Construction & Evaluation of Composite Indices

Group E: Ken Ye, Ejay Lin, Gorden Gao

Composite Indices: Overview

- Definition: mathematical combination of individual indicators that represent different dimensions of a concept.
- Useful for policy making & public communications in conveying information
  - Environment
  - Economy
  - Society
  - Technology

(Nardo et al., 7)

Composite Indices: Pros & Cons

Pros

- Big Picture Perspective
  - Easier to interpret than analyzing numerous separate indicators.
- Efficient Information Handling
  - Assist in reducing the size of indicator lists or incorporating more information within existing size limits.

(Saisana et al., 1)

Cons

- Risk of Misleading Messages
  - May convey misleading policy messages if poorly constructed/misinterpreted.
- Increased Data Requirements
  - Information is required for all sub-indicators.
- Subjectivity in Construction
  - Construction involves judgment in selecting sub-indicators, choosing models, weighting indicators, and handling missing values.

(Saisana et al., 2)

Index Construction
Step 1: Understanding Indicator Structure

- Multivariate statistics
  - Principal Component Analysis (PCA)
- Assess dataset suitability
- Assess implications of methodological choices
  - e.g. weighting and aggregation

(Nardo et al., 9-10)

Index Construction
Step 2: Handling Missing Data

- Case deletion
  - Remove the subject/indicator from analysis.
  - May produce biased estimates if removed records are not a random subsample.
- Single imputation
  - Mean/Median/Mode substitution, Regression Imputation, Expectation-Maximisation Imputation.
- Multiple imputation
  - Markov Chain Monte Carlo algorithm

(Nardo et al., 10-11)

Index Construction
Step 3: Transformation of Raw Data

- Truncation
    - Trim the tails of the sub-indicators’ distributions.
    - Avoid dominance of extreme values.
- Functional Transformation
    - Linear Functional Form
        - Changes in indicator values equally important across all levels.
    - Concave Down Functional Form
        - Logarithmic or nth root transformations.
        - Changes more significant at lower levels of the indicator.
    - Concave Up Functional Form
        - Exponential or power transformations.
        - Changes more important at higher levels of the indicator.

(Nardo et al., 11)

Index Construction
Step 4: Normalization of Indicators

- Indicators have different measurement units.
    - Need to bring indicators to the same unit.
- Methods
    - Standardization/Z-score: Adjusts values to have a mean of 0 and a standard deviation of 1.
    - Re-scaling: Adjusts the range of values to a predefined scale (e.g. 0 to 1).
    - Ranking: Assigns a rank to each value, creating a relative ordering.

(Nardo et al., 11)

Index Construction
Step 5: Weighting

- Different weights to reflect relative significance, data quality, cyclical conformity, etc.
- Techniques
  - Statistical models
    - e.g. correlation coefficients/PCA to avoid double counting
  - Higher weight for statistically reliable data
    - Few missing values
    - Large coverage
    - Sound values
  - Expert opinions

(Nardo et al., 11-12)

Index Construction
Step 6: Aggregation

- Linear Aggregation
  - Addition
  - Same measurement unit
  - Full compensability
    - Poor performance in one indicator can be offset by high values in others.
- Geometric Aggregation
  - Multiplication
  - Positive values & different ratio-scales
  - Partial compensability
    - Lower when the composite index contains indicators with low values.

(Nardo et al., 12)

Index Evaluation: Variable Importance - Part 1

- **Variable importance**
    - Part of Composite Indices
        - Weighting of variables from which the index is composed of in order to aggregate the variables into an index
            - Empirical importance
    - Identify which individual variables have a greater impact on the overall index score
    - Crucial for policymakers and researchers
        - Provides insights into the key drivers of the index
        - Can inform targeted interventions or policy changes
- Key points related to variable importance in index evaluation
    - Weighting Variables
    - Sensitivity Analysis
    - Interpretability
    - Policy Implications

Index Evaluation: Variable Importance - Part 2

- Nominal weights vs. **Importance** (Schlossarek, Syrovátka, Vencálek)
  - Nominal weight assigned to a variable often differs from the degree to which the variable affects the scores of the overall index
  - Do not provide credible information on the importance of variables if it is not accompanied by other methodological procedures, e.g. the method of data normalization
    - Creates a perception that the nominal weight is a measure of the importance of the variable
  - The stronger the correlation is between the original and modified composite index, the lower the importance of the examined indicator
    - Correlation coefficient equal to 1 means that the exclusion of the indicator does not affect (up to a linear transformation) the composite index and therefore has no importance

Index Evaluation: Variable Importance - Part 3

- Nominal weights vs. **Importance** (Schlossarek, Syrovátka, Vencálek)
  - Concept of importance of a given indicator is based on correlation between
    - The original composite index (OCI) which includes the examined indicator
    - The modified composite index (MCI) which excludes the examined indicator

$$OCI_j = \sum_{i=1}^{p} w_i x_{ij}, \quad \text{for } j = 1, \ldots, N,$$

$$MCI_k = \sum_{i \neq k} w_i x_i = OCI - w_k x_k, \quad \text{for } k = 1, \ldots, p.$$

$$r_k = corr(OCI, MCI_k), \quad \text{for } k = 1, \ldots, p.$$

$$MAI_k = (1 - r_k) * 50, \quad \text{for } k = 1, \ldots, p.$$

$$\boxed{MRI_k = \frac{MAI_k}{\sum_{i=1}^{p} MAI_i} \times 100, \text{ for } k = 1, \ldots, p.}$$

- Correlation coefficient between the nominal weight and the importance of indicators in paper's sample is **0.52**
  - A moderately unsurprising strong relationship

Quality Assessment: Sensitivity Analysis - Part 1

$$Y_c = \sum_{q=1}^{Q} I_{q,c} w_q, \qquad \text{where} \begin{cases} I_{q,c} = \dfrac{x_{q,c} - \min(x_q)}{\text{range}(x_q)}, \\ I_{q,c} = \dfrac{x_{q,c} - \text{mean}(x_q)}{\text{std}(x_q)}. \end{cases}$$

- ***Y*** is the composite indicator of country ***c***, ***I*** is the normalized value of sub-indicator ***q*** for country ***c***, w is the weight of sub-indicator ***q*.**
- ***Y*** in this study specifically entails TAI (Technology Achievement Index), and the study finds the ranking of TAI varies across different CI construction methodologies
- Two normalization methods are discussed:
  - **Rescaling** (Min - Max): transforms into a common scale ranging typically from 0 to 1
  - **Standardization** (z-score): adjusts the data to have mean of 0 and deviation of 1
- Two weighting methods are discussed:
  - **Budget Allocation Process (BAP)**:
    - Allocate a fixed budget, conceptualized as 100 points, across sub-indicators
  - **Analytic Hierarchy Process (AHP)**:
    - Compare two sub-indicators at a time using a numerical scale (1-9)
    - Construct a square matrix and relative weights derived with eigenvector

Quality Assessment: Sensitivity Analysis - Part 2

- **Uncertainty Analysis (UA)**: UA examines how uncertainty in inputs (like normalization methods, weighting schemes, and weights) affects the CI values
    - Monte Carlo method: involving generating many sets of weights and applying different normalization methods
    - The output of these simulations is a set of values for the composite indicator. This set is analyzed statistically to understand the distribution of the composite indicator, characterized by metrics such as mean, variance, and confidence intervals.

$$Y_i = f(X_{1i}, X_{2i}, \ldots, X_{ni})$$

where f represents the model (e.g., the composite indicator calculation), and X_{ji} is the value of input j in simulation run i

Quality Assessment: Sensitivity Analysis - Part 3

- **Sensitivity Analysis (SA)**: SA focuses on the contribution of each uncertainty source to the output variance
  - Variance-based techniques (i.e. Sobol Index, Total Effect Index) are employed.
  - By computing these indices for each input, we can quantify the importance of different factors (such as weights and normalization methods) in determining the composite indicator

$$S_i = \frac{\mathrm{Var}[\mathbb{E}(Y|X_i)]}{\mathrm{Var}(Y)}$$

Sobol First Order Index: decompose the variance of the output into fractions attributable to each input or sets of inputs

$$S_{Ti} = 1 - \frac{\mathrm{Var}[\mathbb{E}(Y|X_{\sim i})]}{\mathrm{Var}(Y)}$$

Total Effect Index: measure effect of an input variable including its interactions with other input variables

Quality Assessment: Sensitivity Analysis - Part 4

- **Concluding Thoughts**: Integrating UA and SA allows for an assessment of both the overall uncertainty in the model output and the contribution of individual inputs to this uncertainty
    - UA helps to understand the range of possible outcomes and the overall uncertainty in the model
    - SA identifies which inputs are most responsible for this uncertainty
- **Implications**: For a composite indicator, this means understanding how different weights, normalization methods, and other model inputs contribute to the uncertainty in the final indicator values, providing insights into which aspects of the model are most critical and thus should be the focus of attention and refinement

Reference

Booysen, Frederik. “An Overview and Evaluation of Composite Indices of Development.” Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, Springer, 1 Jan. 1970, ideas.repec.org/a/spr/soinre/v59y2002i2p115-151.html.

Nardo, Michela, et al. Tools for Composite Indicators Building - Europa.Eu, publications.jrc.ec.europa.eu/repository/bitstream/JRC31473/EUR%2021682%20EN.pdf. Accessed 27 Nov. 2023.

Saisana, M., et al. “Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators.” Journal of the Royal Statistical Society Series A: Statistics in Society, vol. 168, no. 2, 2005, pp. 307–323, https://doi.org/10.1111/j.1467-985x.2005.00350.x.

Schlossarek, Martin, et al. “The importance of variables in composite indices: A contribution to the methodology and application to development indices.” Social Indicators Research, vol. 145, no. 3, 2019, pp. 1125–1160, https://doi.org/10.1007/s11205-019-02125-9.