Recebido: 06-06-2023 | Aprovado: 20-10-2023 | DOI: https://doi.org/10.23882/rmd.24160

Geometric mean approach to measure Composite Index

Abordagem média geométrica para medir o Índice Composto

Satyendra Chakrabartty,
Indian Statistical Institute, Indian Maritime University, Indian Ports Association
(chakrabarttysatyendra3139@gmail.com)

Abstract: Composite Index (CI) maps selected *n*-indicators to real line. Existing measures of CIs suffer from limitations. In addition to selection of component indicators, other areas of CI receiving heavy criticism are normalization, weighting and aggregation. The paper proposes CI as geometric mean (GM), avoiding scaling, weighting and outliers with no bias for either developed or under-developed countries. Proposed CI can cover indicators in ordinal or ratio/interval scale or those in percentages and depicts overall improvement/decline in the current year from the base year or from the previous year by a continuous, monotonically increasing function. It helps in better comparisons of countries across time and space, identification of key areas, plotting growth of CI of a country using well defined chain-indices The index with wide applicability can be used for better ranking, estimation of population parameters and testing statistical hypothesis. Separate indices for dimensions can be constructed by considering indicators relevant to a domain. Future studies suggested.

Keywords: Composite index; Monotonically increasing; Geometric mean; Time-reversal test; Chain indices.

Resumo: O Índice Composto (IC) mapeia n-indicadores selecionados para uma linha real. As medidas de IC existentes sofrem de limitações. Para além da seleção dos indicadores componentes, outras áreas do IC que têm sido alvo de fortes críticas são a normalização, a ponderação e a agregação. O presente documento propõe que o IC seja a média geométrica (MG), evitando o escalonamento, a ponderação e os valores anómalos, sem enviesamento, tanto para os países desenvolvidos como para os em desenvolvimento. O IC proposto pode abranger indicadores em escala ordinal ou de rácio/intervalo ou em percentagens e representa a melhoria/declínio global no ano em curso em relação ao ano de referência ou ao ano anterior através de uma função contínua e crescente. Ajuda a melhorar as comparações de países no tempo e no espaço, a identificar áreas-chave, a traçar o crescimento do IC de um país utilizando índices em cadeia bem definidos. O índice, com uma ampla aplicabilidade, pode ser utilizado para uma melhor classificação, estimar parâmetros populacionais e testar hipóteses estatísticas. Podem ser construídos índices separados para as dimensões, considerando indicadores relevantes para um domínio. Sugerem-se estudos futuros.

Palavras-chave: Índice composto; Aumento monotónico; Média geométrica; Teste de inversão temporal; Índices em cadeia.

JEL classification: C43, O15

1. Introduction

Composite Index (CI) combines measures of several indicators on economic, social, environmental variables, government initiatives, wellbeing of health, psychological, satisfaction of basic needs, social/cultural, etc. Large numbers of CIs have been developed for measuring complex overall performance for comparing entities and orienting policy-decisions. CIs are being used for policy makers, academics, etc. in various areas like Economics, Social sciences (OECD, 2008), Economic Freedom Index (Johnston and Sheehy, 1995), Poverty Index (OPHI, 2018), Well-being and quality-of-life (Cendrero et al. 2003), and in other areas like Logistics Performance Index (LPI) (Arvis et al., 2016), Environmental Sustainability Index (ESI) (World Economic Forum, 2002), Life cycle index (Khan et al., 2004), Corruption Perceptions Index (Lambsdorff, 2006), etc. European Parliamentary Research Service (EPRS) (2021) selected following 10 CIs to support policy makers:

- Climate Change Performance Index (CCPI) (www.germanwatch.org/en/ccpi_bame): Assesses countries' efforts to combat climate change and reduction of greenhouse gas (GHG) and to evaluate a country's compatibility with the Paris Agreement goals. Each of 14 indicators of CCPI, distributed over four dimensions is converted to a 0-100 scale. However weights of individual indicator varies from 10% (for on GHG emissions, Climate policy) to 5% for each indicator under Renewable energy; Energy used. CCPI uses overall EU target and not the targets of the Member States, different baseline year for different countries and relative GHG emission reduction targets which gives rise to positive or negative skews.
- Commitment to Development Index (CDI): Combines six components to assess contribution of 40 countries to global development, with emphasis on improved 'footprint'. Inclusion of OECD countries (with sound reporting systems) and non-OECD countries in CDI has generated data gaps, mitigation of which is problematic (Birdsall and Roodman, 2003).
- Ease of Doing Business Index (EoDBI): Has 41 indicators grouped in 10 equally weighted topics. EoDBI measure regulations directly affecting businesses. EoDBI has been criticized for not considering social and human rights situations in

the countries. The World Bank conducted an internal audit in 2020 for rectification of flawed data and suspended the 2021 edition (Djankov, 2016)

- Ecological Footprint (EF): Finds how much biologically productive land and water an individual, population or activity requires producing all it consumes including absorption of the generated wastes. Five sub-components common to EF and bio-capacity are (cropland, grazing land, built-up land, fishing grounds, forest areas). In addition, EF also considers carbon footprint. Ecological deficit occurs when the footprint of a population exceeds the bio-capacity of the area available to that population. Major criticisms of the EF are over-focus on forests and covering only part measures of sustainability and not exhausting all dimensions of environmental quality (Wackernagel, 1996).

- EU Regional Competitiveness Index (RCI): With 74 indicators, RCI measures differences in regional competitiveness across Europe with focus on major factors of competitiveness. RCI, with a broad set of indicators tracks ability of a region to provide an attractive environment for businesses and residents to live and work. It is based on Global Competitiveness Index (GCI), which monitors competitiveness at national level (Annoni & Dijkstra, 2017).

- EU Social Progress Index (EU-SPI): Considers 55 indicators on health, education, safety, environmental quality and personal rights which are grouped into three dimensions. The selected dimensions may not be sufficient to achieve social progress as a whole. It uses hybrid aggregation method, unweighted arithmetic mean within each component and generalized unweighted mean across components and dimensions (Beltrán-Esteve et al. 2023).

- Human Development Index (HDI) (UNDP, 2010): HDI aggregates three chosen dimensions at country level by geometric mean without showing inequalities in a particular country.

- Media Pluralism Monitor Index (MPM): MPM attempts to assess the potential weaknesses in national media systems that may not be conducive for media pluralism. Through a questionnaire consisting of 200 questions on 20 indicators, it covers four dimensions. There is no overall score (Brogi, 2020).

- Normandy Index (NI): Attempts to measures level of threats to peace, security and democracy based on interconnectedness of threats and linkages between such threats. Despite certain weakness, NI demonstrates some positive evolution, in particular the sustainable convergence of some countries in the

European neighborhood. NI complements the annual Peace and Security Outlook produced by the European Parliamentary Research Service (EPRS).

- Sustainable Development Report's SDG Index: Provides an assessment of progress made towards the SDG goals by all UN Member States. The Report includes scores on a scale of 0 to 100 and can be interpreted as a percentage towards optimal SDG performance. Therefore, the difference between 100 and a country's SDG Index score is the distance, in percentage points, that must be overcome to reach optimum SDG performance. Major limitations are significant missing data, equal weights, different number of indicators for different SDG-goals and normalization by Min-Max transformation (Lafortune et al. 2022)

Different CIs with different purpose differ in number of items, item formats; scoring methods, dimensions covered, etc. and thus, CI-scores are not comparable. Lower rank of a country with respect to a CI-value has often been debated since CIs are used as documentary evidence of success of policies adopted by national governments. Properties of CI, values and country-ranks are affected by the changes in methodology (EPI Report, 2016). Two areas receiving heavy criticism are weighting and aggregation (Greco et al. 2019)

Stages of construction of CI are (i) Selection of indicators and domains. (ii) Scaling or normalizing raw data and (iii) Aggregation of transformed scores to get a single value of CI. Lack of agreements at each stage of construction of CI makes such indices useless if not misleading (Böhringer and Jochem, 2007). For a set of chosen variables or indicators $X_1, X_2, \ldots, X_n$, CI finds a function *f* where the vector $(X_1, X_2, \ldots \ldots, X_n)^T$ is mapped on the real line. The value of the function indicates the overall position of a country/region based on which the countries may be compared across time and space. Evaluation of such a function f: $R^n \rightarrow R$ i.e. from n-dimensional space to real line involves a sound methodological approach which is critical for better comparisons and ranking in terms of combined measure of social, economic, political, environmental areas (Freudenberg, 2003). The selected indicators are measurable and reflect levels of achievement of components or dimensions/drivers of the Index in question expressed by objective and also subjective measures. In general, indicators are in different units, discrete or continuous, interval/ratio scale or categorical, independent or correlated with different degrees. The chosen indicators ($X_i's$) could be positively related to CI, where higher $X_i \Longrightarrow$ higher CI or negatively related where lower the $X_i \Longrightarrow$ higher the CI. Aggregating such positive and negative variables may require prior adjustments. Variations in stages of construction of CI may impact significantly values and properties of the resultant CI, affecting quality and appropriateness of such indices even for the same purpose. Jacob, et al. (2004) observed that methodological issues associated with construction of CI give rise to potential manipulations and misinterpretations and suggested to address issues like reliability, validity, appropriateness of such CIs.

The paper describes major methodological issues regarding construction of CI and suggests an assumption free method of CI avoiding scaling, weighting and reducing level of substitutability among the component indicators. Properties satisfied by the proposed CI and benefits are deliberated.

2. Stages of construction of CI:

2.1 Selection of dimensions, indicators:

The selected indicators within a dimension should be relevant to the purpose for which CI is being developed. "Hedonic" measures focuses on the "feeling" components (e.g. happiness) in contrast to "eudaimonic" measures which focuses on the "thinking" components (Ryff and Keyes, 1995). A narrow selection of indicators may fail to represent the overall picture as conceptualized by the CI in question. Broad selection of variables is likely to increase the problems of multicolinearity and complexity in transforming data and aggregation. Greco, et al. (2019) opined to enlarge the range of indicators for constructing CI

Examples of CI with small number of variables include Measures of America containing three domains and four indicators. HDI uses six items. LPI uses six number of 5-point Likert items to assess six dimensions of trade. Overall Life Satisfaction (International Wellbeing Group, 2013) with a single item gives poor discrimination and makes internal consistency reliability inappropriate. Examples of CIs with large number of indicators include: Worldwide Governance Indicators project considered over 300 indicators and 33 data sources (Kaufmann, et al. 2007).

Quality of Life (QOL) with 100 items covers six main domains (WHOQOL-100). Health-related quality of life (HRQOL) includes number of areas associated with well-being. Patient-Reported Outcomes Measurement Information System (PROMIS) considers large number of self-reported items covering multiple HRQOL domains (Cella et al. 2019). The 10-item global HRQOL assesses selected domains of physical/mental/social health including pain, fatigue and perceptions of QOL. Social Progress Index (SPI) combines 51 social and outcome indicators to calculate overall score for countries (www.socialprogressindex.com/methodology).

CIs have been criticized for not considering all relevant dimensions or indicators. OECD Better Life Initiative (BLI) (www.oecd.org/ betterlifeinitiative) covers 11 domains, but ignores usual air–water–noise pollutions and may not fully capture what is truly important for the CI. Kasparian and Rolland (2012) suggested inclusions of more number of variables in BLI. Gross National Happiness (GNH) with nine equally important domains and 33 indicators was later modified to include 72 indicators (www.grossnationalhappiness.com). Canadian Index of Wellbeing (CIW) with 8 indicators for each of the 8 domains ignores subjective Well-being (Michalos et al. 2011). Happy Planet Index (HPI) (www.happyplanetindex.org) ignores QOL, income equality, literacy rate, abuse and violation of human rights, environmental issues, etc. EU's Quality of Life Index (Distaso, 2007), National Accounts of Wellbeing (NEF, 2008) ignore environmental issues, time use and extent of meeting human needs like shelter, safety, security etc.

High correlations between component variables imply multicolinearity, which is inevitable in CI (Smith, 2002). High correlation of one variable with CI implies no need of multidimensional CI and instead the former could well be used (McGillivray and White, 1993).

2.2 Nature of data

Depending on the purpose, CI may consider population-based data and also primary data obtained from survey using (say) self- reported Likert scales or rating scales. Orff, et al. (2007) observed that self-reported insomnia differ with objective evidences of sleep disturbance. Data generated by Likert/rating scales are discrete and ordinal where items may have equal or different number of levels or combination of k-point scales (K= 3, 4, 5, 7) plus Yes –No type binary items like SF-36 (Ware & Sherbourne, 1992). SF-36, CIW, Wellbeing index by ElSarawy, (2016) transforms raw data to percentages before taking average. However, average of percentages is wrong, when the denominator$d_i \neq k * d_j$. Pooled average of 80% (8 out of 10) and 40% (4.8 out of 12) is 58.18% which is different from average of 80% and 40%. Similarly, actual value by a country (in %) – minimum value in a group of countries (in %) could be problematic. To combine percentage figures, Human Poverty Index (HPI) (UNDP, 2007) first finds average of figures (in percentage) and then takes 3-rd root and 4-th root of such average, to get respectively HPI-1 and HPI-2. Happiness surveys often results in negative and positive happiness which are subjectively determined by the respondents. SPI contains around 50 indicators of which one is ordinal (from 0 to 6). HPI considers Life expectancy (secondary data), and experienced wellbeing (Survey based with one item from 0 to 10). Personal Wellbeing Index (PWI) has seven Likert items, each with 11 levels, plus another item on 'spirituality or religion' (International Wellbeing Group, 2006). Combining secondary data and Likert scores could be problematic.

2.2.1 Problems with Polytomus items:

- Ordinal scores from Likert/Rating scales are not additive as they are not equidistant (Hobart et al. 2007)
- Equal importance to the items is not justified since items show different values of correlations with total scores, correlations between a pair of items and factor loadings.
- Arithmetic mean (AM) requiring equidistant scores are not meaningful for ordinal item scores (Jamieson, 2004) and $\bar{X}$ > or <$\bar{Y}$ is meaningless (Hand, 1996). If addition is not meaningful, standard deviation (SD), correlation, Cronbach α, etc. are not meaningful and analysis like regression, Principal component analysis (PCA), Factor analysis (FA), etc. with ordinal item scores may result in distorted results.
- Mean, SD, skew, kurtosis of scales get distorted if "Zero" is used as an anchor-value of Likert items (e.g. SPI, PWI, Insomnia Severity Index (ISI)). Better

could be to assign positive integers 1, 2, 3 … and so on to the levels avoiding zero, without disturbing nature of data.

- Summative Likert scores often generate tied scores and cannot discriminate the subjects with tied scores.
- Distribution of item scores and test scores are not considered to compute summative Likert scores.
- Mean and SD of Likert scales with *K*-number of levels (*K*-=3, 4, 5, …. *)* increase as *K* increases (Finn,1972) and may affect item/test parameters(Lim, 2008).

2.3 Transformation/normalization

Raw scores of selected variables in different units are usually scaled to have unit free values before aggregation. However, different methods of normalization can affect the CI. Kovacevic (2011) found $r_{Life\ expectancy,HDI} = 0.92 > r_{Life\ expectancy,GDP} = 0.71$. The inequality got reversed on taking logarithmic transformations. Illustrative methods of normalization in CI are:

i) Normalization to have equal score ranges by Min – Max function like

$Z = \frac{X - X_{Min}}{X_{Max} - X_{Min}}$ where $0 \leq Z \leq 1$.

Example: HDI, City Development Index (CDI), etc. It depends heavily on X_{Max}andX_{Min} and changes in X_{Max} or X_{Min}will change *Z*-scores and country rankings. Clearly, *Z*-score is relative to performance of others, Gain in *Z* due to unit increase in *X* varies for different values of *X*. If each indicator is improved uniformly by one unit, *Z*-scores and CI=ΣZ_i may remain unchanged. If the variable is in ratio scale, it has a non-arbitrary fixed zero point. Such zero-point gets altered by Min – Max function.

ii) Standardizing by $Z = \frac{X - Mean(X)}{SD(X)} \sim N(0,1)$ has been used in CIs like ESI, WHO index of health system performance (SPRG, 2001), etc. Mazziotta-Pareto Index (MPI) (2013) defines standardized matrix **Z** as

$$Z_{ij} = 100 \pm \frac{(S_{ij} - \bar{X}_i)}{S_{X_j}}.10$$

where the sign is "+" if the indicator is a positive dimension and "–"for negative dimension.

iii) $Z_i = \frac{X_i}{\bar{X}} \times 100$. It is also affected by outliers.

iv) $Z_i = \frac{X_i}{X_{Max}} \times 100$. This again depends on X_{Max}. Scores of each of 11 dimensions of BLI are normalized by relative score and not absolute measurement and thus, comparison of two countries may not be meaningful since their respective performances on each indicator depend on performances of all other countries.

v) For longitudinal data, assuming elimination of cyclical variability, standardization is done by:

$$Y_i^t = \frac{X_i^t - X_i^{t-1}}{X_i^t} \text{ where } t \text{ denotes time period}$$

Such standardization was used to assess potential benefits of the European internal market (European Commission, 2001).

vi) Logarithmic transformation by $Y_i = ln\,(X_i)$. CDI, SPI used such transformation for some indicators using natural log, while other indicators are in percentages.

vii) Percentage distance from base period: For a longitudinal data on year $t_0, t_1, \ldots\ldots, t_k$ CIW computed percentages $\frac{t_0}{t_0} \times 100$, $\frac{t_1}{t_0} \times 100$, $\frac{t_2}{t_0} \times 100, \ldots\ldots\ldots, \frac{t_k}{t_0} \times 100$ for each chosen positive indicator/dimension. For negative indicators/dimensions, standardization were done by taking transformation as $\frac{t_0}{t_0} \times 100$, $\frac{t_0}{t_1} \times 100$, $\frac{t_0}{t_2} \times 100$ and so on. Average of all such transformations was taken as CIW score of each year ensuring $\text{CIW} = 100 \text{ for } \mathrm{t}_0$. Considering n-indicators of which n_1 positive indicators and n_2 negative indicators, and assigning equal importance, CIW is computed by

$$\text{CIW} = \frac{1}{n}[\left(\sum_{i=1}^{n_i} \frac{X_i^t}{X_i^0} \times 100\right) + \left(\sum_{j=1}^{n_2} \frac{X_j^0}{X_j^t} \times 100\right)].$$

However, such non-linear transformations distort trends of the data. Addition of figures in percentages may give wrong results in construction and interpretation of CI as can be seen from the Table – 1 containing hypothetical data involving one positive indicator and one negative indicator where each indicator registered increasing trend:

Table -1

Illustrative CIW scores based on hypothetical data

Year	Original data		Transformed data using formula of CIW		Average
	Positive indicator	Negative indicator	Positive indicator	Negative indicator	
0	104	22	100	100	100
1	107	22.4	102.8846	98.21429	100.5495
2	114	24.3	109.6154	90.53498	100.0752
3	117	25	112.5	88.0	100.25
4	120	26	115.3846	84.61538	100
5	123	26.5	118.2692	83.01887	100.644

Clearly, CIW is not monotonic. Different methods of scaling/normalization may affect CIs differently. Thus, it is desirable to construct CI avoiding normalization.

2.4 Aggregation of the indicators:

Aggregation of indicators may have major implications for the CI. Various CIs used variety of aggregation methods. Addition of indicators assumes measurements are in interval or in ratio level and indicators are equally important i.e. perfect substitutability among the indicators, which can be criticized for arbitrariness. Perfect substitutability implies low value of an indicator can be compensated by high value of another indicator. ESI (2002) used simple AM. Dimension score of CWI is obtained by AM despite the indicators are in percentages and giving equal weights to the dimensions. ElSarawy (2016) also used average of normalized values in percentages of the domains. MPI used AM of normalized variables as $MPI_i^{+/-} = Mean_{Z_i}(1 \pm CV_i^2)$ where CV_i denotes coefficient of variation of the i-th indicator.

Here, sign of MPI_i depends on nature and direction of individual indicators. Higher mean and lower SD ⟺MPI is high.

2.5 Weighted sum approach

Here, CI = $\sum_{i=1}^{n} W_i X_i$ where $0 < W_i < 1$ and $\sum_{i=1}^{n} W_i = 1$ (Example: EPI). For CDI, Wellbeing Index (WI), Human Well-being Index (HWI), Ecosystem Well Being Index (EWI), CI = $\frac{1}{n}\sum_{i=1}^{n} W_i X_i$

Selected weights could be proportional to perceived importance and indicate 'trade-off' between pairs of criteria. $\frac{W_1}{W_2}$ may be interpreted as the amount to be sacrificed by indicator-2 to gain an extra unit of indicator -1. Changing weights to indicators affect CI of the countries being evaluated (Saisana et al. 2005) and can manipulate country rankings (Grupp and Schubert, 2010).

Appleby & Mulligan, (2000) decided weights to the indicators based on public preferences, but sum of the six weights exceeded unity and weight for one indicator was negative. $\sum_{i=1}^{n} W_i \neq 1$ violates the convex property of measurement and properties of CI as weighted sum are not known. Approaches to select weights to the indicators are: normative (subjective weights), data-driven and hybrid (Decancq and Lugo, 2013). PCA weights were used by SPI, internal market index in EU countries (European Commission, 2001b), composite sustainability indicators (Li, et al. 2012), etc. However, PCA weights ignore indictors which are poorly correlated with the CI even if they are practically important. Moreover, PCA deals with normally distributed scores, violation of which distorts results. For covariance matrix, PCA gives more weight to variables with larger variances. PCA weights are data specific and may vary across time. PCA approach to weighting was found inappropriate in constructing CI to measure business climate by OECD and the index of environmental sustainability (WEF, 2002). Use of PCA is not favoured by ESI 2002. For Economic Sentiment Indicator and index of environmental sustainability, PCA and FA failed (Nardo et al.2005). Sava (2016) found PCA weights for data accumulated for 11-years which were different from the PCA weights found after averaging year-wise data. LPI Report 2014 used PCA weights where sum of weights$>$ 1.

No weight or equal weights are wrong (Greco et al. 2017; Mikulić et al. 2015). Equal weights treat equally the essential indicators and less important indicators. It assumes a constant trade-off rate between each pair of criteria (Tofallis, 2014). GNH assigned 50%, 40%, 33%, 30%, 20%. 17% and 10% weights to indicators under each dimension but weights to the dimensions were equal. Subjective indicators have been given lower weights than objective indicators. HPI attempts to combine human well-being and environmental impact gives high weights for the carbon footprint. Thus, HPI is higher for countries with lower ecological footprints which try to measure human demand on nature is a much-criticized concept. Hartung et al.(2008) mentioned about ideal weights. Chakrabartty (2017) found the weight vector $W = (W_1, W_2, \ldots\ldots, W_n)^T$ satisfying $\sum_{i=1}^{n} W_i = 1$ and minimizing variance of the weighted sum $Y = \sum_{i=1}^{n} W_i X_i$. The author proved that if $X_1, X_2, \ldots\ldots\ldots, X_n$ are replaced by standardized scores $Z_{ij} = \frac{X_{ij} - \overline{X_j}}{s_{X_j}}$, then Y is equi-correlated with each Z_i $\forall\, i = 1, 2, \ldots.., n$. However, determination of methodologically sound weights for computation of CI as weighted sum is a difficult task (Yang et al.2018) and there is no weighting system which is beyond criticism (Greco, et al.2019).

2.6 Methods other than weighted sums:

Ease of doing business computes distance to frontier score from 12 areas of business regulations and aggregates such distances (Meng et al. 2013). But distance may not be invariant under change of scale. Blancard and Hoarau, (2013) obtained CI using data envelopment analysis (DEA), to develop an efficiency frontier by optimizing the weighted output/input ratio, ensuring that this ratio $\ngtr 1$. However, DEA results vary with changes in number of input variables and output variables. The best specification cannot be tested. DEA cannot be applied directly where indicators are outputs, with no data on inputs (Herman, 2008). Moreover, countries differ in policy goals for different areas and use of resources and thus, create problems in computation of weights.

Yang et al. (2018) developed model to have a common set of weights for constructing growth potential index for emerging market, using DEA approach which performed equally to the average of each indicator transformed to Z-scores (a reference point) and derived efficiency scores by minimizing the sum of the differences between the countries that are above average and those that are below average. The model suffers from limitations of DEA models. Fusco (2015) proposed Directional Benefit-of-Doubt using DEA and PCA. CI by this approach is heavily affected by outliers, like DEA and PCA approaches.

Multiple Linear Regressions presumes linear relationship of the dependent variable and the set of independent variables, which is rather rare with CIs (Saisana et al. 2005). Moreover, the dependent variable is not observable. Porter and Stern (2001) considered logarithmic transformation of the dependent variable in the National Innovative Capacity Index. Ács et al. (2014) proposed 'penalty for a bottleneck' for penalizing unbalances in the CI. Here, 'penalty' amount depends on the data set, presence of outliers and thus, the solution is not always optimal. Tarabusi and Guarini (2013) addressed penalization of the unbalances in CI using mean–min function. However, estimation of parameters of the model for proper penalization is an issue.

GNH is calculated as $(1 - H_n A_n)$ where H_n denotes percentage of "Not-yet-happy" people and $A_n$denotes percentage of dimensions in which $H_n -$ people enjoy sufficiency. Higher GNH $\Longrightarrow$ greater happiness. Computation of HPI involves Atkinson Index of experienced wellbeing, Life expectancy, Ecological footprints and four other parameters. However, Atkinson inequality index (Atkinson, 1970) assumes higher the mean income, higher the social welfare and is influenced heavily by country-level socio-ecological indicators like output per worker, number of pensioners, proportion of persons with secondary education and above, etc. Moreover, measures of life-span inequality do not distinguish between healthy and unhealthy years of life (Murray et al. 1999) and weak positive associations between Gini's measure of income inequality and Atkinson index of Life expectancy across all ages and weak negative associations among those aged ≥ 65 suggest no association between income inequality and lifespan inequality (Hertog, 2013).

HDI has shifted from arithmetic aggregation to geometric aggregation from 2010 and thus avoided perfect substitution across dimensions (UNDP 2010). Living Planet Index by Loh, (2002) also uses geometric aggregation considering ratios of indicator at successive time periods.

Just like no consensus of perfect weighting scheme, no perfect aggregation scheme is there (Arrow and Raynaud, 1986). Tofallis (2014) addressed the issues of additive aggregation and also multiplicative aggregation and favoured the later to avoid pitfalls of the former. Segovia-González and Contreras (2023) proposed geometric aggregation to evaluate gender effect in educational systems of OECD countries. McDonnell et al. (2023) reviewed methodologies at each stages of development of CI for quality and safety in healthcare and found no universally agreed approach on the design, development and reporting of such CIs.

3. Proposed method

Ignoring the stage of selection of indicators, the proposed method of CI avoiding scaling and selecting weights are described below:

3.1 Pre-processing of data

I: Ensure that increase in each selected indicator increases CI. For negative indicator, if any, consider reciprocal of values.

II. Mark the levels of items as 1, 2, 3, 4, 5, and so on avoiding zero

3.2 Method

For a country, let value of the *i*-th indicator at *t*-th time be $X_{it} > 0$ $\forall\, i = 1,2,\ldots n$ and value of the same indicator in the base period be X_{i0}. The unit free ratio $\frac{X_{it}}{X_{i0}}$ reflects progress or decline with respect to the *i*-th indicator at *t*-th time period in comparison to the base period. CI for the country at the current time-period is proposed as the Geometric mean of the ratio $\frac{X_{it}}{X_{i0}}$

$$CI_{c0} = \sqrt[n]{\frac{X_{1c}, X_{2c}, \ldots\ldots, X_{nc}}{X_{10}\, X_{20}\, \ldots\ldots X_{n0}}} \qquad (1)$$

or avoiding the *n*-th root, $CI_{c0} = \frac{X_{1c}, X_{2c}, \ldots\ldots, X_{nc}}{X_{10}\, X_{20}\, \ldots\ldots X_{n0}}$ (2)

CI_{c0} exceeding one, as per equation (2) implies overall improvement registered by the country from the base year. Similarly, $\frac{CI_{i_t}}{CI_{i_{(t-1)}}} > 1$ indicates the progress in period *t* over the previous period. For two successive time periods, critical indicators are those with $\frac{X_{it}}{X_{i(t-1)}} < 1$ and can be ordered for policy purpose to decide appropriate action. Each of (1) and (2) can cover all indicators measured in ordinal scale or ratio/interval scale or those in percentages and depicts overall improvement/decline in the current year from the base year by a unit-free function which is continuous and monotonically increasing. Growth of CI on year-to-year basis can be obtained by considering the previous year instead of the base period.

Properties of CI by equation (2):

1. Unaffected by change of scale in one or many indicators

2. Possible to plot curve showing percentage gain in the *i*-th indicator and corresponding percentage gain in CI.

3. Avoids scaling, selection of weights and significantly reduces substitutability among the component indicators.

4. Avoids effect of outliers implying no bias for either developed or under-developed countries

5. Satisfies time-reversal test since $CI_{t0}.CI_{0t} = 1$

6. Facilitates formation of chain indices since$CI_{20} = CI_{21}.CI_{10}$. This helps to plot path of a country to achieve overall progress across time which may help in inter-regional comparisons.

7. Facilitates construction of dimension-wise indices considering indicators relevant to that dimension. Aggregation of dimensions to get CI requires no selection of weights.

8. Easy to find relative importance of each indicator.

9. Facilitates estimation of population *GM*, since $\log GM = \frac{1}{n}\sum_{i=1}^{n} \log Y_i$ where $Y_i = \frac{X_{it}}{X_{i0}}$ Geometric standard deviation (GSD) (S_{GM}) is given by

$\log S_{GM} = [\frac{1}{n}\sum_{i=1}^{n}(log\ Y_i - \log GM)^2]^{\frac{1}{2}}$

Thus, log (GSD of $Y_1, Y_2, \ldots \ldots, Y_n$) = SD of log Y_1, log Y_2,, ………, logY_n

For large sample, population estimate of *GM* is sample *GM* and estimate of standard error of the *GM* is $GM.(\frac{\text{Log}\, S_{GM}}{\sqrt{n-1}})$. Confidence interval of GM for (1 – α) % are (e^U, e^L) where U=$logGM$ + $S_m.t_{(\frac{\alpha}{2},df)}$ and L=$logGM$- $S_m.t_{(\frac{\alpha}{2},df)}$ (Alf and Grossberg, 1979).

Thus, H_0: $CI_{ti} = CI_{tj}$ for two different countries i and j or H_0: $CI_{ti} = CI_{(t-1)i}$ for the *i*-th country can be tested by *t*-tests using logarithms of the observations.

3.3 Benefits

- The proposed CI is applicable even for skewed data for two different time periods. Thus, it facilitates meaningful comparison of a set of regions/countries.
- Sub-groups of a country like ethnic groups, economically/socially backward groups, elderly people with specific morbidity, etc., can be compared by the proposed index with pre-determined indicators.
- Can be well used for classification of countries.
- The plot of progress/decline of CI over time helps to find impact of various socio-economic measures adopted over time.

4. Limitations:

In case a new indicator is induced subsequently, values of the new indicator need to be estimated since the base period. GM fails if value of an indicator ≤ 0

5. Conclusions:

The existing methods of measuring CI have several disadvantages. The proposed CI as function of geometric mean avoids scaling, weighting and compensability inherent in the linear aggregation setting. It can consider all chosen indicators and corresponding dimensions to reflect overall improvement/decline of a country in current year from the base year or previous year. Thus (2) avoids the criticisms on weighting, aggregation and possibly selection of indicators. It helps in undertaking statistical tests to infer significant difference if any, in overall improvement between a pair of countries at a given time period or a region/country at two different time periods. In addition, it assesses improvement-path registered by each country over time.

It satisfies desired properties like unit-free monotonic continuous function, time-reversal test, formation of chain indices, identification of the critical dimensions/indicators requiring attention and can be applied irrespective of distributions of the indicators. Besides, the measure helps in estimation of population *GM*, confidence interval of *GM* and statistical testing of H_0: $CI_{ti} = CI_{tj}$ for two countries i and j where $i \neq j$ at time period *t* or H_0: $CI_{ti} = CI_{(t-1)i}$ for the i-th country.

The proposed measure satisfying the above said desired properties and avoiding the criticisms of linear aggregation is an improvement over the existing methods. However, still there is room for improvement. Simulation study using multi data set is suggested to investigate finer points including robustness of the proposed CI and other recent methods and techniques.

Declaration

Acknowledgement: Nil

Funding: No funds, grants, or other support was received

Conflicts of interest/Competing interests: The author has no conflicts of interest to declare

Financial or non-financial interests: Nil

Availability of data and material: Nil. The article did not use any dataset

Code availability: No application of software package or custom code

CRedit statement: Conceptualization; Methodology; Analysis; Writing and editing the paper by the Sole Author

References:

Ács, Z., Autio, E., & Szerb, L. (2014): National systems of entrepreneurship: Measurement issues and policy implications. *Research Policy,* 43(3), 476–494.

Alf, EF and Grossberg, JM. (1979): The geometric mean: Confidence limits and significance tests, *Perception &Psychophysics*, 26 (5), 419-421

Annoni, P. & Lewis Dijkstra, L.(2017): Measuring and monitoring regional competitiveness in the European Union, in: Robert Huggins & Piers Thompson (ed.), Handbook of Regions and Competitiveness, chapter 3, pages 49-79, Edward Elgar Publishing.

Appleby, J. & Mulligan, J. (2000) *How well is the NHS performing? A composite performance indicator based on public consultation*, King's Fund: London**.**

Arrow, KJ & Raynaud, H.(1986): *Social choice and multicriterion decision-making*. Cambridge: MIT Press.

Arvis, JF, Saslavsky, D, Ojala, L, Shepherd, B, Busch, C, Raj, A, & Naula, T. (2016): *Connecting to Compete: Trade Logistics in the Global Economy,* World Bank, Washington DC

Atkinson AB (1970): On the Measurement of Inequality, *Journal of Economic Theory*, 2, 244-263.

Beltrán-Esteve M, Peiró-Palomino J., Andrés J. Picazo-Tadeo, Vicente Rios (2023): Is the European Social Progress Index robust? Implications for the design of European Union regional Cohesion Policy. https://doi.org/10.6084/m9.figshare.21966949.v1

Birdsall, Nancy and David Roodman (2003): *The Commitment to Development Index: A Scorecard of Rich Country Policies*, Center for Global Development, Washington, DC

Blancard, S. & Hoarau, JF. (2013): A new sustainable human development indicator for Small Island developing states: A reappraisal from data envelopment analysis, *Economic Modelling,* 30(C), 623-635

Bohringer, C & Jochem, PE (2007): Measuring the immeasurable: A survey of sustainability indices. *Ecological Economics, 63*(1), 1–8.

Brogi, E. (2020). The Media Pluralism Monitor: Conceptualizing media pluralism for the online environment. Profesional De La information, 29(5). DOI: 10.3145/ epi.2020.sep.29

Cella D, Choi SW, Condon DM, Schalet B, Hays RD, Rothrock NE, Yount S, Cook KF, Gershon RC, Amtmann D, DeWalt DA, Pilkonis PA, Stone AA, Weinfurt K, Reeve BB. (2019): PROMIS® Adult Health Profiles: Efficient Short-Form Measures of Seven Health Domains. *Value Health*.22 (5):537-544. doi: 10.1016/j.jval.2019.02.004.

Cendrero, A., Francés, E., Del Corral, D., Fermán, JL, Fischer, D., Del Río, L., Camino, M., and López, A. (2003): Indicators and indices of environmental quality for sustainability assessment in coastal areas: Application to case studies in Europe and the Americas. *Journal of Coastal Research,* 19(4), 919-933.

Chakrabartty, Satyendra Nath (2017): Composite Index: Methods and Properties, *Journal of Applied Quantitative Methods,* 12(2), 31-41

Competitiveness: a critique of the Institutionalist literature, *Socio-Economic Review*, 6(4), 637–669.

Cummins, R., Eckersley, R., Pallant J., Van Vugt, J., & Misajon, R. (2003): Developing a national index of subjective wellbeing: The Australian unity Wellbeing Index. *Social Indicators Research*, 64(2), 159-190. doi:10.1023/A:1024704320683

Decancq, K., & Lugo, MA (2013): Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews, 32*(1), 7–34.

Distaso, A. (2007): Well-being and/or quality of life in EU countries through a multidimensional index of sustainability. *Ecological Economics*, *64*(1), 163-180.

Djankov, Simeon (2016). The Doing Business Project: How It Started: Correspondence. *Journal of Economic Perspectives*, 30 (1): 247-48.DOI: 10.1257/jep.30.1.247

ElSarawy, MM (2016): Use of new technique to measure wellbeing index for North Africa countries in year 2012, Presented in the 34th IARIW General Conference Dresden, Germany, August 21-27.

Environmental Sustainability Index (2002): *Yale Center for Environmental Law & Policy*, New Haven.

EPI Report (2016): *Global Metrics for the Environment,* http://www.epi.yale.edu

European Commission (2001) *Summary Innovation Index*, DG Enterprise, European Commission: Brussels.

European Commission (2001b): *Internal Market Scoreboard*, DG MARKT, European Commission: Brussels

Finn, RH (1972): Effects of some variations in rating scale characteristics on the means and reliabilities of ratings. *Educational and Psychological Measurement*, 32(7), 255–265.

Freudenberg, M. (2003): *Composite indicators of country performance: A critical assessment*, OECD STI Working paper DSTI/DOC 2003/16, OECD: Paris.

Fusco, E. (2015): Enhancing non-compensatory composite indicators: A directional proposal, *European Journal of Operational Research, 242*(2), 620–630.

Greco, S, Ishizaka, A, Tasiou, M and Torrisi, G. (2019): On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness, *Soc Indic Res* 141:61–94 https://doi.org/10.1007/s11205-017-1832-9

Greco, S., Ishizaka, A., Matarazzo, B., & Torrisi, G. (2017): Stochastic multi-attribute acceptability analysis (SMAA): An application to the ranking of Italian regions. *Regional Studies*. https://doi.org/10.1080/00343404.2017.1347612

Grupp, H., & Schubert, T. (2010): Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy,* 39(1), 67–78.

Hand, D. J.(1996): Statistics and the Theory of Measurement, *J. R. Statist. Soc. A*; 159, Part 3, 445-492

Hartung, J., Knapp, G. and Sinha, B.K. (2008): *Statistical Meta-Analysis with Applications*.

Herrmann, M. (2008): Rethinking the link between labor market flexibility and corporate

Hertog, Sara (2013): The Association between Two Measures of Inequality in Human Development: Income and Life Expectancy, Technical Paper No. 2013/7, Population Division, United Nations

Hobart JC, Cano SJ, Zajicek JP, Thompson AJ (2007): Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *Lancet Neurol* 6(12):1094–1105

International Wellbeing Group (2006): *Personal Wellbeing Index (Manual).* Melbourne: Deakin University, Australian Centre on Quality of Life.

International Wellbeing Group (2013): *Personal Wellbeing Index*: 5th Edition. Melbourne: Australian Centre on Quality of Life, Deakin University. http://www.acqol.com.au/instruments#measures

Jacobs, R., Smith, P. and Goddard, M. (2004): Measuring performance: An examination of composite performance indicators, *CHE Technical Paper Series 29,* University of York

Jamieson, S. (2004): Likert scales: How to (ab) use them. *Medical Education,* 38, 1212 -1218

John Wiley & Sons, Hoboken, 44. https://doi.org/10.1002/9780470386347

Johnston, BT & Sheehy, TP (1995): *The Index of Economic Freedom*. Heritage Foundation, Washington, pp. ix-21

Kasparian, J, and Rolland, A. (2012): OECD's Better life index: can any country be well ranked? *Journal of Applied Statistics*, 39 (10), 1-8, DOI: 10.1080/02664763.2012.706265

Kaufmann, D., Kraay, A. and Mastruzzi, M. (2007): Growth and Governance: A Reply, *Journal of Politics,* 69(2), 555-562

Khan, F., Sadiq, R. and Veitch, B., (2004): Life Cycle iNdeX (LInX): a new indexing procedure for process and product design and decision-making. *Journal of Cleaner Production*. 12, 59–76

Kovacevic, M. (2011): *Review of HDI Critiques and Potential Improvements*, The Human Development Research Paper 2010/33, UNDP

Lafortune, G., Fuller, G., Bermont Diaz, L., Kloke-Lesch, A., Koundouri, P., Riccaboni, A. (2022). Achieving the SDGs: Europe's Compass in a Multipolar World. Europe Sustainable Development Report 2022. France: Paris.

Lambsdorff, JG. (2006): Causes and Consequences of Corruption: What Do We Know from a Cross-Section of Countries? Chapters in: Susan Rose-Ackerman (ed.), International Handbook on the Economics of Corruption, chapter 1, Edward Elgar Publishing.

Li, T., Zhang, H., Yuan, C. et al.(2012): A PCA-based method for construction of composite sustainability indicators, *Int J Life Cycle Assess,* 17: 593.

Lim, Hock-Eam (2008): The Use of Different Happiness Rating Scales: Bias and Comparison Problem? *Soc Indic Res* . 87:259–267

Loh, J. (Ed.) (2002): *Living Planet Report 2002*, World Wildlife Fund, Gland, Switzerland

Matteo Mazziotta, M. and Pareto, A. (2013): Methods for constructing composite indices: One for all or All for one? Rivista Italiana di Economia Demografia e Statistica, 67(2), 67-80

McDonnell T, Cosgrove G, Hogan E, Martin J, McNicholas T, O'Dowd M, Rizoaica F, McAuliffe E.(2023): Methods to derive composite indicators used for quality and safety measurement and monitoring in healthcare: a scoping review protocol. BMJ Open. 14;13(7):e071382. doi: 10.1136/bmjopen-2022-071382.

McGillivray M, White H. (1993): Measuring development? The UNDP's human development index. *J Int Dev.*5:183–92. doi: 10.1002/jid.3380050210

Meng, FY, Fan, LW, Zhou, P. and Zhou, DQ (2013): Measuring environmental performance in China’s industrial sectors with non-radial DEA, *Mathematical and Computer Modelling*, 58(5-6), 1047–1056.

Michalos, A., Smale, B., Labonté, R., et al. (2011): The Canadian Index of Wellbeing. *Technical Report 1.0*. Waterloo: CIW.

Mikulić, J., Kožić, I., & Krešić, D. (2015): Weighting indicators of tourism sustainability: A critical note. *Ecological Indicators, 48,* 312–314.

Murray CJ, Gakidou EE, Frenk J. (1999): Health inequalities and social group differences: what should we measure? Bull World Health Organ. 77(7):537-43

Nardo, M., Saisana, M., Saltelli, A. and Tarantola, S. (2005): Handbook on Constructing Composite Indicators: Methodology and User Guide, *OECD Statistics Working Papers*, 2005/3, OECD Publishing.

NEF (2008): National Accounts of Well-being: Bringing Real Wealth onto the Balance Sheet. www.nationalaccountsofwellbeing.org/public-data/files/national-accounts of-well-being-report.pdf

OECD (2008): *Handbook on constructing composite indicators: Methodology and user guide*. Paris: OECD Publishing.

OPHI (2018): *Global Multidimensional Poverty Index 2018: The Most Detailed Picture to Date of the World’s Poorest People.* Oxford Poverty and Human Development Initiative, University of Oxford.

Orff, HJ, Drummond, SP, Nowakowski, S, Perlis, ML (2007): Discrepancy between subjective symptomatology and objective neuropsychological performance in insomnia. *Sleep,* 30, 1205–1211.

Pichon E, Widuto A, Dobreva A, Jensen L.(2021): Ten composite indices for policymaking. European Parliamentary Research Service (EPRS). www.europarl.europa.eu/RegData/etudes/IDAN/2021/6962 03/EPRS_IDA (2021)696203_EN.pdf

Porter, ME and Stern, S (2001): Innovation: Location Matters. *MIT Sloan Management Review*, 42, 28-36.

Ryff, CD and Keyes CLM (1995): The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69(4):719–727

Saisana,M., Saltelli, A. and Tarantola, S. (2005): Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators, *Jr. of Royal Statistical Society Series A*, 168(2),1-17

Sava, AM.(2016): Weighting method for developing composite indices. Application for measuring sectoral specialization, *Journal of Applied Quantitative Methods*,11(3),77–84.

Scientific Peer Review Group (SPRG) (2001): *Report of the Scientific Peer Review Group on Health Systems Performance Assessment*, WHO: Geneva. http://www.who.int/health-systemsperformance

Segovia-González, MM and Contreras, I. (2023): A Composite Indicator to Compare the Performance of Male and Female Students in Educational Systems. *Soc Indic Res* 165, 181–212. https://doi.org/10.1007/s11205-022-03009-1

Smith, P. (2002): Developing composite indicators for assessing health system efficiency, in Smith, P.C. (ed.) *Measuring up: Improving the performance of health systems in OECD countries*, OECD: Paris.

Tarabusi, C., & Guarini, G. (2013): An unbalance adjustment method for development indicators. *Social Indicators Research, 112*(1), 19–45.

Tofallis, Chris (2014): Add or Multiply? A Tutorial on Ranking and Choosing with Multiple Criteria. INFORMS Transactions on Education 14(3):109-119. http://dx.doi.org/10.1287/ited.2013.0124

UNCHS (2001): *The State of the Worlds Cities*, UN, Nairobi.

UNDP (2007): Fighting climate change: Human solidarity in a divided world. *Human development report 2007/2008*. New York: Palgrave Macmillan

UNDP (2010): *Human development report: The real wealth of nations: Pathways to human development*. New York.

Wackernagel, M.; Rees, W (1996): *Our Ecological Footprint: Reducing Human Impact on the Earth*; New Society Publishers: Gabriola, BC, Canada

Ware, JE & Sherbourne, CD (1992): The MOS 36-item Short-Form Health Survey (SF-36): Conceptual framework and item selection. *Medical Care*, 30, 473-83.

World Economic Forum (WEF) (2002): An initiative of the Global Leaders of Tomorrow Environment Task Force. Pilot Environment Performance Index. www.ciesin.columbia.edu/indicators/ESI/EPI2002_11- FEB02.pdf

Yang, FC, Kao, RH, Chen, YT, Ho, YF, Cho, CC, & Huang, SW (2018): A common weight approach to construct composite indicators: The evaluation of fourteen emerging markets. *Social Indicators Research*, 137, 463 – 479. https://doi.org/10.1007 /s112 05-017-1603 -7.