra payments/bribes _2006	Trafficking of people Convictions and investiagations_2006	0.060
28	2006	Extra payments/bribes _2006	N_Crimes recorded by the police Theft of a motor vehicle_2006	0.521
29	2005	Extra payments/bribes _2005	Pervasiveness of Money Laundering through Banks_2005	0.796
30	2004	Extra payments/bribes _2004	Pervasiveness of Money Laundering through Nonbank Channels_2004	0.838
31	2008	Organized crime_2008	Reliability of police services: Q: Police services in your country _2008	0.671
32	2007	Organized crime_2007	Corruption in the Police_2007	0.291
33	2005	Organized crime_2005	Irregular payments in Import/Export permits_2005	0.634
34	2006	Organized crime_2006	Irregular payments in judicial decisions_2006	0.592
35	2006	Organized crime_2006	Irregular payments in Public Contracts_2006	0.510
36	2007	Organized crime_2007	Corruption in Local Institutions_2007	0.516
37	2007	Organized crime_2007	Corruption in National Institutions_2007	0.491
38	2008	Organized crime_2008	Favouritism in decisions of government officials_2008	0.613
39	2008	Organized crime_2008	Drugs data UNODC	0.111
40	2006	Organized crime_2006	Trafficking of people Convictions and investiagations_2006	0.117
41	2006	Organized crime_2006	N_Crimes recorded by the police Theft of a motor vehicle_2006	0.180
42	2005	Organized crime_2005	Pervasiveness of Money Laundering through Banks_2005	0.812
43	2004	Organized crime_2004	Pervasiveness of Money Laundering through Nonbank Channels_2004	0.857

3.2.1. Cluster Analysis

In order to reach a final variant of clustering of the countries, a total of 332 different iterative experiments have been conducted with the aid of the hierarchical cluster analysis, wherein various indicators for the countries have been involved. ***Usually, the change in the number and type of indicators by which the analysis is made may lead to the formation of various clusters.*** The inclusion (or exclusion) even of a single indicator may have a substantial impact on the formulation and the composition of the clusters.

For the above reason, the most important and essential stage in the application of cluster analysis is the proper set of variables by which the individual groups of objects may be formed (in this case, the member countries of the EU).

The final clustering of the countries was based on the following indicators that measure corruption, organized crime, the effectiveness of government, macro-economic indicators, and the size of the grey economy:

1) *Control of Corruption Index from 2000, 2002 to 2007 Year*
2) *Extra payments/bribes from 2000 to 2006 Year*
3) *Organized crime from 2001 to 2008 Year*
4) *Corporate Ethics Index for 2004 Year*
5) *Rule of Law Index from 2000 to 2007 Year*
6) *Crimes recorded by the police: Drug Trafficking from 1999 to 2006 Year*
7) *Crimes recorded by the police: Theft of a motor vehicle from 1999 to 2006 Year*
8) *Cocaine (UNODC – drugs data) from 2003 to 2006 Year*
9) *Size of Government Index from 2000 to 2006 Year*
10) *Government Effectiveness Index from 2000 to 2007 Year*
11) *Overall Economic Freedom Score from 1999 to 2009 Year*
12) *GDP per capita in PPS from 1999 to 2008 Year*
13) *Share of Envelope Wages for 2007 Year*

The next stage in a correct and proper application of the cluster analysis is the transformation of the data in such a way that they could be represented on the same scale.

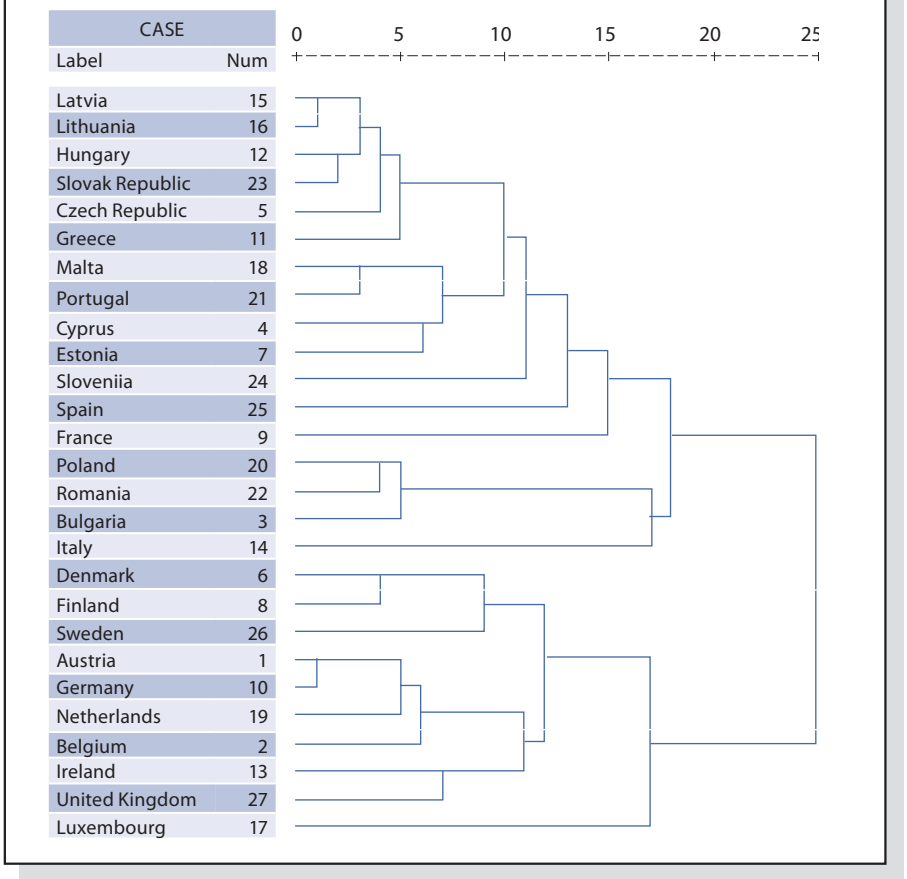
Initially, the values of the selected variables by which the respective groups of countries have been formed, are transformed through a mini-max transformation. What's typical for it is that the values of the standardized variables receive a minimum in the zero and a maximum in the one.

Through the use of the same indicators, a new clustering has been done, using a Z-transformation this time. In comparison with the mini-max transformation, the clusters obtained by a Z-transformation are more stable. In other words, irrespective of the algorithm and metrics that have been used, in most cases the composition of the groups of countries thus formed remains the same.

With the Z-transformation, the values of the variables receive a mean of 0 and a standard deviation of 1. With the hierarchical cluster analysis the results can be visualized with the aid of a tree-type diagram (dendogram). It is possible to track the clustering process on it, every single country forming a separate cluster in the beginning. Gradually, the countries begin to be grouped according to their degree of similarity in terms of the investigated characteristics. The more similar the countries, the quicker they group together, while the more different they are, the more iterations of the algorithm are necessary in order to unite them. The clustering process is concluded when all the countries have been united into one single cluster.

As already mentioned, during the first step of clustering, every country is separated in a different cluster. During the second step, 11 clusters of countries have been formed, but they are not clearly expressed. During the next step, one can already clearly distinguish 6 clusters that satisfy the goal of the study. The countries have been grouped in the following way:

FIGURE 21. DATA CLUSTERING USING Z – TRANSFORMATION



Cluster 1 – Austria, Belgium, Denmark, Finland, Germany, Ireland, Netherlands, Sweden and UK

Cluster 2 – Bulgaria, Poland, Romania

Cluster 3 – Cyprus, Czech Republic, Estonia, Greece, Hungary, Lithuania, Latvia, Malta, Portugal, Slovakia, Slovenia, Spain

Cluster 4 – France

Cluster 5 – Italy

Cluster 6 – Luxemburg

When six separate groups of countries had been obtained, then a comparison was made with the average values of the indicators that measure the effectiveness of government, the financial and economic indices, and those of the gray economy. Thus, a profile is obtained (a portrait) of the selected clusters, and it answers the question why the countries of the European Union have grouped themselves in such a way in terms of the above listed indicators. Besides, with the help of the results thus obtained a comparison may be made both between the individual EU countries as well as between the clusters themselves.

Cluster 1 is described as one with the highest level of control of corruption and the lowest of the organized crime, the highest value of the ethic index, effective rule of law and also effectiveness of the government, as well as the highest GDP per capita. At the same time countries in this cluster have the highest value of the drug trafficking and the theft of motor vehicle.

Cluster 2 countries have the worst scores for the following indicators – the lowest levels of control of corruption and rule of law, the lowest GDP per capita. It also has the highest level of undeclared income (envelope wages) but the lowest value of the theft of motor vehicle.

Cluster 3 is similar to Cluster 2, but with better scores for control of corruption and rule of law.

The last three independent clusters consist of one state each which speaks of the uniqueness of the respective countries and of the impossibility to include them in any of the remaining clusters. Luxembourg **(Cluster 6)** is the country with the highest results for all of the indicators (high degree of control on corruption, low rates of organized crime, the highest GDP per capita of the population). At the same time, it demonstrates a highly developed market for drugs and car thefts. High results are also exhibited by **Cluster 4** (France). Besides, the indicator measuring drug traffic and car theft in France is sufficiently low. Italy **(Cluster 5)** is the country with some of the poorest results – similar to cluster 2, which contains Bulgaria, Romania, and Poland.

3.3.1. Neural Networks (Self-Organizing Kohonen Maps)

3.3.1.1. Architecture of the Kohonen Maps

A number of experiments were carried out to come up with the optimal network architecture. Maps of various sizes were tested, starting from 8x8 (64 neurons) to 50x50 (2500 neurons). We selected a map size of 50x50 for the Kohonen layer. We tested the selected architecture to guarantee sustainability of results.

3.3.1.2. Results of the Kohonen Maps – Interpretation

The values of the following indicators were used to group the EU countries:

1) Control of Corruption Index for 2007 Year
2) Extra payments/bribes for 2006 Year
3) Organised crime for 2008 Year
4) Corporate Ethics Index for 2004 Year
5) Rule of Law Index for 2007 Year
6) Crimes recorded by the police: Drug Trafficking for 2006 Year
7) Crimes recorded by the police: Theft of a motor vehicle for 2006 Year
8) Cocaine usage (UNODC – drugs data) from 2003 to 2006 Year
9) Size of Government Index for 2006 Year
10) Government Effectiveness Index for 2007 Year
11) Overall Economic Freedom Score for 2009 Year
12) GDP per capita in PPS for 2008 Year
13) Share of Envelope Wages for 2007 Year
14) GINI coefficient for 2007 Year

The figure, below, presents a Unified Distance Matrix. The resulting complex surface has peaks (representing long distances between neurons), and valleys (representing small distances).

Information that can be extracted from this map is related to the placement of countries on it and the colour of the "border" neurons, which are interpreted as a similarity or dissimilarity between them. Slovenia (SVN) is placed at the top left corner of the map, immediately adjacent to Luxembourg (LUX), Netherlands (NLD) and Belgium (BEL). The distance between these countries is relatively small (identified by the beige and light-brown colour of their neurons on the chart). Not far from the above countries are placed Germany (DEU), Austria (AUT), Ireland (IRL), United Kingdom (GBR) and Spain (ESP).

In other group belong Latvia (LTU), Lithuania (LTV), Estonia (EST) and Portugal (PRT). Greece (GRC), Hungary (HUN), Czech Republic (CZE), Cyprus (CYP), Slovakia (SVK) and Malta (MLT) form another cluster.

In an independent group are Bulgaria (BGR), Romania (ROM), Poland (POL) and they are far from the cluster, composed of Denmark (DNK), Finland (FIN), Sweden (SWE).

FIGURE 22. UNIFIED DISTANCE MATRIX (U-MATRIX)

France (FRA) and Italy (ITA) are placed in separate clusters, because of their dissimilarity to the neighbours (these two countries are surrounded by neurons coloured in dark-brown and red).

3.3.1.3. Cluster characteristics

Moreover analysis of the composition of clusters is necessary to clarify the reasons for their formation. The self-organising maps created for each indicator allowed to provide a detailed description of the resulting clusters. The colouring is the same principle – the areas in which the index has low values are coloured in green and those in which the values are high – in brown-red.

To visualise the resulting clusters, each cluster is compared with the average values of the government effectiveness indicator, the financial and economic indexes, the grey economy index and GINI coefficient.

Cluster 1: Denmark, Finland and Sweden. These countries are with the best scores of the almost all indicators – low levels of organised crime and corruption and effectiveness of the government.

Cluster 2: UK, Ireland, Austria, Germany, Netherlands, Belgium, Luxembourg, Slovenia and Spain. These are the countries with best control of corruption, lowest level of organised crime, highest scores for rule of law, and highest GDP per capita. Like Cluster 1, but there is GDP is higher.

Cluster 3: France. Statistical analysis demonstrated that France is unique by some key indicators and therefore cannot be included in any of the other clusters. It is characterised by high level of control of corruption, low frequency of bribes, and effective rule of law. The size of government is the lowest, and effectiveness of government is one of the highest in EU. France also has one of the highest GDP per capita. At the same time, by these same indicators (corruption, rule of law, GDP) France scores distinctively lower than most of its west European neighbours (Cluster 2) and the Scandinavian countries (Cluster 1).

Cluster 4: Latvia, Lithuania, Estonia and Portugal. It has one of the lowest levels of organised crime. With better indicators than Italy, Greece, and considerably worse than the countries in Clusters 1 and 2. This cluster is also characterised by one of the lowest GDP.

Cluster 5: Greece, Hungary, Czech Republic, Cyprus, Slovakia and Malta. There is characterised by relatively low control of corruption and frequent use of bribes. At the same time, low level of organised crime.

Cluster 6: Italy. Similar to France, Italy's key indicators have significantly different values that justify placing it in a cluster of its own. Italy has the highest level of organised crime among the member states, combined with one of the lowest scores for control of corruption and rule of law. It also has the highest level of undeclared income (envelope wages).

Cluster 7: Bulgaria, Romania and Poland where corruption control is lowest and organised crime is at a level similar to Italy. These countries have the worst scores for rule of law, effectiveness of government and corporate ethics. They also have the lowest GDP per capita among the member states.

The self-organising maps created for each indicator allowed to provide a detailed description of the resulting clusters. The colouring of the Unified Distance Matrix (Figure 3) follows the same principle: low values of each indicator are coloured in green, while zones with high values are coloured in brown/red.

To visualise the resulting clusters, each cluster is compared with the average values of the government effectiveness indicator, the financial and economic indexes, the grey economy index, as well as the additional indicator, GINI coefficient.

The table below provides a detailed numerical description of each of the identified clusters. The table consists of the average values of each indicator taken to construct each cluster. For instance, the average value of the Control of Corruption Index for the countries in Cluster 1 (Denmark, Sweden, and Finland) is the highest (2.459), while for the countries in Cluster 7 (Bulgaria, Romania, and Poland), it has the lowest value (0.091). The clusters are described inside the report

TABLE 25. AVERAGE VALUES OF CLUSTER CHARACTERISTICS

CLUSTER #INDICATOR	1	2	3	4	5	6	7
Control of Corruption Index 2007	2.459	1.721	1.320	0.639	0.541	0.449	- 0.091
Extra payments/ bribes 2006	9.470	8.361	8.240	6.918	6.397	6.290	5.397
Organised crime 2008	6.600	6.056	5.900	6.150	5.983	3.600	4.433
Corporate Ethics Index 2004	0.825	0.662	0.597	0.430	0.376	0.409	0.229
Rule of Law Index 2007	1.905	1.587	1.317	0.750	0.836	0.426	- 0.013
Drug Trafficking 2006 (Eurostat police data)	60.941	91.711	9.137	42.740	47.258	54.634	19.833
Theft of a motor vehicle 2006 (Eurostat)	439.481	183.841	383.570	129.698	139.851	473.805	31.427
Cocaine prevalence use (EMCDDA)	0.567	1.289	0.600	0.350	0.350	2.100	0.200
Size of Government Index 2006	6.423	5.166	4.110	6.348	6.308	5.990	5.277
Government Effectiveness Index 2007	2.079	1.561	1.304	0.850	0.934	0.329	0.131
Overall Economic Freedom Score 2009	74.867	73.356	63.300	69.475	67.217	61.400	62.700
GDP per capita in PPS 2008	116.133	131.689	105.700	63.525	78.500	97.600	45.700
Share of Envelope Wages 2007	0.070	0.172	0.540	0.420	0.194	0.629	0.555
GINI 2007	24.330	28.330	28.000	36.750	28.330	33.000	31.000

On the basis of the resulted groups specific statistical methods will be applied such as regression and correlation analysis.

3.3. Determining relationship between corruption and organised crime using regression and correlation analysis

By means of the analyses that were used, the relations between the basic indicators and the specific indicators of corruption and organized crime were investigated. Besides, with their aid, it is proper to show the impact of indicators such as government effectiveness, the economic and financial indicators, and others on variables measuring corruption and crime.

By the respective analysis, a relationship is sought both between absolute values of the given indices and between their accretions for two consecutive years.

In Table 1, the value of ***R square –*** *the coefficient of determination* is 0.520. The correlation coefficient is **r = 0.721.** Therefore, it can be admitted that the examined indices are very closely related to each other, and that 52% of the changes to the extra payments and bribes in the member countries of the EU are due to changes in organized crime.

TABLE 26. RELATIONSHIP BETWEEN CORRUPTION ('EXTRA PAYMENTS/ BRIBES) AND ORGANISED CRIME

Year	Model Summary				Coefficients			
	Dependent Variable	Independent Variable	r	R square	Coefficients	B	Std. Error	Significance
2006	Extra payments/ bribes	Organized crime	0.721	0.520	constant	2.223	1.015	0.038
					coefficient	0.949	0.182	0.000

The remaining percentages up to 100 (in this case, 48%) are accounted for as the impact of other factors.

B – *the regression coefficient* is 0.949. This shows that if organized crime would decrease by one point, it can be expected that the indicator of corruption (Extra payments/ bribes) would decrease significantly (by 0,949 points).

The ***Significance*** of *the regression coefficient* is less than 0.01. Therefore, it can be asserted with a probability of 99% that the basic index which measures organized crime exerts a substantial impact on the bribes in the EU countries for the year 2007.

The last table shows a significant interrelationship between organized crime and the effectiveness of government.

TABLE 27. RELATIONSHIP BETWEEN ORGANIZED CRIME AND GOVERNMENT AND INSTITUTIONAL EFFECTIVENESS (JUDICIAL/LEGAL EFFECTIVENESS)

Year	Model Summary				Coefficients			
	Dependent Variable	Independent Variable	r	R square	Coefficients	B	Std. Error	Significance
2004	Organized crime	Judicial/Legal Effectiveness	0.872	0.761	constant	3.042	0.277	0.000
					coefficient	3.731	0.418	0.000

In Table 29, the coefficient of ordinary correlation is very close to one. That is why, it is possible to admit that the indicators of crime and the effectiveness of the judicial system are closely interrelated.

The ***Significance*** of *the regression coefficient* in the models that are shown is less than 0,01, and with a probability rate of 99% it can be asserted that organized crime exerts a substantial impact on corruption and vice versa. Besides, the effectiveness of government and the gray economy have a significant impact on the development of corruption and organized crime in the member countries of the EU.