

Regional economic discrepancies and the impact of the underground economy in Romania

Nicoleta-Luminița Popescu
PhD Student, Bucharest University of Economic Studies,
Nicoleta.groaznicu@gmail.com

Abstract: The article explores the impact of the underground economy in Romania's regions, using statistical analysis to identify key economic factors. Developed regions with higher GDPs and larger formal workforces have a smaller underground economy, while less developed regions face larger informal sectors, marked by significant income-expenditure discrepancies. The study highlights two main dimensions: overall economic performance and financial imbalances. The findings emphasize the need for region-specific policies, where stronger regions focus on protecting formal economies and weaker regions require stronger interventions to reduce the underground economy and promote formal business practices.

Keywords: Underground economy, regional economy, informal sector, economic discrepancies, gross domestic product (GDP), labor force, principal component analysis (PCA), financial imbalances, economic policy

Introduction

The underground economy, often referred to as the informal economy, represents one of the major challenges faced by modern states, including Romania. Although the underground economy may seem marginal, its influence on the formal economy is significant, negatively affecting budget revenues, economic competitiveness and the effectiveness of public policy measures. In Romania, the underground economy varies by region, reflecting significant differences in terms of economic development, employment and degree of business formalization.

The aim of this article is to explore these regional variations of the underground economy through a detailed statistical analysis, using data on regional gross domestic product, average number of employees, turnover officially declared by companies and discrepancies between reported income and expenditure. In this endeavor, we focus on identifying the economic factors that contribute to the growth or decline of the underground economy at the regional level.

The analysis of the underground economy is carried out by means of modern statistical methods, such as principal component analysis (PCA) and factorial analysis, which allow the identification of the essential variables that characterize this phenomenon. The adopted methodology is based on the processing of complex

data sets, using the SPSS computer program, and provides a clear vision of how the underground economy manifests itself in Romania's regions.

The article aims to contribute to a better understanding of the size and structure of the underground economy at the regional level, providing a solid basis for the formulation of economic policies adapted to the specifics of each region. It also aims to highlight regions where the underground economy has a greater impact on the formal economy, in order to guide policy interventions to reduce this phenomenon.

By identifying and quantifying the factors that influence the underground economy, the article offers a significant contribution to the discussion about the effectiveness of economic policies in combating it, emphasizing the need for specific and coordinated regional measures. This analysis not only highlights the economic discrepancies between Romania's regions, but also emphasizes the importance of an adequate response to reduce the negative impact of the informal economy on long-term economic development.

Empirical results :

The table presents descriptive statistics related to four essential variables for evaluating the underground economy at the level of regions in Romania. These variables are: the average number of employees, the index of the discrepancy between income and expenses, the regional gross domestic product and the turnover officially declared by the companies.

Table 1: Descriptive statistics for key economic indicators at the level of regions in Romania

Table 1: Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Average number of employees	651186.63	223290,315	8
Income-expenditure discrepancy index	13.255942	2.6998741	8
Regional gross domestic product	176,112	92.1136	8
The turnover officially declared by the companies	310857.13	239126.255	8

Source: Pauthor rework, results obtained through the SPSS computer program.

Average number of employees: Mean: 651,186.63 employees, Standard deviation: 223,290,315, N: 8 analyzed regions. This indicator shows the average number of employees in each region. A high value indicates a stronger economic base and wider labor force participation.

Index of discrepancy between income and expenses: Mean: 13.25, Standard deviation: 2.6998741, N: 8 regions analyzed. This index shows the discrepancy between income and expenditure recorded in each region, an indicator of the informal economy. A higher value suggests a possible more extensive underground economy.

Regional gross domestic product: Mean: 176.112 million monetary units (possibly billions), Standard deviation: 92.1136 million monetary units, N: 8 regions analyzed, regional GDP reflects the total economic output of each region. Regions with a higher GDP have more intense economic activity.

Turnover officially declared by companies: Mean: 310,857.13 million monetary units, Standard deviation: 239,126.255 million monetary units, N: 8 regions analyzed. This indicator reflects the official turnover reported by companies. Large discrepancies between reported turnover and reality may signal undeclared or underreported economic activities (part of the underground economy).

The table suggests the existence of significant variations between regions in terms of economic activity and indicators related to the underground economy. The high variability of GDP and declared turnover indicates discrepancies between more developed and less developed regions, which may affect the magnitude of the underground economy in these regions.

Table 2: Empirical results of the Bartlett test and the KMO statistic

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,586
Bartlett's Test of Sphericity	Approx. Chi-Square	38,341
	df	6
	Sig.	,000

Source: Pauthor rework, results obtained through the SPSS computer program.

Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy= 0.586. This KMO value indicates that the data are moderate for applying a factor analysis. A KMO value closer to 1 indicates a better fit of the data for factor analysis. In general, a value of 0.5 is considered an acceptable minimum, and a value above 0.6 is considered decent.

Bartlett's Test of Sphericity: $Chi-Square = 38.341$, df (degrees of freedom) = 6, $p < 0.001$. Bartlett's test measures whether the variables are correlated well enough for factor analysis to be applied. A significant result ($p < 0.05$) indicates that there are significant correlations between the variables and the covariance structure is

not an identity matrix. In this case, the significant value of the test ($p = 0.000$) confirms that the economic variables are sufficiently correlated to justify a factor analysis. This means that there are patterns of interdependence between variables that can be explored to better understand how the underground economy works in the regions under analysis.

KMO and Bartlett test results indicate that the regional economic data are moderate to good for a factor analysis. The correlations between the analyzed variables, such as the average number of employees, the discrepancy between income and expenditure, the regional GDP and the declared turnover, are sufficient to be included in a deeper analysis, which suggests that these variables can contribute to the identification of common factors that influence the underground economy at the regional level.

Table 3: Utilities for key economic indicators in Principal Component Analysis

Table 3: Communalities

	Initial	Extraction
Average number of employees	1,000	,975
Income-expenditure discrepancy index	1,000	1,000
Regional gross domestic product	1,000	,996
The turnover officially declared by the companies	1,000	,986

Extraction Method: Principal Component Analysis.

Source: Pauthor rework, results obtained through the SPSS computer program..

Sthe meaning of Communal Values (Communalities):

- The initial values of communalities are always 1 for all variables because they represent the entire variable.
- Extraction values reflect the proportion of the variance of each variable that can be explained by the factors extracted from the factor analysis.

Analysis of the variables in the table:

*Average number of employees:*Extraction = 0.975: About 97.5% of the variance of this variable is explained by the extracted factors. This means that the variable "average number of employees" has a very strong connection with the factors extracted in the factor analysis and contributes significantly to the understanding of the underground economy at the level of the regions.

*Index of discrepancy between income and expenses:*Extraction = 1.000: This value indicates that the entire variance of this variable is explained by the extracted factors. The "income-expenditure discrepancy

index" plays a crucial role and is completely related to the extracted factors, which suggests that it is an essential variable in identifying the phenomena of the underground economy.

*Regional gross domestic product (GDP):*Extraction = 0.996: About 99.6% of the variance of regional GDP is explained by the extracted factors. This shows that the regional GDP is closely related to the underground economy factors identified in the factor analysis.

*Turnover officially declared by companies:*Extraction = 0.986: About 98.6% of the variance of the turnover officially declared by the companies is explained by the extracted factors. This suggests that turnover is a very relevant indicator for identifying and explaining the underground economy at the regional level.

The communalities values for all variables are very high (between 0.975 and 1.000), indicating that the variables included in this factor analysis are very well explained by the extracted factors. The analyzed variables - average number of employees, discrepancies between income and expenditure, regional GDP and declared turnover - are relevant and strongly related to the underground economy at the level of the regions.

Table 4: Eigenvalues and variance recovered

Component	Total Variance Explained					
	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.971930	74.298257	74.298257	2.949797	73.744929	73.744929
2	0.985893	24.647321	98.945578	1.008026	25.200650	98.945578

Source: Author processing, *Extraction Method:* Principal Component Analysis, results obtained through the SPSS computer program.

The table "Total Variance Explained" presents information about how much of the total variation of the observed variables is explained by each component (factor) extracted in the factor analysis. This table provides data on both the initial "Eigenvalues" and the cumulative variance explained after rotating the components.

Component 1: Initial Eigenvalue: 2.972 – This is the eigenvalue (Eigenvalue) of the first component, meaning that it explains 2.972 times more of the total variation than an individual variable. % of Variance: 74.298% – The first component explains 74.3% of the total variation in the data. This means that the first component

captures most of the variance related to the underground economy in the analyzed regions. Cumulative %: 74.298% – This shows that up to this point component 1 explains 74.3% of the cumulative variance.

Component 2: Initial Eigenvalue: 0.986 – This is the eigenvalue for the second component, which shows that this component adds an important contribution to the explanation of variation. % of Variance: 24.647% – Component 2 explains 24.6% of the total variation. Cumulative %: 98.946% – Together, the first two components explain about 99% of the total variation in the data. This shows that the first two components are extremely strong in capturing the total variation, almost completely explaining the observed economic phenomenon.

Component 3 and Component 4: Very small Eigenvalues (0.039 and 0.003) – These components do not make a significant contribution to explaining the variance, having very small values and explaining only an extremely small fraction of the variance. Very Low % of Variance: 0.976% and 0.078% – Components 3 and 4 explain only insignificant percentages of the variation and do not contribute significantly to understanding the underground economy.

The first two components extracted explain almost all the variation in the data set (about 99%). This means that most of the information regarding the underground economy in the analyzed regions is captured by these two components. These components reflect the main factors contributing to economic variations between regions and suggest that the data set is dominated by a few key factors, such as the average number of employees, discrepancies between income and expenditure, regional GDP and declared turnover. These variables explain most of the observed economic phenomena.

The underground economy at the regional level can be effectively explained by the two main components identified in the analysis, which together explain almost 99% of the total variation in the data. The rest of the components have no significant contribution.

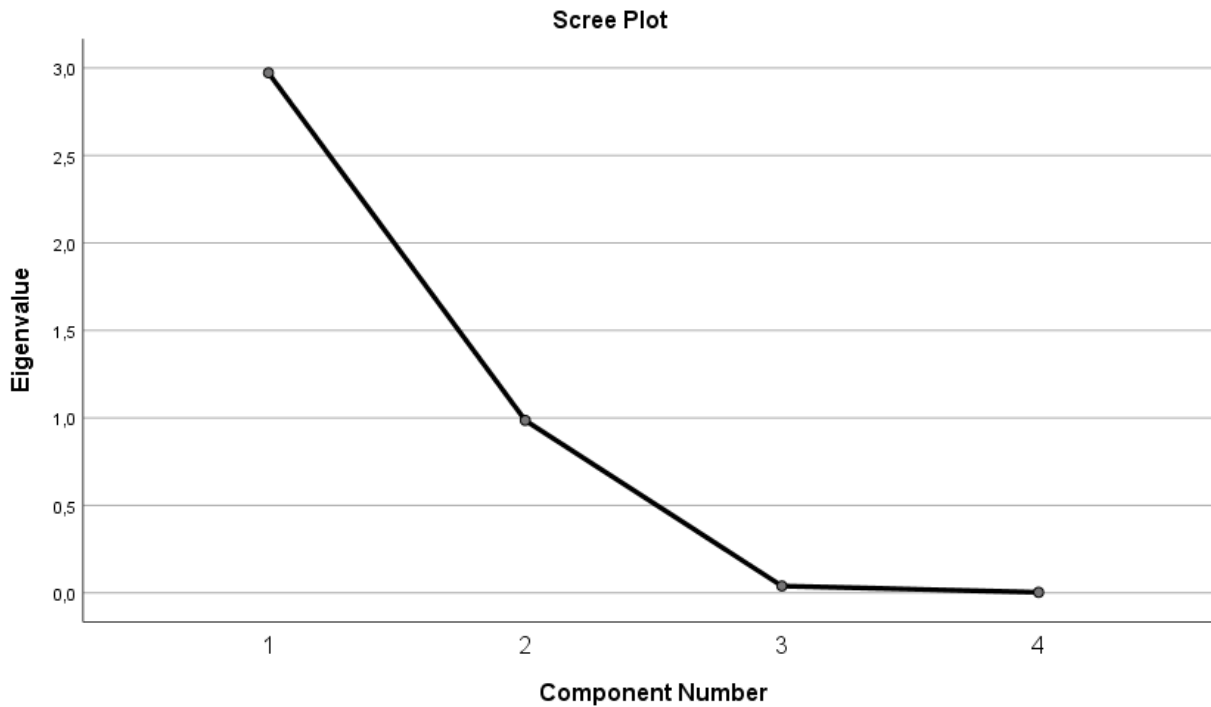


Figure 1: Scree Plot: Principal Components Analysis for the Regional Underground Economy
 Source: Author processing, results obtained through the computer program SPSS version 26.

The presented Scree Plot graph illustrates the "Eigenvalue" value for each component extracted within the factor analysis. This plot is used to determine how many principal components are sufficient to explain a large portion of the variation in the data.

Description of the Scree Plot Graph. Y-Axis: Eigenvalue – This represents the amount of variation explained by each component. X-Axis: Component Number – Each component extracted from the factor analysis.

In this graph, we notice the following: Component 1 has an "Eigenvalue" of about 3, indicating that this component explains a large part of the variation in the data. Component 2 has a smaller, but still significant Eigenvalue, just below 1. Component 3 and Component 4 have very small values (close to 0), indicating that they do not contribute significantly to the total variation in the data.

The first two components are essential for explaining the underground economy at the regional level, as they explain most of the variation in the data, according to the high "Eigenvalue" values. Component 1 appears to be most important, explaining almost all of the variation in the data. Component 3 and Component 4 do not make a significant contribution and can be further ignored in the interpretation of the results.

The Scree Plot confirms that the first two components are sufficient to explain the major variations in the economic data regarding the underground economy at the level of the analyzed regions.

Table 5: Matrix of components for key economic indicators in Principal Component Analysis

Table 5: Component Matrix

	Component	
	1	2
Regional gross domestic product	,997	,041
The turnover officially declared by the companies	,989	,093
Average number of employees	,987	,024
Income-expenditure discrepancy index	-,158	,987

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Source: Author processing, results obtained through the SPSS computer program

The Component Matrix table presents the factor coefficients for the two components extracted in the Principal Component Analysis. The coefficients indicate the correlation between the observed variables and the extracted components.

Regional gross domestic product (GDP):Component 1: Coefficient of 0.997, suggesting a very strong correlation between GDP and Component 1.Component 2: Coefficient of 0.041, indicating a weak correlation with Component 2.

Turnover officially declared by companies (CADOC):Component 1: Coefficient of 0.989, indicating a very strong correlation with Component 1. Component 2: Coefficient of 0.093, indicating a weak correlation with Component 2.

Average number of employees (NMS):Component 1: Coefficient of 0.987, indicating a strong correlation between NMS and Component 1. Component 2: Coefficient of 0.024, suggesting a non-significant correlation with Component 2.

Income-expenditure discrepancy index (IDVC):Component 1: Coefficient of -0.158, indicating a weak and negative correlation between IDVC and Component 1. Component 2: Coefficient of 0.987, suggesting a very strong correlation between IDVC and Component 2.

Component 1 has a very strong correlation with GDP, CADOC and NMS, suggesting that this component largely captures the variation related to these leading economic indicators. Therefore, Component 1 can be interpreted as a strong economic factor that reflects the economic performance recorded in the regions.

Component 2 is dominated by the Income-Expenditure Discrepancy Index (IDVC), with a very high coefficient of 0.987. This suggests that Component 2 mainly reflects issues related to discrepancies between income and expenditure, and is not as influenced by other economic indicators.

In conclusion, the first two components extracted reflect two major dimensions of the regional underground economy: one related to the economic performance indicators (PIBR, CADOC, NMS) and the other related to the discrepancies between revenues and expenditures (IDVC).

Table 6: Matrix of rotated components for key economic indicators in ACP
Rotated Component Matrix

	Component	
	1	2
Regional gross domestic product	,996	-,065
The turnover officially declared by the companies	,993	-,012
Average number of employees	,984	-,081
Income-expenditure discrepancy index	-,053	,999

Extraction Method: Principal Component Analysis.
 Rotation Method: Varimax with Kaiser Normalization.a
 a. Rotation converged in 3 iterations.

Source: Author processing, results obtained through the SPSS computer program

The Rotated Component Matrix table shows the results obtained from principal component analysis (PCA) and component rotation using the Varimax method with Kaiser normalization. This matrix shows how much each variable contributes to each of the two extracted components.

Regional gross domestic product (GDP): Component 1: Coefficient of 0.996, indicating a very strong correlation between PIBR and Component 1. Component 2: Coefficient of -0.065, suggesting a weak and negative relationship between PIBR and Component 2.

Turnover officially declared by companies (CADOC): Component 1: Coefficient of 0.993, indicating a very strong correlation between CADOC and Component 1. Component 2: Coefficient of -0.012, indicating an almost negligible relationship with Component 2.

Average number of employees (NMS): Component 1: Coefficient of 0.984, suggesting a very strong correlation between NMS and Component 1. Component 2: Coefficient of -0.081, indicating a weak and negative correlation with Component 2.

Income-expenditure discrepancy index (IDVC): Component 1: Coefficient of -0.053, indicating a very weak and negative correlation between IDVC and Component 1. Component 2: Coefficient of 0.999, indicating an extremely strong correlation between IDVC and Component 2.

Component 1 is strongly correlated with PIBR, CADOC, and NMS. This suggests that this component is dominated by economic performance indicators at the level of regions. Thus, Component 1 can be interpreted as a major economic factor that reflects the overall economic health of regions, based on economic output and labor force.

Component 2 is almost entirely associated with the Income-Expenditure Discrepancy Index (IDVC), having a correlation of 0.999. This indicates that Component 2 represents aspects related to financial imbalances, more precisely the discrepancies between revenues and expenditures in the analyzed regions.

The table suggests that the underground economy at the level of the regions can be interpreted along two major dimensions:

Component 1: Focus on economic performance, given strong linkages with PIBR, CADOC and NMS.

Component 2: Aspects related to financial imbalances, in particular reflected by the income-expenditure discrepancy index.

This shows that regions are characterized by both economic performance and financial balance, with each component capturing these two distinct dimensions of the underground economy.



Figure 2: Plot of Components in Rotated Space: Representation of Latent Components in the Underground Economy
 Source: Author processing, results obtained through the computer program SPSS version 26.

The "Component Plot in Rotated Space" graph represents the distribution of variables in the space of rotated principal components, the result of principal component analysis (PCA). This plot shows how each variable is positioned on the two extracted components after rotation.

Axis Component 1: This is the first main component. The variables that are located closer to the right end of the axis have a high load on Component 1. In this case, the variables "PIBR" (Regional Gross Domestic Product), "NMS" (Average Number of Employees), "CADOC" (Figure of Business Officially Declared by Companies) are all positioned on the right-hand side, suggesting that these variables contribute substantially to this component.

Axis Component 2: This is the second main component. The variable "IDVC" (Index of Discrepancy between Income and Expenditure) has a high loading on Component 2, being placed near the top of the vertical axis.

Arrangement of variables: The variables PIBR, NMS and CADOC are positioned very close to each other and to the Component 1 axis, suggesting that they are highly correlated with each other and contribute significantly to the first component. IDVC is separate from the other variables, placing at the top of the

Component 2 axis, suggesting that this variable is a major contributor to the second component and is relatively independent of the other variables.

The graph indicates two major dimensions of the underground economy at the level of the regions. The first component (Component 1) appears to be strongly influenced by general economic variables (PIBR, NMS, CADOC), while the second component (Component 2) is mainly defined by the variable related to financial discrepancies (IDVC). These components are distinct and provide a different perspective on the underground economy depending on the region.

Table 7: Matrix of Coefficients of Component Scores for Key Economic Indicators in ACP

Table 7: Component Score Coefficient Matrix

	Component	
	1	2
Average number of employees	.333	-.011
Income-expenditure discrepancy index	.053	1,002
Regional gross domestic product	.338	.006
The turnover officially declared by the companies	.341	.058

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Component Scores.

Source: Author Processing, results obtained through the SPSS computer program

The "Component Score Coefficient Matrix" table provides the coefficients that are used to calculate the scores of each principal component for each region based on the variables in the analysis.

Component 1 and Component 2 represent the two main dimensions identified by principal component analysis (PCA) after Varimax rotation. Each variable has a coefficient for each component, which indicates the relative importance of that variable within that component.

Average number of employees: It contributes positively to Component 1 with a coefficient of 0.333 and negatively, but very weakly, to Component 2 with a coefficient of -0.011. This indicates that the average number of employees is important for the first component and has an insignificant influence on the second component.

Income-expenditure discrepancy index (IDVC): This variable has a small contribution to Component 1 with a coefficient of 0.053, but strongly dominates Component 2 with a coefficient of 1.002. So, this variable almost completely defines Component 2, being the most influential in this dimension.

Regional gross domestic product (GDP): It contributes significantly to Component 1 with a coefficient of 0.338 and has very little influence on Component 2 with a coefficient of 0.006. Regional GDP is therefore an essential variable in Component 1.

Turnover officially declared by companies (CADOC): It has the greatest influence on Component 1 with a coefficient of 0.341, and contributes moderately to Component 2 with a coefficient of 0.058.

Component 1 is dominated by general economic variables, such as the average number of employees, regional GDP and turnover declared by companies. These variables are essential for the description of this dimension and indicate the official economic performance of the regions.

Component 2 is almost entirely dominated by the Income-Expenditure Discrepancy Index (IDVC), which reflects underground savings or discrepancies between declared income and expenditure.

This table suggests that the underground economy at the regional level can be described in terms of two major dimensions: one related to official economic performance and another related to financial discrepancies.

Table 8: Covariance Matrix of Component Scores in Principal Component Analysis

Component Score Covariance Matrix

Component	1	2
1	1,000	,000
2	,000	1,000

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

Component Scores.

Source: *Author Processing, results obtained through the SPSS computer program*

The "Component Score Covariance Matrix" table presents the covariance matrix of the principal component scores, resulting from PCA analysis and Varimax rotation. This matrix is used to assess the relationships between the scores obtained for the two principal components.

Top Left Item (Component 1 - Component 1): The value is 1.000, indicating that Component 1 scores have an internal variability of 1, which is normal in a principal component analysis. This suggests that component 1 scores are standardized.

Bottom Right Item (Component 2 - Component 2): The value is also 1.000, indicating that Component 2 scores are also standardized and have a variability of 1.

Bottom Left and Top Right Item (Component 1 - Component 2 and Component 2 - Component 1): Both values are 0.000, suggesting no covariance between the two component scores. In other words, the two components are orthogonal (uncorrelated with each other). This is a common expectation when using an orthogonal rotation method such as Varimax.

The two principal components (Component 1 and Component 2) are uncorrelated with each other, indicating that each component captures distinct information from the original variables. In the context of the underground economy at the level of regions, this suggests that there are two major dimensions, independent of each other: one related to the formal economy and another related to economic discrepancies (which could be associated with the underground economy).

The table confirms that the PCA analysis and Varimax rotation were able to isolate two independent dimensions, which is useful for interpreting the results in the context of the underground economy.

Evaluation of the main poles of the underground economy in regional profile in 2022

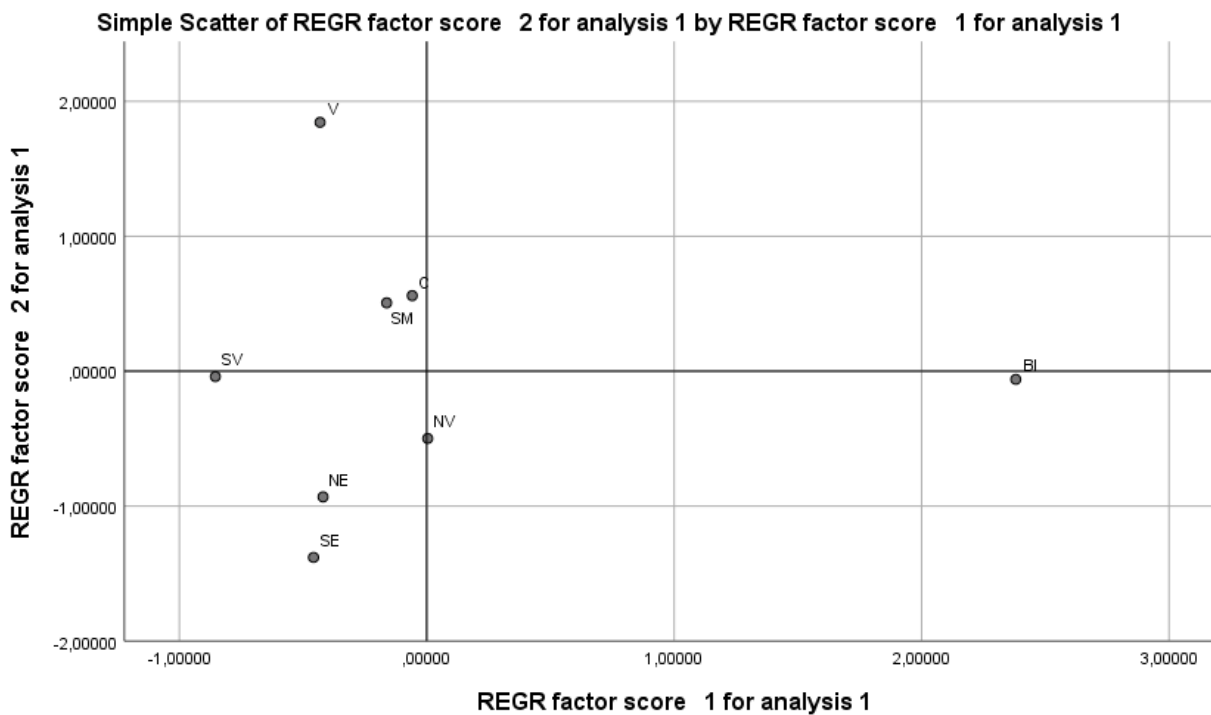


Figure 3: Global trends in underground economy regions

Source: Author processing, results obtained through the computer program SPSS version 26.

The graph shown in the image "Simple Scatter of REGR factor score 2 for analysis 1 by REGR factor score 1 for analysis 1" is a scatter graph that shows the positioning of the different regions according to the two principal components obtained from the PCA analysis.

Axes: Component 1 (horizontal axis) represents the first principal component that captures the major variability in the data set. Component 2 (vertical axis) represents the second principal component that captures the remaining variability independent of the first component.

Positioning regions on the graph:

Regions on the right side of the graph (with high scores on Component 1) show a strong correlation with official economy indicators such as regional Gross Domestic Product, average number of employees and declared turnover. These regions tend to have a strong formal economy and less activity associated with the underground economy. BI (positioned on the extreme right) indicates a region highly correlated with the official economy.

Regions on the left side of the graph (with low scores on Component 1) suggest a lower degree of formal economic development and could be more correlated with the underground economy. NE and SE are in this area, suggesting greater economic discrepancies and a possible stronger presence of the underground economy.

Regions placed above the horizontal axis (with positive scores on Component 2) may indicate discrepancies between revenues and expenditures in the region, more associated with the underground economy. V and SM show such characteristics, perhaps indicating greater informal economic activities.

Regions placed below the horizontal axis (negative scores on Component 2) indicate a lower correlation with economic discrepancies, suggesting a balance between formal income and expenditure. NE and SE are examples of regions with negative scores on Component 2, indicating economies less affected by income-expenditure discrepancies.

The graph provides a picture of the underground economy at the regional level through principal component analysis. The positioning of the regions reflects the degree of discrepancy between the formal and informal economy. Regions on the right suggest a stronger formal economy, while those on the left may indicate a greater presence of the underground economy.

The Proximity Matrix table (Appendix 3) presents a matrix of squared Euclidean distances between the various regions analyzed within the underground economy. These values indicate the degree of dissimilarity or distance between each pair of regions in terms of the analyzed indicators.

Description of the Proximity Matrix table: 1: NV - Represents one of the analyzed regions. 2: C - Another region included in the analysis. Array Values - The array values represent the squared Euclidean distances between each pair of regions. The higher the value, the more different the regions are in terms of the analyzed economic characteristics.

Zero Diagonal: The value on the diagonal is always zero because it represents the distance between a region and itself.

Large distances: For example, the distance between the region NW (case 1) and BI (case 6) is very large (578,403,659,111,440), which suggests a significant difference between these two regions in terms of the economic indicators considered.

Small distances: In contrast, the distance between V (case 8) and SE (case 4) is relatively small (2,264,992,850,210), suggesting that these two regions are more similar in terms of the analyzed indicators.

Moderate distances: For example, the distance between NW (case 1) and W (case 8) is 41,866,717,307,996, indicating a moderate difference between these regions.

Regions with large values: Regions with large distances between them (such as NV and BI) can have very different economies in terms of formal and informal economic activity, income, expenditure and other economic indicators.

Regions with small values: Regions with smaller distances (such as W and SE) suggest similarities in the underground economy, and it is possible that these regions face similar economic challenges.

The proximity matrix gives us a clear understanding of how different regions relate to each other from the perspective of the analyzed economic indicators. This can help to identify groups of regions that share similar characteristics and that, consequently, may require similar economic policies to combat the underground economy.

Table 9: Clustering Program for Combining Clusters in Hierarchical Analysis

Agglomeration Schedule						
Interns hip	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	4	8	2264992850.210	0	0	3
2	1	2	3029977864.109	0	0	5

3	3	4	4065634988,187	0	1	4
4	3	5	7894649117,842	3	0	5
5	1	3	20758215566,433	2	4	6
6	1	7	53828657978.508	5	0	7
7	1	6	781155535280,323	6	0	0

Source: Author Processing, results obtained through the SPSS computer program

The Agglomeration Schedule table describes the process of agglomeration (hierarchical agglomeration) of regions in clusters, using the distances between them according to the economic indicators considered in the analysis of the underground economy. This process helps group similar regions into clusters to highlight similarities and differences between them.

Description of the table Agglomeration Schedule: Stage: The table presents several stages of agglomeration, each stage involving the combination of two distinct clusters/regions into one.

Cluster Combined: At each step, two clusters (or individual regions) are combined into a new cluster. For example, at stage 1, regions 4 and 8 are combined into one cluster.

Coefficients: The coefficient associated with each stage represents the distance or dissimilarity between the clusters combined at that stage. As the stages progress, the coefficients increase, indicating that the combined regions/clusters are less and less similar. For example, at stage 1, the coefficient is 2,264,992,850.210, which indicates a relatively small distance between regions 4 and 8. In contrast, at the final stage (stage 7), the coefficient is 781,155,535,280.323, which indicating a significant dissimilarity between the last combined clusters.

Stage Cluster First Appears: This column indicates at which stage each cluster involved in the combination first appeared. For example, at stage 5, clusters 1 and 3 are combined, where cluster 1 first appeared at stage 2 and cluster 3 appeared at stage 4.

Next Stage: This column shows at what stage the merged cluster will be involved in another merge. For example, the cluster formed in step 1 (4 and 8) will be combined again in step 3.

Similar regions: Regions 4 and 8 are combined in the first stage, suggesting that they have the most similar economic characteristics in the context of the underground economy. Similarly, regions 1 and 2 are combined at step 2, indicating a strong similarity between them.

Increasing distance between clusters: As the clustering stages progress, the coefficients increase significantly, suggesting that as clusters are combined, regions become increasingly different. For example, at the final stage, the coefficient is extremely high, indicating a substantial difference between the last clusters combined.

In the early stages, regions with similar underground economies quickly combine, and in the later stages, the differences between the groups become more obvious. This suggests the existence of regional variations in the degree of dissimilarity of the underground economy in various areas of the country.

The Agglomeration Schedule provides a clear picture of how regions are grouped based on their economic similarities, particularly in terms of the underground economy. Early stages indicate regions that are very similar, while later stages highlight regions that are more economically distinct. This information can be useful for the development of regional economic policies and for identifying areas that require special attention in combating the underground economy.

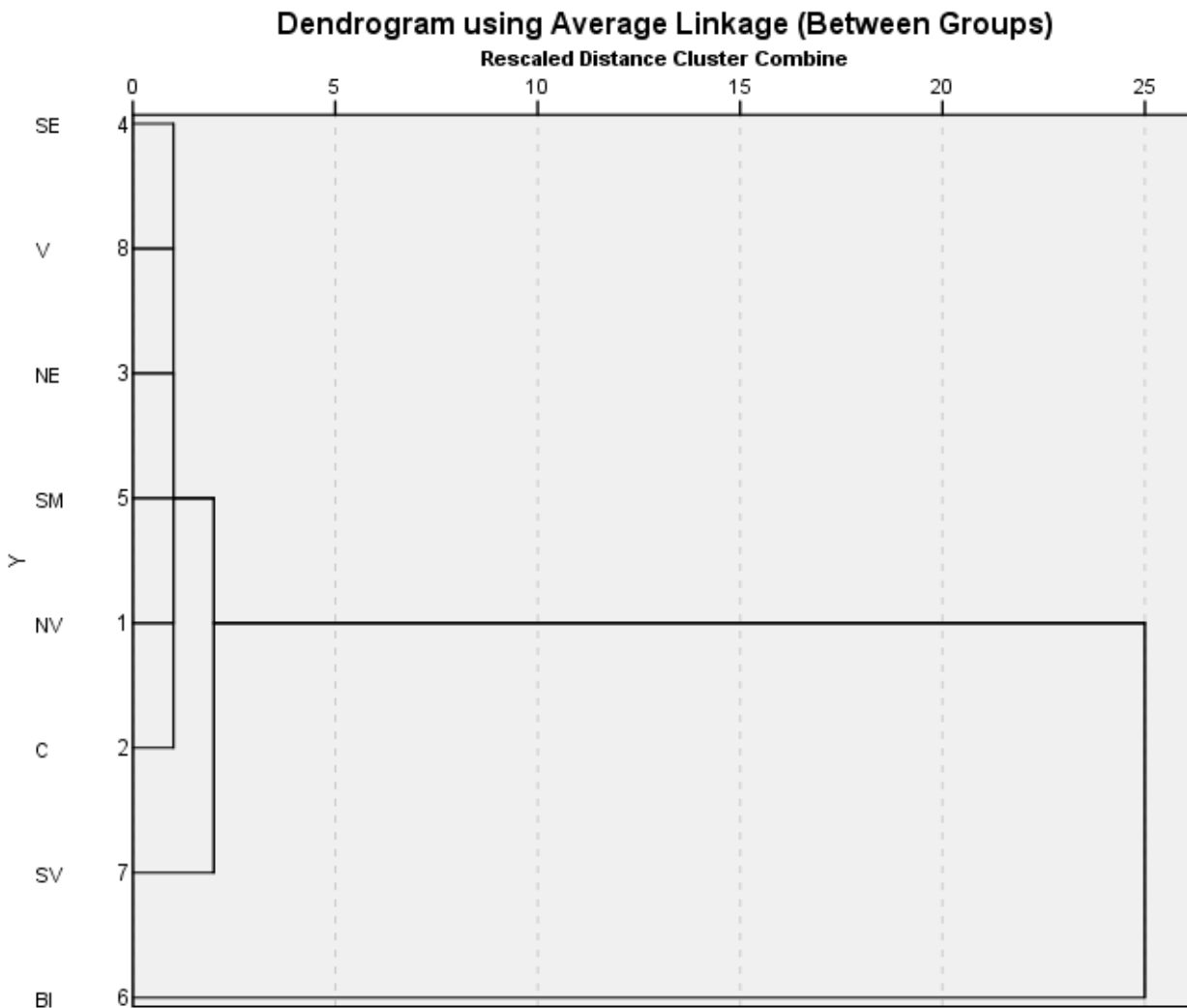


Figure 4: Dendrogram of the performance of the underground economy in regional profile in 2022

Source: Author processing, results obtained through the computer program SPSS version 26

This dendrogram provides a visual representation of the ranking of regions according to the similarities between them, based on the economic indicators related to the underground economy.

Binding method used: The dendrogram uses the "Average Linkage (Between Groups)" linkage method to combine regions. This means that the average distances between all pairs of points in two clusters are used to decide when to combine two clusters.

Similarity between economic regions: The dendrogram shows the order in which the regions were grouped, starting from the most similar to the most different, based on the underground economy. It is observed that the SE (South-East), W (West), NE (North-East), and SM (Satu Mare) regions are the first to combine with each other, suggesting that they have quite similar underground economies in relation to the indicators use. The BI (Bihor) region, on the other hand, is distinct and does not combine with other regions except at a greater distance, suggesting that it has unique characteristics in terms of the underground economy.

Combination order: To the left of the dendrogram, regions are combined according to their economic similarity. The lower the connecting line is in the dendrogram, the more similar the respective regions are. As we move up the dendrogram, regions are combined at greater distances, indicating that they are less similar to each other. Closest to the region in terms of the underground economy appear to be the SE and W, which are clustered very early in the process, followed by the NE and SM regions.

Distances between clusters: The rescaled distance between clusters is represented on the horizontal axis (Rescaled Distance Cluster Combine). The greater the distance between the clusters, the more economically different the combined regions are from each other. BI (Bihor) only combines at a large distance, suggesting that it has a very different underground economic profile than the other regions.

The dendrogram suggests that the SE, W, NE, and SM regions are similar in terms of economic indicators related to the underground economy, while BI (Bihor) appears to be a distinct region in this context. This may indicate that economic policy measures that work well in some regions may not be as effective in others, especially in regions with unique underground economic profiles.

Conclusions:

The analysis of the underground economy at the level of the regions in Romania, carried out by means of the method of factor analysis and the main components, highlights a series of important conclusions regarding the distribution and characteristics of the underground economy in the various regions of the country. First of all, the results suggest the existence of significant discrepancies between Romania's regions in terms of formal and underground economic activity. Economically more developed regions, with a higher regional gross domestic product (GDP) and a higher average number of employees, have a smaller underground economy.

These regions are characterized by robust formal economic activity, where the turnover officially declared by companies is significant. This finding suggests that regions with stronger formal economies have a better ability to limit the expansion of the underground economy.

In contrast, less developed regions, which have a lower regional GDP and a lower average number of employees, tend to be more affected by the underground economy. These regions are characterized by major discrepancies between declared income and expenditure, indicating a greater presence of the informal economy. In particular, the index of the discrepancy between income and expenses proves to be an essential indicator for identifying these phenomena.

The principal components analysis highlighted two major dimensions of the underground economy in Romania: one related to general economic performance, including regional GDP and official turnover, and another related to economic imbalances, measured by the income-expenditure discrepancy index. These dimensions suggest that the underground economy is not only a problem of general economic development, but is also influenced by deeper aspects, such as the internal financial balance of the regions.

Another important aspect revealed by this research is the fact that economic policy measures must be adapted to the specifics of each region. For example, regions with strong formal economies could benefit from measures to strengthen and protect this formal economy, while regions affected by economic disparities and informal activities would require more aggressive policies to reduce the underground economy and promote business formalization .

In conclusion, this study provides a detailed understanding of how the underground economy manifests itself in Romania at a regional level and underlines the need for well-directed economic policies that take into account the economic realities specific to each region. A differentiated approach, adapted to local conditions, is essential to reduce the magnitude of the underground economy and to stimulate sustainable economic development in Romania.

Bibliography:

- Statistical Yearbook of Romania, National Institute of Statistics, Bucharest, 2021.
- Ciutacu, C., Chivu, L., Guga, Ş., **Underground economy and the labor market in Romania**, Economic Publishing House, Bucharest, 2020.
- Dobre, I., Alexandru, A., Păuna, B., **Analysis of the phenomenon of the underground economy in Romania**, Revista Română de Economie, Vol. 37, No. 2, 2021, pp. 65-80.
- Georgescu, G., **Methods of measuring the underground economy: applications for Romania**, Romanian Academy Publishing House, Bucharest, 2019.
- Ghizdanu, I., **Evaluation of the size of the underground economy in the regions of Romania**, Center for Economic Research and Regional Analysis, Bucharest, 2022.

-
- Williams, CC, **The Shadow Economy**, **International Journal of Social Economics**, Vol. 48, No. 6, 2021, pp. 764-784.
 - **SPSS Statistical Analysis Software Manual**, Version 26.0, IBM Corp, 2020.
 - National Bureau of Economic Research, **Measuring the Informal Economy: A Guide for Policy Makers**, Washington DC, 2018.
 - OECD, **Shadow Economies in Europe**, OECD Publishing, Paris, 2020.