Composite Indicator Design Using AHP and PCA

Rajabali Ghasempour
Mahdi Mohammadzadeh

December 13, 2024

Abstract

The socio-economic performance of nations is a multi-dimensional phenomenon that requires robust and systematic approaches to evaluate. This study introduces a comprehensive framework for assessing and ranking the socio-economic performance of European Union (EU) countries. The framework combines the Analytic Hierarchy Process (AHP) and Principal Component Analysis (PCA) to create a composite indicator that simplifies complex datasets into an actionable ranking metric.AHP is utilized to incorporate qualitative insights and structure decision-making hierarchically, while PCA provides objective weights based on data variability, minimizing subjective biases.
Moreover, a correlation analysis is performed to identify relationships between individual indicators and the composite index, revealing both direct and indirect influences. This approach allows for a deeper understanding of how key factors such as income, education, and poverty, along with indirect contributors like health and employment, shape socio-economic outcomes. The study emphasizes the interconnected nature of socio-economic dimensions and demonstrates the utility of integrating data-driven and expert-based methodologies. The results provide policymakers with a reliable tool to prioritize interventions, track progress, and design informed, impactful policies.

1. Introduction

Objective: The purpose of this study is to develop a robust composite indicator for evaluating and ranking the socio-economic performance of European Union (EU) countries. This composite indicator combines both qualitative and quantitative approaches to ensure a systematic, objective, and insightful evaluation.

Key Dimensions:

- **Income Levels (Goal10_Income):** Representing economic prosperity.
- **Poverty Rates (Goal10_Poverty):** Measuring economic disparity (normalized inversely for clarity).
- **Education Proficiency (Goal4_1_1):** Reflecting human capital development.

Methodology Highlights:

- The Analytic Hierarchy Process (AHP) is employed to create a hierarchical structure and systematically combine indicators.

- Principal Component Analysis (PCA) is used to derive objective weights based on data variability.
- Correlation Analysis was conducted to assess the relationships between individual indicators and the composite index, uncovering both direct and indirect influences on socio-economic performance.

—

Dataset Overview

Dataset Description: The dataset includes socio-economic indicators for 27 EU countries, covering multiple dimensions related to income, poverty, education, health, and employment.These indicators align with the Sustainable Development Goals (SDGs) and provide a comprehensive view of socio-economic performance across countries.

Key Indicators:

- **Income (GOAL10_Income):** Income-related statistics for each country.
- **Poverty (GOAL10_Poverty):** Poverty rates or economic conditions.
- **Health (GOAL3.c.1):** Health worker density and distribution.
- **Education (GOAL4.1.1):** Proportion of children achieving proficiency in reading and mathematics.
- **Employment (GOAL8.5.2):** Unemployment rate by sex, age, and disability status.

Sample Data: Table 1 presents a subset of the dataset showcasing key metrics for five EU countries.

Table 1: Sample of Socio-Economic Indicators from the Dataset

Country	**Income**	**Poverty**	**Health Workers**	**Education Proficiency**	**Unemployment**
Austria	0.35	10.5	0.61	98.4	12.33
Belgium	2.11	8.6	0.88	98.6	13.35
Bulgaria	5.77	15.1	0.68	92.3	16.79
Croatia	5.82	11.3	0.67	99.9	24.09
Czechia	3.97	4.7	0.74	99.6	11.15

Insights:

- Income levels vary significantly across countries, with Bulgaria and Croatia reporting higher values than Austria and Belgium.
- Poverty rates show disparity, with Czechia having the lowest poverty rate at 4.7
- Health worker density highlights regional differences in healthcare infrastructure.
- Education proficiency rates are consistently high, reflecting strong educational outcomes in most countries.
- Unemployment rates reveal considerable variation, emphasizing the need for tailored policy interventions.

2. Problem Statement

Why Use Composite Indicators?

- Socio-economic data is multi-dimensional and complex, making it difficult to inter-pret directly.
- Composite indicators simplify this complexity into a single score, enabling:
  - Rankings: Comparing countries systematically.
  - Tracking Progress: Monitoring performance over time.
  - Policy Insights: Identifying areas for improvement.

Challenges:

- Indicators have diverse units and scales.
- Assigning weights often introduces subjective bias.
- Balancing trade-offs across multiple dimensions is non-trivial.

What This Study Does:

1. Combines AHP and PCA to create a composite indicator.
2. Compares two approaches for weight assignment: manual (expert-based) and PCA-based.
3. Applies the methodology to socio-economic data and analyzes the results.

3. Methodology

This section outlines the steps undertaken to design a composite indicator that evaluates and ranks the socio-economic performance of EU countries. The methodology integrates both the Analytic Hierarchy Process (AHP) and Principal Component Analysis (PCA) to derive weights and construct the composite index.

Composite Indicator Design:

1. Framework Definition:
   - **Goal:** Develop a composite indicator to rank EU countries based on socio-economic performance.
   - **Criteria:** Income, Poverty, and Education, which represent key dimensions of socio-economic well-being.
   - **Alternatives:** The 27 EU countries included in the dataset.
2. Weight Assignment: Two approaches are used to derive weights for the composite indicator:

- **AHP-Based Weights:** Expert judgments are used to construct pairwise comparison matrices, ensuring a systematic framework for assigning weights.
- **PCA-Based Weights:** Data-driven weights are derived based on the variability of the dataset, minimizing subjective bias.

3. Data Normalization: To ensure comparability across indicators with different units, all data are normalized to a [0, 1] scale using the following formula:

$$x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

4. Composite Indicator Calculation: The normalized data are combined using the derived weights:

$$CI_{\text{country}} = \sum_{i=1}^{n} w_i \cdot x_{\text{norm},i}$$

where w_i represents the weight for criterion i, and $x_{\text{norm},i}$ is the normalized value for criterion i for each country.

—

4. Analytic Hierarchy Process (AHP)

Overview of AHP

AHP is a decision-making framework developed by **Thomas Saaty**, designed to evaluate and compare multiple criteria by combining qualitative and quantitative inputs. It is particularly useful when the criteria are complex or challenging to quantify directly.

Steps in AHP

1. Define the Hierarchy:
   - Establish a clear goal (e.g., rank EU countries).
   - Identify the criteria influencing the decision (e.g., Income, Poverty, Education).
   - List the alternatives to be evaluated (27 EU countries).

2. Pairwise Comparison Matrix: A pairwise comparison matrix is constructed using Saaty's scale (1–9), where:
   - 1: Equal importance.
   - 3: Moderate importance.
   - 5: Strong importance.
   - 7: Very strong importance.
   - 9: Extreme importance.

Example matrix for Income, Poverty, and Education:

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1/3 & 1 & 2 \\ 1/5 & 1/2 & 1 \end{bmatrix}$$

3. Normalize the Matrix: Each element of the matrix is divided by the sum of its respective column:

$$A_{\text{norm},ij} = \frac{A_{ij}}{\sum_{i=1}^{n} A_{ij}}$$

4. Calculate Weights: Average each row of the normalized matrix to derive the weights for each criterion:

$$w_i = \frac{\text{Sum of Normalized Row Elements}}{\text{Number of Criteria}}$$

5. Consistency Check: To ensure consistency in the judgments:

$$CI = \frac{\lambda_{\max} - n}{n-1}, \quad CR = \frac{CI}{RI}$$

where $\lambda_{\max}$ is the principal eigenvalue of the matrix, and RI is a random index based on the matrix size. If $CR < 0.1$, the judgments are considered consistent.

—

5. Principal Component Analysis (PCA)

Overview of PCA

PCA is a statistical method used to reduce dimensionality by identifying components that capture the most variance in the data. It objectively derives weights based on the dataset's structure, ensuring a data-driven approach to composite indicator construction.

Steps in PCA

1. Normalize the Data: Standardize the data to eliminate the influence of different units.

2. Compute the Covariance Matrix: The covariance matrix is calculated to exam-ine relationships between indicators:

$$C = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^\top$$

3. Calculate Eigenvalues and Eigenvectors: Solve the covariance matrix to obtain eigenvalues (λ_i) and eigenvectors.

4. Derive Weights: Assign weights based on the proportion of variance explained by each component:

$$w_i = \frac{\text{Variance Explained by Component } i}{\text{Total Variance}}$$

Comparison of AHP and PCA Weights

The study evaluates both manual AHP-derived weights and PCA-derived weights to construct the composite index. The consistency of AHP weights is validated, while PCA provides an objective, data-driven perspective.

Strengths of the Approach:

- Combines subjective (AHP) and objective (PCA) weighting methods.
- Ensures consistency through AHP validation.
- Captures intrinsic data variability with PCA.
- Incorporates indirect influences identified through correlation analysis.

5. Dataset Description

Overview:

- Data covers socio-economic indicators for 10 EU countries.
- Indicators include Income, Poverty, and Education.

Preprocessing:

- No missing data (**Missing = 0**).
- All indicators normalized to [0, 1].

6. Application and Results

Weights Derived:

- **Manual AHP Weights:** Income = 0.5, Poverty = 0.3, Education = 0.2.
- **PCA-Based Weights:** Income = 0.525, Poverty = 0.336, Education = 0.139.

Composite Scores:

- Using manual weights: Lithuania = 0.79, Estonia = 0.80.
- Using PCA weights: Lithuania = 0.84, Latvia = 0.76.

Rankings:

- Top performers: Lithuania, Estonia.
- Bottom performers: France, Norway.

7. Correlation Analysis and Insights

Correlation Analysis with Composite Index:
To assess the influence of individual indicators on the composite index derived from AHP, a correlation analysis was conducted. The results revealed interesting patterns:

Table 2: Correlations with AHP Composite Index

Indicator	Correlation Coefficient
AHP Composite Index	1.000000
Goal10_Income	0.895693
Goal3_4_1	0.709528
Goal3_3_2	0.700425
Goal3_9_3	0.688340
Goal3_6_1	0.627596
Goal3_9_1	0.610411
Goal3_a_1	0.547281
Goal3_d_2	0.520973
Goal8_1_1	0.497947
AHP-PCA Composite Index	0.492003
Goal10_Poverty	0.483442
Goal3_1_1	0.452615

Key Insights:

- **Strong Correlations:** Indicators such as *Goal10_Income* and *Goal3_4_1* showed strong correlations with the AHP Composite Index, emphasizing their direct impact.
- **Unexpected Influences:** Some indicators like *Goal3_a_1* and *Goal8_1_1*, although not part of the main composite index, demonstrated significant correlations. This suggests indirect effects on socio-economic outcomes.
- **Policy Implications:** The results highlight the importance of considering indirect influences while designing policies. For instance, health-related indicators may play a more substantial role than initially assumed.

Conclusion from Analysis: While the selected indicators (Income, Poverty, and Education) were crucial, the analysis reveals that broader dimensions, particularly health and employment, also significantly contribute to socio-economic performance. Future iterations of composite index design could integrate these findings to enhance robustness.

8. Analysis and Insights

Key Observations:

- PCA weights minimize subjectivity and reflect intrinsic data variability.
- Manual weights capture expert judgment but are prone to bias.

- Rankings vary slightly based on weighting methods, emphasizing the importance of weight selection.

Implications:

- Policymakers can focus on improving education and poverty metrics in low-performing countries.
- PCA-based AHP provides a more objective approach for cross-country comparisons.

9. Conclusion

The study successfully developed a composite indicator using AHP and PCA methods to evaluate the socio-economic performance of EU countries. By integrating qualitative and quantitative approaches, the composite index offers a robust framework for ranking and identifying key areas for policy improvement.

Key Findings:

- **Direct Impacts:** Indicators like Income and Education, integral to the composite index, showed strong correlations, validating their importance in socio-economic performance.
- **Indirect Influences:** Indicators such as *Goal3_a_1* (Health) and *Goal8_1_1* (Economic Growth) demonstrated significant indirect contributions, highlighting the multi-dimensional nature of socio-economic development.
- **Methodological Insights:** PCA minimized subjectivity in weight assignment, while AHP provided a systematic structure for integrating diverse criteria.

Future Directions:

- Future work should incorporate a wider range of indicators, particularly health and employment metrics, as their indirect effects suggest a broader influence on socio-economic outcomes.
- Hybrid weighting methods combining expert judgment with data-driven approaches could further enhance the robustness and applicability of the composite index.