



THE 21ST EDITION OF THE INTERNATIONAL CONFERENCE EUROPEAN INTEGRATION REALITIES AND PERSPECTIVES

A Principal Component Analysis for the Social Sustainable Development Goals Indicators Framework

Alina Georgeta Ailincă¹, Gabriela Cornelia Piciu²

Abstract: Sustainable development is a theoretical construct with more and more supporters in the last three decades, be it in the media, the political system, researchers or the public policy environment. The Sustainable Development Goals generally propose an approach to a specific sector through numerous indicators, with interferences often existing between objectives and indicators. An analysis focused on a certain pattern of indicators (e.g. economic, social, environmental), for a critical destination can better channel efforts to achieve it starting from the goal, rather than from the manner of intervention or the path to be followed, and eliminating as much as possible the elements of interference with an ambiguous, multi-univocal character. Therefore, in addition to clarifying the conceptual framework, the study methodologically attempts to approach the Sustainable Development Goals of a social nature based on the Principal Components Analysis (PCA). Thus, based on a multivariate analysis, mainly SDGs indicators 4, 5, 10 and 17 are analyzed, starting from Eurostat data, for the period 2012-2025. The use of a large number of variables for the European Union countries (EU27) for each target and for the social field as a whole, against the background of the complexity of the indicators and the different units of measurement for each of them, revealed the choice of the Principal Components Analysis methodology as appropriate. The results show the strengths and weaknesses of the social SDGs starting from the PCA analysis. The paper, through the original systematization proposed, can be considered a landmark in the analysis of the social SDGs in order to better identify the gaps that need to be corrected in the analysis of the SDGs.

Keywords: sustainability; statistical description; Principal component analysis (PCA)

JEL Classification: C38; O15; R13

1. Introduction

Current theoretical and practical frameworks increasingly and deeply focus on the human being and his contemporary needs. In addition to the basic elements of a better life, good education, equal rights and the reduction of discrimination, the reduction of social inequalities and the constant harmonization of policies with human needs represent not only declarative, theoretical elements, but also deep anchors for legislation and policies well-grounded in what the future of human society entails. In recent decades, the Sustainable Development Goals (SDGs) have received increasing attention both in academic literature and in practical implementation, which has led to the development of dedicated assessment tools and performance indicators. In this context, Principal Component Analysis (PCA), through an

¹ PhD., 3rd degree scientific researcher, “Victor Slăvescu” Centre for Financial and Monetary Research, INCE, Romanian Academy, Bucharest, Romania, Corresponding author e-mail address: alinageorgetaailinca@gmail.com.

² PhD., 2nd degree scientific researcher, “Victor Slăvescu” Centre for Financial and Monetary Research, INCE, Romanian Academy, Bucharest, Romania, E-mail address: gabriela_piciu@yahoo.com.



orthogonal transformation, can serve as a valuable statistical technique for examining complex and large data sets, offering the identification of patterns, trends and discrepancies related to the indicators of the Sustainable Development Goals. In addition to reducing dimensionality, Principal Component Analysis (PCA) provides information about the relationships between variables and underlying patterns in the data set and allows the elimination of irrelevant and redundant features induced by multicollinearity.

Thus, the aim of the article is to reduce dimensionality by eliminating variations, correlating factors through PCA analysis, in order to demonstrate how the social elements within the SDGs connect and potentiate each other. The paper is ordered as follows: first, the general framework of analysis is described, then the related literature and the statement of the problem, the methodology used and the description of the data, then in the second part the analysis of econometric and statistical tests is carried out, and finally the conclusions are presented.

2. Related Work

The analysis of the social dimension in the framework of the Sustainable Development Goals has increasingly relied on composite indicators and multivariate statistical techniques to capture complex and interconnected phenomena, often capturing shortcomings and controversial elements, such as the risk of oversimplification (Saltelli, 2007). In contexts where direct data on income or expenditure are limited or unreliable, early contributions, such as those of Filmer & Pritchett (2001), demonstrate how asset-based indices, derived through Principal Component Analysis, can serve as important indicators of socio-economic status. This approach has been further refined by Vyas & Kumaranayake (2006), who provide a methodological framework for the construction and interpretation of PCA-based indices in social and health research. Beyond the construction of indices, the literature highlights important methodological considerations regarding the aggregation, weighting and interpretability of multidimensional indicators. Studies such as Decancq & Lugo (2010) highlight the normative and empirical implications of weighting schemes in composite indices, while Greco et al. (2019) provide a comprehensive analysis of the methodological frameworks used in the construction of such indices, including the role of dimensionality reduction techniques such as PCA. These contributions highlight the importance of transparency and robustness when combining heterogeneous social indicators into a single measure. In parallel, broader assessments of development measurement, such as Booyesen's (2002), provide critical insights into the strengths and limitations of composite indices in capturing social progress, including that composite indices satisfactorily capture cross-national availability but less satisfactorily capture intertemporal availability. These insights are particularly relevant in the context of SDG monitoring, where multidimensionality is a defining feature. Institutional reports, including methodological ones, produced by the European Commission, OECD, United Nations, etc. (Saisana & Tarantola, 2002; Nardo et al., 2008; OECD, 2018; Sachs et al, 2018; OECD, 2024; UNDP, 2022; Sachs et al., 2023; Hijman, R. et al., 2025) further demonstrate how composite indicators are operationalized to assess countries' progress, including towards achieving the SDGs.

In general, the existing literature supports the integration of PCA as a valuable tool for reducing dimensionality (Jolliffe & Cadima, 2016) and identifying latent structures within SDG-relevant social indicators. By combining methodological rigor with practical applicability, these studies provide a solid basis for empirical analyses that aim to assess social sustainability and inequality through data-driven approaches.

3. Problem Statement

Although there is a rich literature on the Sustainable Development Goals and Principal Component Analysis, the clear linkage of the social component of the SDGs to PCA is rarely addressed in the literature in a coherent, unfragmented and standardized manner. At the same time, beyond data, statistical and econometric analyses do not have the capacity to fully capture the dynamics and causality changes in social sustainability over time. PCA is often used to construct socio-economic indices, but there are no integrated frameworks that link the use of PCA to the social Sustainable Development Goals in a way that explicitly connects PCA indicators to policy-relevant interpretations, reducing their practical utility for decision-making and subsequent reflection in improving social statistics of the SDGs.

4. Solution Approach

This paper explores the interconnections between the indicators of the social Sustainable Development Goals by implementing Principal Components Analysis (PCA). The paper thus aims to analyze the indicators of the social Sustainable Development Goals, in particular SDG4, SDG5, SDG10 and SDG17. To these, where expressly provided, some relevant indicators from other SDGs (e.g. SDG1 and SDG8) have been added. The methodology aims to apply a principal components analysis (PCA), using a linear orthogonal method so that the data set can adjust its size. The analysis is carried out for the period 2012-2025, and the initial data belong to Eurostat. The approach is based on a panel data set for the European Union (EU27). It should be noted that, by using various data completion and remediation techniques (e.g. proxies, interpolation, extrapolation), the results should be viewed with caution.

Since the variables included in the analysis have different measurements, even if they come from the same domain (social sphere), a relatively simple normalization of the data sets was previously performed, taking into account their positive or negative significance:

- for indicators with positive significance, the formula is:

$$N_i = ((X_i - X_{\min}) / (X_{\max} - X_{\min})) \quad (1)$$

- for indicators with negative significance, the formula is:

$$N_i = ((X_{\max} - X_i) / (X_{\max} - X_{\min})) \quad (2)$$

Notations: N_i - normalization of the value for year i of the selected indicator, X_i - current value of the indicator in year i , $X_{\max}$ – maximum value of the indicator, $X_{\min}$ - minimum value of the indicator.

Table 1 presents the selected initial variables.

Table 1. Variables used for social SDGs

Name of Variable	Abbreviation	Measure Unit	Source of Data
Share of persons at risk of poverty or social exclusion by degree of urbanisation (Cities)	SPRPU	Percentage	Eurostat, [sdg_01_10a]
Persons at risk of monetary poverty after social transfers by citizenship	PRMPAST	Percentage	Eurostat, [sdg_01_20a]
Early leavers from education and training by sex	ELETS	Percentage	Eurostat, [sdg_04_10]
Persons aged 25-34 with tertiary educational attainment level	PATEA	Percentage	Eurostat, [sdg_04_20]

Participation in early childhood education by sex (children aged 3 and over)	PECE	Percentage	Eurostat, [sdg_04_31]
Low achieving 15-year-olds in reading, mathematics or science	LARMS	Percentage	Eurostat, [sdg_04_40]
Adult participation in learning by sex	APL	Percentage	Eurostat, [sdg_04_60]
Gender pay gap in unadjusted form	GPGUF	Percentage	Eurostat, [sdg_05_20]
Gender employment gap, by type of employment	GEGTE	Percentage point	Eurostat, [sdg_05_30]
Persons outside the labour force due to caring responsibilities by sex	POLFDCR	Percentage of total population	Eurostat, [sdg_05_40]
Seats held by women in national parliaments and governments	SHWNPG	Percentage of women	Eurostat, [sdg_05_50]
Positions held by women in senior management	PHWSM	Percentage	Eurostat, [sdg_05_61]
Young person's neither in employment nor in education and training (NEET) by citizenship	YPNENET	Percentage	Eurostat, [sdg_08_20a]
Employment rate by citizenship	ERC	Percentage	Eurostat, [sdg_08_30a]
Purchasing power adjusted GDP per capita	PPAGDP	Percentage	Eurostat, [sdg_10_10]
Relative median at-risk-of-poverty gap	RMARPG	Percentage	Eurostat, [sdg_10_30]
Income distribution	ID	Ratio	Eurostat, [sdg_10_41]
Income share of the bottom 40 % of the population	ISBP	Share of national equivalised income	Eurostat, [sdg_10_50]
Asylum applications by state of procedure	AASP	Per million inhabitants	Eurostat, [sdg_10_60]
Official development assistance as share of gross national income	ODASGNI	Percentage of gross national income (GNI)	Eurostat, [sdg_17_10]
High-speed internet coverage, by type of area	HSICTA	Percentage of households (Rural areas)	Eurostat, [sdg_17_60]

Source: Eurostat indicators; authors' selection

For the adequacy of the sampling, a Kaiser-Meyer-Olkin (KMO) test is used, as it can guide us in the selection of the most relevant variables, allowing us to select the variables that we can consider the most important for a subsequent analysis within the Principal Components Analysis (PCA). The interpretation of KMO is that a value in the range of 0.50-0.60 can be considered poor, above 0.60 but up to 0.70 inclusive can be considered acceptable, and above 0.70 to 0.80 good and very good if it is above 0.80.

In the first phase we can see that there is an acceptable adequacy of the sampling, but that some variables can be excluded (especially if they are below 0.60) for an even better relevance of the analysis. Thus, the chosen model is based on KMSA 2, and the variables selected (16) for PCA analysis are the following: SPRPU, PRMPAST, ELETs, PATEA, PECE, APL, GEGTE, SHWNPG, PHWSM, YPNENET, ERC, PPAGDP, RMARPG, ID, ISBP, ODASGNI.

Table 2. Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMSA) for social SDGs

	KMSA 1	KMSA 2
SPRPU	0.764327	0.826938
PRMPAST	0.810302	0.855069
ELETS	0.631948	0.619515
PATEA	0.755887	0.762101
PECE	0.685395	0.758128
LARMS	0.565792	-
APL	0.867116	0.890924
GPGUF	0.395867	-
GEGTE	0.780809	0.803868
POLFDCR	0.484028	-
YPNENET	0.816184	0.798558
ERC	0.711199	0.742789
PPAGDP	0.786039	0.788971
RMARPG	0.867415	0.868280
ID	0.742655	0.753794
ISBP	0.709504	0.713923
AASP	0.521467	-
ODASGNI	0.848585	0.847346
HSICTA	0.597973	-
PHWSM	0.741188	0.691250
SHWNPG	0.797061	0.771659
Kaiser's MSA	0.743912	0.784596

Source: Eurostat indicators; authors' selection

5. Analysis of Results

Descriptive statistics are presented for the selected variables (only 16 variables from the initial ones) such as mean, median, standard deviation, Kurtosis and Skewness. Given the panel distribution of the analyzed data, the Jarque-Bera statistic may prove to be less relevant, but nevertheless we observe that the values of the variables for this test are different from zero, and the probabilities demonstrate that they are also mostly different from zero. If we assume that the null hypothesis requires Kurtosis to be 3 and Skewness to be 0, we can observe that with some exceptions the null hypothesis is respected. Thus, Skewness is generally positive and negative, around the value 0, above 1 being only the indicators PECE, APL, ODASGNI and PPAGDP, and Kurtosis is positive and generally below 3 or around this value, the exception being PECE and PPAGDP with extremely high values, indicating without these variables, generally a platykurtic distribution.

If we consider that for a normal distribution the mean and median values should be close, suggesting a relatively symmetrical distribution of the series, the null hypothesis is confirmed. At the same time, the standard deviation also oscillates around the mean and median values, being below them, thus indicating that the values are tightly grouped in a narrow range around the mean.

Table 3. Selective descriptive statistics for social SDGs indicators

	Mean	Median	Std. Dev.	Skewness	Kurtosis	Jarque-Bera	Probability	Sum	Sum Sq. Dev.
SPRPU	0.4897	0.5000	0.2046	-0.1383	2.5147	4.9134	0.0857	185.1091	15.7831
PRMPAST	0.5251	0.6006	0.2353	-0.4038	2.1249	22.3317	0.0000	198.4941	20.8739
ELETS	0.6908	0.7225	0.1743	-0.9082	3.9952	67.5664	0.0000	261.1233	11.4493
PATEA	0.4382	0.4357	0.2164	0.1360	2.5385	4.5196	0.1044	165.6546	17.6505
PECE	0.7873	0.8272	0.1694	-1.3547	4.9545	175.7831	0.0000	297.6057	10.8238
APL	0.2943	0.2388	0.2150	1.0521	3.5479	74.4650	0.0000	111.2639	17.4203
GEGTE	0.6808	0.7219	0.1695	-0.9718	3.8926	72.0477	0.0000	257.3375	10.8265
SHWNPG	0.4937	0.4792	0.2365	0.0280	2.0969	12.8946	0.0016	186.6210	21.0894
PHWSM	0.4446	0.4272	0.2293	0.3607	2.2312	17.5055	0.0002	168.0610	19.8252
YPNENET	0.6357	0.6688	0.2052	-0.7318	3.0534	33.7797	0.0000	240.2890	15.8730
ERC	0.6394	0.6725	0.2059	-0.5606	2.8715	20.0598	0.0000	241.6967	15.9848
PPAGDP	0.2226	0.1876	0.1761	2.1425	8.5235	769.7110	0.0000	84.1406	11.6859
RMARPG	0.6328	0.6600	0.2096	-0.5121	2.6014	19.0218	0.0001	239.2033	16.5557
ID	0.6626	0.7127	0.2137	-0.7838	2.8577	39.0238	0.0000	250.4565	17.2089
ISBP	0.5316	0.5663	0.2279	-0.2468	2.1923	14.1120	0.0009	200.9303	19.5844
ODASGNI	0.2390	0.1643	0.1926	1.2982	3.9164	119.3947	0.0000	90.3294	13.9818

Source: Eurostat data; EViews 9 software authors' processing; 378 number of obs.

In general, with some small exceptions, analyzing the values of the presented correlation matrix, most of the correlation coefficients between macroeconomic variables are below 0.5 and show a relatively weak positive and negative linear correlation, thus eliminating questions related to collinearity. Only, for example, ID is positively and strongly correlated with PRMPAST (0.842), ISBP is positively and strongly correlated with PRMPAST (0.847), PHWSM is positively and strongly correlated with SHWNPG (0.736), and ERC is positively and strongly correlated with YPNENET (0.792), ID is positively and strongly correlated with RMARPG (0.732), but also ISBP is positively and strongly correlated with ID (0.978). Thus, we can observe that in particular, the income distribution and especially the income distribution of the poorest 40% of the population substantially affects people at risk of monetary poverty after social transfers and the Relative median at-risk-of-poverty gap indicator, suggesting indeed that solving financial problems plays a fundamental role in social de-escalation and ensuring social sustainability. At the same time, regarding the correlation matrix, we can observe that the analyzed variables present positive correlations among themselves, suggesting a potentiation between them. At the same time, we can note as a significance that there is an equal weight between the indicators with positive significance and those with negative significance, which makes it difficult to draw a general line of analysis. However, we can still see that where there are negative correlations between the variables, these are logical and expected. For example, Seats held by women in national parliaments and governments (SHWNPG) and Positions held by women in senior management (PHWSM) are negatively but weakly correlated with Share of persons at risk of poverty or social exclusion by degree of urbanisation (SPRPU), indicating that greater women's power can contribute to reducing the risk of poverty at the urban level.

Table 4. Correlation matrix for social SDGs indicators selected

	SPRPU	PRMPAST	ELETS	PATEA	PECE	APL	GEGTE	SHWNPG	PHWSM	YPNENET	ERC	PPAGDP	RMARPG	ID	ISBP	ODASGNI
SPRPU	1															
PRMPAST	0.331	1														
ELETS	0.478	0.249	1													
PATEA	0.152	0.238	0.418	1												
PECE	-0.138	0.149	-0.235	0.258	1											
APL	0.122	0.413	0.119	0.383	0.408	1										
GEGTE	0.127	-0.002	0.272	0.401	0.288	0.421	1									
SHWNPG	-0.161	0.211	0.063	0.276	0.449	0.662	0.435	1								
PHWSM	-0.100	0.156	0.113	0.219	0.441	0.466	0.334	0.736	1							
YPNENET	0.308	0.480	0.335	0.443	0.461	0.619	0.369	0.446	0.251	1						
ERC	0.360	0.257	0.241	0.338	0.443	0.543	0.511	0.341	0.262	0.792	1					
PPAGDP	0.132	0.486	0.198	0.573	0.304	0.437	0.092	0.319	0.258	0.535	0.275	1				
RMARPG	0.328	0.664	0.377	0.450	0.271	0.459	0.185	0.204	0.177	0.573	0.444	0.457	1			
ID	0.495	0.842	0.421	0.173	0.105	0.418	0.045	0.200	0.179	0.522	0.351	0.339	0.732	1		
ISBP	0.499	0.847	0.399	0.096	0.076	0.389	0.040	0.188	0.179	0.484	0.346	0.274	0.661	0.978	1	
ODASGNI	-0.013	0.442	0.113	0.391	0.385	0.690	0.256	0.594	0.427	0.600	0.388	0.699	0.384	0.332	0.289	1

Source: Eurostat data, EViews 9 software authors' processing

In addition to the correlation matrix analysis, the stationarity of the series is also analyzed by applying a summary test of grouped unit roots to calculate and compare the statistical values with p-values. In order to reduce the number of ADF tests from 16 to 1, the summary test of grouped unit roots was used. Thus, using a common method of unit roots, the study observed that all statistical methods mentioned in Table 4 have significant statistics and probabilities. For example, ADF - Fisher Chi-square has a statistical value of 224.154 and a probability of 0.0000. Based on the above, we can conclude that the selected indicators can be presented as a stationary series.

Table 5. Pool unit root test for social SDGs selected indicators

Group unit root test: Summary Series: SPRPU, PRMPAST, ELETs, PATEA, PECE, APL, GEGTE, SHWNPG, PHWSM, YPNENET, ERC, PPAGDP, RMARPG, ID, ISBP, ODASGNI Sample: 1 378 Exogenous variables: Individual effects, individual linear trends Automatic selection of maximum lags, Automatic lag length selection based on SIC: 0 to 14 Newey-West automatic bandwidth selection and Bartlett kernel				
Method	Statistic	Prob.**	Cross-sections	Obs
Null: Unit root (assumes common unit root process)				
Levin, Lin & Chu t*	-7.6202	0.0000	16	6017
Breitung t-stat	-9.8905	0.0000	16	6001
Null: Unit root (assumes individual unit root process)				
Im, Pesaran and Shin W-stat	-12.6026	0.0000	16	6017
ADF - Fisher Chi-square	224.1540	0.0000	16	6017
PP - Fisher Chi-square	253.7340	0.0000	16	6032
** Probabilities for Fisher tests are computed using an asymptotic Chi -square distribution. All other tests assume asymptotic normality.				

Source: Authors' calculation based on EViews 9. Significant p-values recorded at the 5% significance level

The PCA table results in the first section, summarizing the eigenvalues or standardized variance associated with each factor, and the second section shows the eigenvector loadings. For example, Factor 1 explains 40.31% of the total variance, while Factor 2 accounts for 17.10%. The contribution of the first factor is obtained by dividing its eigenvalue (6.449763) by the total number of variables (16), resulting in a value of approximately 0.4031. Together, the first two principal components capture 57.42% of the overall variability in the data set. When the third component is included, the cumulative variance explained increases to 66.95%.

The patterns in the data explain the eigenvector loadings. For example, the eigenvector loading of the first principal component labeled PC1 is 0.1454 for SPRPU and 0.2914 for ID. The first component has positive eigenvector loadings for all variables. The principal component PC2 has six negative and ten positive values. The decision on how many factors to retain can be based on the eigenvalues. As we can see from Table 6, Factors 1,2,3,4 and 5 have eigenvalues greater than 1, indicating that they explain a significant part of the variation in the data, so only these factors are retained for analysis.

Table 6. Principal Components Analysis results for the selected social SDGs indicator

Sample: 1 378, Included observations: 378, Computed using: Ordinary correlations, Extracting 16 of 16 possible components																
Eigenvalues: (Sum = 16, Average = 1)																
Number	Value			Difference			Proportion			Cumulative Value			Cumulative Proportion			
1	6.449763			3.713017			0.4031			6.449763			0.4031			
2	2.736746			1.210770			0.1710			9.186509			0.5742			
3	1.525976			0.340513			0.0954			10.712480			0.6695			
4	1.185463			0.149651			0.0741			11.897950			0.7436			
5	1.035812			0.364821			0.0647			12.933760			0.8084			
6	0.670991			0.133896			0.0419			13.604750			0.8503			
7	0.537094			0.096155			0.0336			14.141850			0.8839			
8	0.440940			0.095236			0.0276			14.582790			0.9114			
9	0.345704			0.082101			0.0216			14.928490			0.9330			
10	0.263603			0.021719			0.0165			15.192090			0.9495			
11	0.241883			0.049053			0.0151			15.433980			0.9646			
12	0.192831			0.040830			0.0121			15.626810			0.9767			
13	0.152001			0.039513			0.0095			15.778810			0.9862			
14	0.112488			0.017104			0.0070			15.891290			0.9932			
15	0.095384			0.082063			0.0060			15.986680			0.9992			
16	0.013321			---			0.0008			16.000000			1.0000			
Eigenvectors (loadings):																
Variable	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	PC 9	PC 10	PC 11	PC 12	PC 13	PC 14	PC 15	PC 16
SPRPU	0.1454	-0.3625	0.3220	0.1306	-0.0781	-0.2278	0.4827	0.5428	-0.1594	-0.0253	0.1223	0.2218	0.1956	0.1045	-0.0402	0.0037
PRMPAST	0.2804	-0.2669	-0.3211	-0.0392	0.0211	0.0713	-0.1533	0.1593	0.2760	0.2380	-0.2123	-0.1593	0.1191	0.6686	0.1252	-0.1048
ELETS	0.1672	-0.2431	0.4331	-0.1054	0.4431	0.0273	0.1250	-0.4472	0.0846	0.0103	0.4030	-0.1421	-0.1510	0.1900	0.2295	-0.0151
PATEA	0.2209	0.0961	0.3203	-0.5013	0.0693	0.3584	-0.0868	0.1603	-0.2571	0.4251	-0.1815	-0.1599	0.2660	-0.1976	-0.0743	0.0326
PECE	0.1822	0.3213	-0.1347	0.1247	-0.4143	0.4728	0.2922	0.0705	0.1249	0.1321	0.5287	-0.0755	-0.0504	0.0404	0.1479	0.0030
APL	0.3032	0.1843	-0.0594	0.0960	0.0212	-0.3179	-0.2984	0.2155	-0.5546	0.0583	0.2159	-0.2673	-0.4224	0.1287	-0.0175	0.0007
GEGTE	0.1710	0.2286	0.4707	0.2295	0.0895	0.0881	-0.4240	0.3538	0.5126	-0.1714	0.0038	0.0531	-0.1054	-0.0300	-0.1281	-0.0091
SHWNPG	0.2322	0.3520	-0.1176	0.1893	0.3414	-0.0989	-0.0135	-0.0681	-0.0730	0.3074	-0.0401	0.6446	0.1383	-0.0134	0.3243	-0.0008
PHWSM	0.1911	0.3024	-0.1146	0.2351	0.4602	0.2124	0.4607	0.0159	-0.0722	-0.2594	-0.2992	-0.2738	0.0008	0.0802	-0.3013	-0.0128
YPNET	0.3300	0.0352	0.1193	0.0118	-0.3097	-0.1867	0.0620	-0.4244	0.0759	0.2202	-0.0351	0.2180	-0.0501	0.1477	-0.6573	0.0383
ERC	0.2704	0.0766	0.2981	0.2548	-0.4036	-0.1397	0.0881	-0.2469	-0.0486	-0.0903	-0.4524	-0.2393	0.0680	-0.0779	0.4788	-0.0610
PPAGDP	0.2572	0.0602	-0.1160	-0.5766	-0.0442	-0.1215	0.2646	0.1382	0.2475	-0.1802	-0.1800	0.1873	-0.5313	-0.1116	0.1487	0.0217
RMARPG	0.2989	-0.1874	-0.0432	-0.0807	-0.0976	0.4025	-0.2557	-0.1099	-0.3287	-0.6010	0.0141	0.3321	0.1215	0.1267	0.0036	0.0874
ID	0.2914	-0.3397	-0.1938	0.1628	0.0707	0.0890	-0.0545	-0.0103	0.0455	0.0725	0.0518	-0.0115	-0.0635	-0.4754	-0.0804	-0.6877
ISBP	0.2756	-0.3492	-0.2058	0.2348	0.0845	0.0484	-0.0408	0.0097	0.1249	0.1461	-0.0134	-0.1083	-0.0659	-0.3730	0.0135	0.7077
ODASGNI	0.2834	0.2025	-0.1861	-0.2545	0.0373	-0.4304	-0.0455	-0.0203	0.1833	-0.2703	0.2933	-0.2175	0.5747	-0.1403	-0.0032	0.0179

Source: Authors' calculation on EViews 9.

Combining Figure 1 and Table 6, we can see the number represented by eigenvalues. Thus, we should only retain five factors for which the eigenvalues are greater than 1. If we consider only two factors for simplification, the results are presented as in Figure 2.

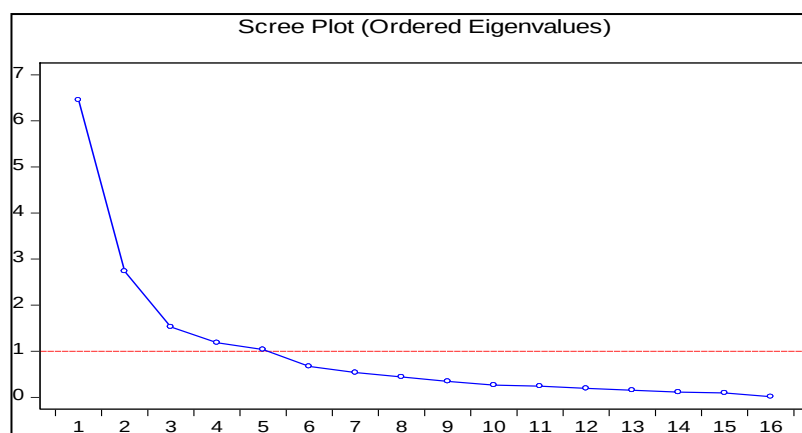


Figure 1. The observed eigenvalue matrix of the principal component analysis (PCA) for social SDGs indicators

Source: Authors' calculation based on EViews 9

As illustrated in Figure 2, the component scores are represented by circles, while the loadings of the macroeconomic indicators are represented as vectors. The first principal component accounts for the largest share of the total variance — 40.3% — and displays positive loadings for all variables analyzed. The second component explains 17.1% of the total variance, showing positive loadings for ten analyzed indicators and negative loadings for six indicators: SPRPU, PRMPAST, ELETS, RMARPG, ID, and ISBP.

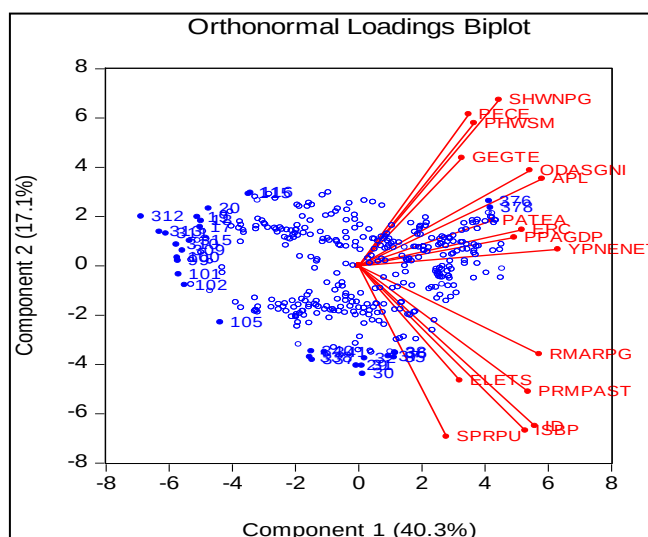


Figure 2. The orthonormal loadings of the principal component analysis (PCA) for social SDGs indicators

Source: Authors' calculation based on EViews 9

6. Conclusions

The article tested the correlation relationships between the normalized values of the main indicators of the Sustainable Development Goals with a social character. The dataset includes annual observations covering the period 2012-2025, with systematized panel data for the 27 EU countries, the data source being Eurostat, thus the total number of observations is 378.

To analyze the structure of the indicators of the Sustainable Development Goals (SDGs) with a social character, the principal component analysis (PCA) was performed using EViews 9. This allowed the examination of the eigenvalues and the eigenvector loadings derived from the correlation matrix. The correlation matrix showed that most of the indicators present positive linear correlations suggesting an amplification between them, and where there are negative correlations these are explainable from a logical perspective, in general the correlation matrix having a relatively equal weight between indicators with positive and negative meanings.

In terms of dimensionality reduction, the analysis found that five principal components had eigenvalues greater than 1. Specifically, Factor 1 recorded an eigenvalue of 6.449763, Factor 2 had 2.736746, and Factor 3 showed a value of 1.525976, Factor 4 indicated a value of 1.185463, and Factor 5 recorded a value of 1.035812. Based on this criterion, five components were retained for further interpretation. In terms of explained variance, Factor 1 accounted for 40.31% of the total variance, Factor 2 for 17.10%, Factor 3 for 9.54%, Factor 4 for 7.41%, and Factor 5 for 6.47%, with the cumulative contribution of these five factors totaling 80.84% of the overall variance.

Orthonormal loadings indicated that the first principal component, which explains the largest share of the variance (40.31%), is positively associated with all selected variables. The second component

explains only 17.10% of the variance and displays negative loadings for six of the sixteen indicators analyzed.

7. Future Work

Although the paper focuses mainly on the social Sustainable Development Goals indicators proposed by Eurostat, the list of indicators tracked may be longer or shorter depending on how they are selected. For example, indicators from other SDGs other than those chosen in this study may be included (specifically differently from SDGs 4, 5, 10 and 17).

At the same time, the geographical area and the period of analysis could be expanded or narrowed depending on other possible research directions. Other techniques and methodologies may also be addressed, for a better deciphering of the relationships between the SDGs.

References

- Booyesen, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research*, 59(2), 115–151. <https://doi.org/10.1023/A:1016275505152>
- Decancq, K., & Lugo, M. A. (2010). *Weights in multidimensional indices of well-being: An overview*. SSRN. <https://doi.org/10.2139/ssrn.1571124>
- Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography*, 38, 115–132. <https://doi.org/10.1353/dem.2001.0003>
- Greco, S., Ishizaka, A., Tasiou, M., & Torrisi, G. (2019). On the methodological framework of composite indices: A review of the issues of weighting, aggregation, and robustness. *Social Indicators Research*, 141(3), 1–34. <https://doi.org/10.1007/s11205-017-1832-9>
- Hijman, R., et al. (2025). *Sustainable development in the European Union: Monitoring report on progress towards the SDGs in an EU context* (Eurostat statistical books, 2025 edition). Publications Office of the European Union. <https://doi.org/10.2785/1111373>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., & Giovannini, E. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing.
- OECD. (2018). *Measuring distance to the SDG targets 2017: An assessment of where OECD countries stand*. OECD Publishing. <https://doi.org/10.1787/9789264308183-en>
- OECD. (2024). *How's life? 2024: Well-being and resilience in times of crisis*. OECD Publishing. <https://doi.org/10.1787/90ba854a-en>
- Sachs, J. D., Lafortune, G., Fuller, G., & Drumm, E. (2023). *Sustainable development report 2023: Implementing the SDG stimulus; includes the SDG index and dashboards*. Dublin University Press. <https://doi.org/10.25546/102924>
- Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., & Fuller, G. (2018). *SDG index and dashboards report 2018*. Bertelsmann Stiftung & Sustainable Development Solutions Network (SDSN).
- Saisana, M., & Tarantola, S. (2002). *State-of-the-art report on current methodologies and practices for composite indicator development*. European Commission, Joint Research Centre.
- Saltelli, A. (2007). Composite Indicators between Analysis and Advocacy. *Soc Indic Res* 81, 65–77. <https://doi.org/10.1007/s11205-006-0024-9>
- United Nations Development Programme. (2022). *Human development report 2021–22: Uncertain times, unsettled lives: Shaping our future in a transforming world*.
- Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic status indices: How to use principal components analysis. *Health Policy and Planning*, 21(6), 459–468. <https://doi.org/10.1093/heapol/czl029>