Behind the Number: A Review of Index Methodologies to Improve Innovation Measurement in Africa

Islam Hassouna

Working Paper 11
Published: 8 February 2018

openAIR
AFRICAN INNOVATION RESEARCH

Author

Islam Hassouna
Consultant, Access to Knowledge for Development Center (A2K4D)
The American University in Cairo (AUC)

Abstract

This paper reviews the methodologies of 16 indices in innovation, information and communication technologies, economic environment, governance, and development. It looks at the different techniques used by these indicators to aggregate data into a single number. The paper presents index structure, data, weighing of indicators, assessment, and ends with a focus on the measurement of innovation in the reviewed indices.

Acknowledgements

The author gratefully acknowledges the team at A2K4D especially Dr. Nagla Rizk and Mennatalah Badreldin for their continuous support.

Keywords

Index, Methodology, Innovation, Weights, Aggregation, Assessment.

I. Introduction

Indices employ the power of numbers to reflect and influence societies. Despite the subjectivity that is included in the formulation of that number, it gathers a form of objectivity by virtue of its existence. The objectivity arises from the quantitative breakdown of what is perceived to be abstract aspects. The societal aspects indices attempt to quantify are abstract either due to their qualitative nature, complexity, or difficulty in measurement. However, their quantification makes them concrete. That materializes their status and specifies areas of weakness, making these aspects harder to ignore by policy makers.

The different ways to compose an index range from what to include in an index, to how to gather the data, to the procedure of calculating the final number. This paper presents the methodologies of 16 indices as an effort to review the common ideas used in index development. The methodological areas that will be covered are index structure, data preparation and manipulation, qualifications for data to be included and imputation techniques, methods used for assigning weights, and techniques of index assessment. Following that, there will be a review of how innovation is measured in the indices covered.

The ideas gathered in this paper will form a starting reference to a sequel paper. That paper will be a proposed methodology of an index that is to be developed by Access to Knowledge for Development

Center (A2K4D) at the American University in Cairo (AUC). A2K4D is developing the index in partnership with Open Africa Innovation Research network (Open AIR) as part of its efforts to focus "on finding alternative ways to assess knowledge production and use in African contexts, with special focus on innovation."[1] Although the index will attempt to measure innovation in the developing world, this paper reviews indices in a number of different fields. The indices reviewed in this paper are shown in Table 1 categorized in their broad fields.

The mathematical scope of this paper does not go beyond a reference to the mathematics used in the methodologies reviewed. A general explanation will usually be given to show the purpose of the mathematics used, but not an explanation of the mathematics itself. Where the explanation of the mathematics is understood, and it is deemed significantly relevant to context, an explanation of the mathematics will be provided as an exception.

Table 1: Indices reviewed

	Index	Report	Publisher
Innovation Indices			
1	Global Innovation Index (GII)	The Global Innovation Index 2016: Winning with Global Innovation	Cornell University, INSEAD, WIPO
2	Summary Innovation Index (SUII)	European Innovation Scoreboard 2016	EU European Commission
3	Social Innovation Index (SII)	Old problems, new solutions: Measuring the capacity for social innovation across the world, 2016	The Economist Intelligence Unit
Information and Communication Technology (ICT) Indices			
4	Network Readiness Index (NRI)	The Global Information Technology Report 2016: Innovating in the Digital Economy	Cornell University, INSEAD, World Economic Forum
5	ICT Development Index (IDI)	Measuring the Information Society Report 2016	International Telecommunication Union
6	Compliance Gap (CG)	The Compliance Gap: BSA	The Software Alliance (BSA)

[1] Open Air, "Open AIR: Metrics and Policies", www.openair.org.za

		Global Software Survey, May 2016	
Economic Environment Indices			
7	Distance to Frontier (DTF)	Doing Business 2017: Equal Opportunities for All	World Bank
8	Global Competitiveness Index (GCI)	The Global Competitiveness Report 2015-2016	World Economic Forum
Governance Indices			
9	Corruption Perception Index (CPI)	Corruption Perception Index 2016	Transparency International
10	Transformation Index (BTI)	BTI 2016 Codebook for Country Assessments	Bertelsmann Stifung
11	African Capacity Index (ACI)	African Capacity Indicators 2013: Capacity Development for Natural Resource Management	African Capacity Building Foundation
Development Indices			
12	Human Capital Index (HCI)	The Human Capital Report 2016	World Economic Forum
13	Social Progress Index (SPI)	Social Progress Index 2016	Social Progress Imperative
14	Human Development Index (HDI)	Human Development Report 2016	UNDP
15	Inequality-adjusted Human Development Index (IHDI)	Human Development Report 2016	UNDP
16	Multidimensional Poverty Index (MPI)	Human Development Report 2016	UNDP

II. Structure

The structure of the indices has an important role in their formulation. There are different levels of considerations that each index utilizes. All indices gather data as indicators at the most basic level. These indicators are then grouped into hierarchies of different levels. The groupings are sometimes used as a framework for calculations, and at other times as a presentation of conceptual classifications. The level of hierarchies varies in number from seven levels to a single level of just ungrouped indicators. The hierarchal setup of indices also relates to the different outputs of their reports. Some publishers provide further indices in addition to the main index, which are based on the index structure.

Publishers sometimes offer a statistical analysis of how sound an index is. The structure of an index is an integral component of the analysis. It explains the conceptual framework of the index in terms of the indicators used and their roles is reflected in the statistical properties of the index. That analysis covers two broad perspectives. The first one is to what extent the indicators and their groupings are distinct from one another, and to what extent do they contribute to the phenomenon the index claims to measure. Although it is not the defining analysis of an index, it gives a view of its quality.

The language used for each level of hierarchy varies between publishers. The language for different levels in this report will largely follow that of the index being discussed.

A. Index Levels and Composing Methods

This section provides the hierarchy of the indices and computational steps taken at each level to arrive at the overall index. Almost all of the indices reviewed employ the use of the weighted average in their calculations. The details and methods of which will be discussed later in the paper.

Frequent mention will be made to the concept of a simple average and weighted average in this section. A simple average is the common arithmetic average. The weighted average (AW) is a generalization of the simple average calculated as

$$AW = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

where x_i is a sequence of numbers of interest, and w_i is referred to as the weight given to number x_i in calculation of the weighted average. In most cases the denominator sums to 1.

B. Indices with Commonly Used Structures

GII (Global Innovation Index)

The GII considers 128 countries by gathering 82 indicators into an index range of 0 to 100. These indicators are first grouped into 21 sub-pillars. The sub-pillars are then grouped into seven pillars with each containing three sub-pillars. Pillars are then grouped into two sub-indices, with five pillars grouped as the Innovation Input Sub-Index and two pillars into the Innovation Output Sub-Index.[2] The GII structure is shown in Figure 1.

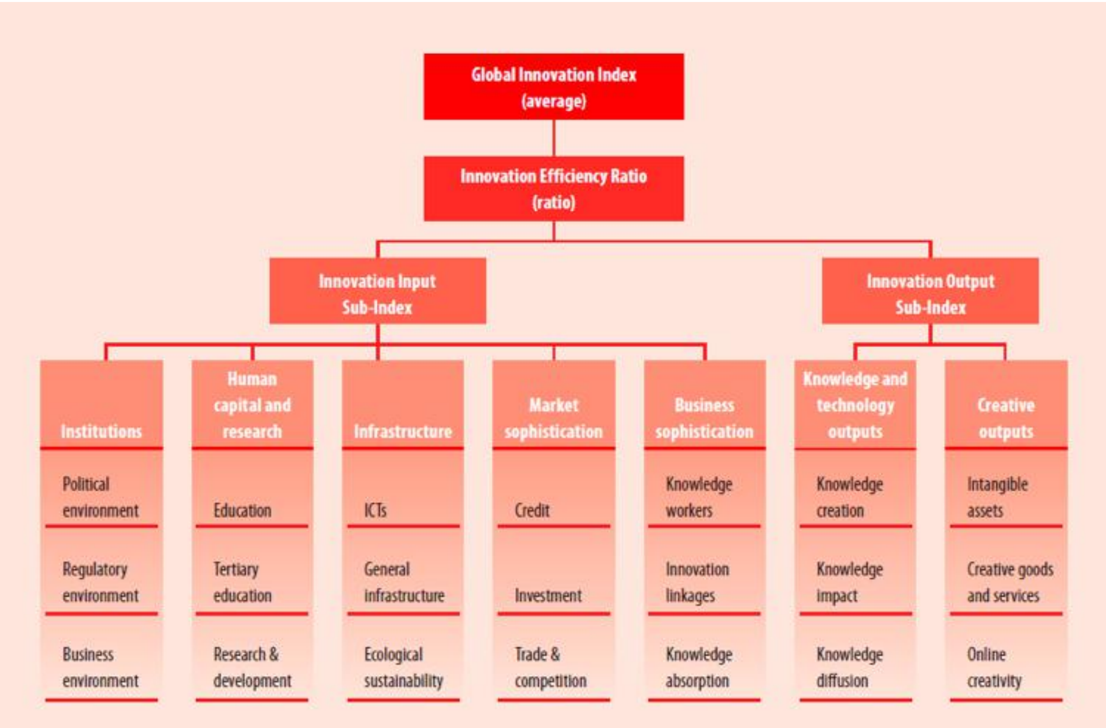

Figure 1: GII structure. Source: Global Innovation Index 2016: Winning with Global Innovation, page 14.

The calculation of the index starts at the sub-pillar level, with each sub-pillar allocated a score of mostly the simple average of its indicators, with a few cases where explicit weights are given to specific indicators. A simple average is then taken to calculate the pillars, followed by a simple average of the pillars to calculate the two sub-indices. Then the GII is calculated as the simple average of the Innovation Input and Output sub-indices. GII report presents four outputs through that structure. They are the GII and its two sub-indices, and the Innovation Efficiency Ratio, which is the ratio of the Innovation Output to the Innovation Input.[3]

SII (Social Innovation Index)

The SII has a similar structure to the GII. It measures social innovation in 45 countries through 17 indicators. These indicators are grouped into four categories. Each category is calculated using mostly a weighted average of its indicators. The SII is also the weighted average of the four categories.[4]

NRI (Network Readiness Index)

[2] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. "The Global Innovation Index 2016: Winning with Global Innovation". Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.
[3] Ibid.
[4] "Old problems new solutions: Measuring the capacity for social innovation across the world: Social Innovation Index 2016". The Economist Intelligence Unit, 2016.

The NRI has an identical hierarchy as the GII with a difference in the range of the index of 1 to 7 instead of 0 to 100, and a different number of indicators.[5]

DTF (Distance to Frontier)

The DTF collects 120 indicators about 190 countries into 11 indicator sets each representing a regulatory area, with 10 sets used in the DTF and the Labor Market Regulation collected for presentation purposes. Aggregation for the DTF is done also for each indicator set first, and then across indicator sets using mostly simple averages.[6]

C. Indices with Mostly Unequal Weights

GCI (Global Competitiveness Index)

GCI has a high number of levels where it groups 114 indicators into categories, sub-pillars, pillars, and three sub-indices. However, not all the levels are available for all indicators. The indicators of the Innovation and Sophistication Factors sub-index are grouped into pillars directly; other sub-pillars group indicators directly without grouping them into categories first. Calculation is done at the lowest grouping level of indicators available using mostly simple averages at each level until the sub-index level. The calculation of the GCI from the three sub-indices of Basic Requirements, Efficiency, and Innovation and Sophistication is done using a variable weighted average. Where each sub-index is given a different weight depending on the development of the countries considered.[7]

HCI (Human Capital Index)

HCI combines 46 indicators for 130 countries. The indicators are grouped into categories, and categories are grouped into two themes of learning and employment across 5 vertical group pillars according to age. So, each age group pillar has two sub-pillars of learning and employment. Although the themes are similar, the categories of each of the themes vary across age groups. Calculation of the HCI starts with the simple average of the categories up to the scores of each age pillar. HCI is the weighted average of the age pillars with each pillar having a specific weight.[8]

BTI (Transformation Index)

The BTI Transformation Index is actually a publication of two indices ranging from 1 to 10 that are left uncombined. BTI reports on 129 countries through 49 indicators that are grouped into 17 criteria, which are grouped into democracy status, market economy status, and management performance dimensions. Calculation is carried as a simple average at each level. With the Status Index being the average of the democracy and economy status, and the Management Index is the management performance weighted by a level of difficulty. The Management Index takes into account how

[5] Baller, Silja, Soumitra Dutta and Bruno Lanvin. “Insight Report: The Global Information Technology Report 2016: Innovating in the Digital Economy”. Johnson Cornell University, World Economic Forum, and INSEAD, 2016.
[6] “Doing Business 2017: Equal Opportunity for All”. World Bank, 2017.
[7] Schwab, Klaus and Xavier Sala-i-Martin. “Insight Report: The Global Competitiveness Report 2016-2017”. World Economic Forum, 2016.
[8] “Insight Report: The Human Capital Report 2016: Technical Notes”. Mercer and World Economic Forum, 2016.

difficult management in a country is, so the higher the difficulty the higher the enhancement of the management performance to the Management Index.[9]

SPI (Social Progress Index)

The SPI has the usual hierarchy, where it reports on the wellbeing of 133 countries without the use of economic indicators. It is made up of 53 indicators that are grouped into 12 components, which are grouped into three dimensions with four components each. Each component score is a weighted average of its measures. The weights of measures in each component are decided by statistical analysis post data gathering. Then the calculation is carried as a simple average of components in each dimension, and the simple average of dimensions to compute the SPI.[10]

IDI (ICT Development Index)

IDI reports on 175 countries on a scale of 1 to 10 through 11 indicators that are grouped into three sub-indices of ICT Access, ICT Usage, and ICT Skills. Each sub-index is calculated as the simple average of its indicators. The IDI is then calculated as the weighted average of its sub-indices. The weights of the sub-indices were decided after a statistical analysis carried out in the first edition.[11]

D. Statistical Analysis Shaping of Structure and Harmonic Mean

ACI (African Capacity Index)

ACI stands out in its structure in its grouping of indicators and in its aggregation computation. The indicators collected for 44 countries are grouped into 4 clusters of policy environment, processes of implementation, development results, and capacity development outcomes. The four clusters were created through statistical analysis of the data collected in the first edition of the index report.[12] So, the data decided how it would be allocated, and not as it commonly is that the design allocates the data. It is a more statistically sound approach, as statistical testing of indices usually compares statistical grouping of indicators with the conceptual grouping of the index. So, when the conceptual grouping is the statistical grouping, it is definitely a statistical advantage.[13]

The score for each cluster is calculated as the simple average of its indicators in the ACI. It is then calculated as the harmonic mean of its clusters as

$$ACI = \frac{1}{\frac{1}{4}\sum_{j=1}^{j=4}\frac{1}{CL_j}}$$

[9] "Bertelsmann Stifung Transformation Index BTI 2016: Methodology". BTI, 2016.
[10] Stern, Scott, Amy Wares, and Tamar Hellman. "Social Progress Index 2016 Methodological Report". The Social Progress Imperative, 2016.
[11] "Measuring the Information Society Report 2016". ITU, 2016.
[12] "African Capacity Report 2017: Building Capacity in Science, Technology and Innovation for Africa's Transformation". African Capacity Building Foundation Knowledge and Learning Department, 2016.
[13] Note that there are a number of different techniques for statistical groupings of data through the field of multivariate analysis in statistics.

where CL_j is the score for cluster j out of the four clusters.[14] In other words, it is the reciprocal of the simple average of the reciprocal of the cluster scores. The African Capacity Report 2017 states the reasoning for that aggregation method as

… capacity development is an indivisible whole of its dimensions. As such, none of the capacity development factors as given by the four clusters should be neglected. Weakness in one of the four components should be easily captured by the harmonic mean formula, which is sensitive to small values.[15]

E. Strictly Equal Weights Across All Indicators

SUII (Summary Innovation Index)

The EIS publishes the SUII through 25 indicators of EU countries in addition to China, Japan, United States, and South Korea. The SUII is calculated on a range from 0 to 1. Data is presented in a hierarchy of 8 innovation dimensions, then 3 types of indicators. However, the calculation is made across that hierarchy with the SUII being the simple average of the 25 indicators.[16]

CPI (Corruption Perception Index)

The CPI also calculates its perception of corruption in 176 countries by taking simple average of its indicators. The CPI indicators are different in that they are a collection of 13 questions about corruption from international and regional surveys.[17]

F. Selection of HDR Indices

HDI (Human Development Index)

The three HDR indices are the only indices considered that use the geometric mean in their computation. The HDI index reports on 188 countries using only four indicators. The four indicators are grouped into the Life Expectancy Index (I_H), Education Index (I_E), and the Standard of Living Index (I_L). Life Expectancy and the Standard of Living have only one indicator each, life expectancy and GNI per capita at USD PPP 2011 rates respectively. While the Education Index is made up of the simple average of expected years of schooling and mean years of schooling.[18] The HDI is then the geometric mean of the three indices calculated as: [19]

$$HDI = (I_H \, I_E \, I_L)^{1/3}.$$

IHDI (Inequality-adjusted Human Development Index)

---

[14] Ibid.
[15] Ibid.
[16] Hollanders, Hugo, Nordine Es-Sadki, and Minna Kanerva. "European Innovation Scoreboard 2016 Methodology Report". Mastricht University (Maastricht Economic and Social Research Institute on Innovation and Technology – MERIT), 2016.
[17] "Corruption Perception Index 2016: Technical Methodology Note". Transparency International, 2016.
[18] There is treatment of the data prior to aggregation in the HDR and other indices that will be discussed later in the paper.
[19] "Human Development Report 2016: Human Development for Everyone: Technical Notes". UNDP, 2016.

The IHDI accounts for inequality in the same three indices of the HDI, with considering only the mean years of schooling in the Education Index. The inequality in each index is accounted for by using the Atkinson measures, which are a family of measures used to measure inequality. The Atkinson measure is a value of 0 to 1, with 0 representing equality and 1 representing inequality. Each index is then adjusted by $I_H^* = (1 - A_H)I_H$, $I_E^* = (1 - A_E)I_E$, and $I_L^* = (1 - A_L)I_L$. Where A_x is the Atkinson inequality measure for index I_x, and I^* is the inequality-adjusted index. The IHDI is calculated as the geometric mean like the HDI. IHDI also presents the Loss Due to Inequality (LDI) measure as

$$L = 1 - [(1 - A_H)(1 - A_E)(1 - A_L)]^{1/3}$$

and the Coefficient of Human Inequality (CHI) by

$$CHI = \frac{A_H + A_E + A_I}{3}$$

as an average of inequality across the three indices.[20]

MPI (Multidimensional Poverty Index)

The MPI measures the deprivation of education, health, and standard of living through 10 components. The components are grouped into three pillars of education, health, and living conditions. The score of each component carries a value of 0 if the household in the survey does not lack the component, or 1 if does lack it. Calculation is carried as a simple average of the components in each pillar, then an average of pillars for each household in the data surveys. A household is considered in multidimensional poverty if it scores 33.3% or more in the pillars average, which means that everyone in the household is considered to be in multidimensional poverty.[21]

The headcount ratio (H) is then calculated as the ratio of the number of people in poverty (q) and the total population of the sample (n), $H = q/n$. The intensity of poverty (A) is then calculated as a ratio of the sum of the deprivation score for each person in multidimensional poverty (c) and q as

$$A = \frac{\sum_i^q c_i}{q}$$

where c_i is the deprivation score of person i. The MPI score is calculated as $MPI = HA$.[22]

The MPI also presents the contribution of each of the pillars to poverty ($Contrib_k$), where k is either the health, education, or living conditions pillar. It uses the score of each component in the pillar in its calculation. That score is the weight of each component in the total deprivation score if the household is a deprivation of that component. Even though a simple average is taken at each level, that creates different weights for the component in each pillar according to the number of

[20] Ibid.
[21] Ibid.
[22] Ibid.

components in that pillar. However, since a simple average is taken, all the components in a single pillar have the same weight (a_k). The contributions are hence calculated as

$$Contrib_k = \frac{a_k}{n\,MPI}\sum_{i=1}^{m} q_i$$

where m is the number of components in pillar k, and q_i is the number of people in deprivation of component i in pillar k.[23]

G. Non-average Structure

CG (Compliance Gap)

CG structure is singled out from all of the reviewed indices. It is an index that presents the percentage of unlicensed software installed in a country. It covers 92 countries with a collection of 182 data points for each country. The data points in CG do not perform a similar role as indicators in the other indices. The collected is the Software Market Value (SMV), Average Software Unit Price (ASP), Number of PCs Getting Software (NPS), and Software Units per PC (SUP).

The Total Software Units Installed (TSU) is calculated as $TSU = NPS \times SUP$, and the Legitimate Software Units Installed (LSU) is calculated as $LSU = SMV/ASP$. These values are then used to calculate the Unlicensed Software Units Installed (USU) as $USU = TSU - LSU$. CG publishes two outputs using this data, the Rate of Unlicensed Software Installation (RUS), and the Commercial Value of Unlicensed Software (CVU). They are calculated through $RUS = USU/TSU$, and $CVU = USU \times ASP$.[24] So, the CG does not break an abstract down into quantifiable indicators to summarize it into a number. It tries to estimate an unmeasured aspect by collecting the data about what is measured readily.

III. Data Preparation and Manipulation

The data used as indicators need to be treated before the calculations are carried out at different levels to reach the overall index. Index indicators are in different units that are for the most part unrelated.

Indicator data has to be transformed into a common scale in order for it to be possible to combine them into a single index. For example, consider an index that consists of the three indicators: number of mobile phones per person, speed of Internet connection, and length of schooling. Common units for these indicators are number count, megabytes per second, and years. To combine them using any kind of average would lead to a senseless number for several reasons. One of those reasons is that megabytes per second is usually a number in the high hundreds, while the two other indicators take

[23] Ibid.
[24] "Seizing Opportunity Through License Compliance: BSA Global Software Survey May 2016". BSA The Software Alliance, 2016.

values that are much lower. If an aggregation were carried before a transformation into a common scale, then the only relevant number would be megabytes per second.

Data is thus transformed into a common scale in what will be referred to as normalization of data. The scale varies between indices from a 0-100 range to a 0-1 range, and ranges in between. The construction of that scale usually depends on the distribution of indicator values. Normalization is sensitive to outliers because if one value is a lot higher than all of the other values, that value would get the highest score on the scale, while all the other values would be much lower. In effect, the outlier value would push down all the other values on the scale. Furthermore, standardization can also be misleading if the data for an indicator is too far apart across countries, or data is leaning to one side. The data is treated for those indicators first and then normalized.

A. Skewness, Outliers, and Limits

GII detects indicators with outliers through skewness and kurtosis.[25] "Skewness is a measure of the degree of asymmetry of a distribution."[26] There are a number of different ways for calculating skewness. The GII report cites a Groenveld and Meeden paper for its skewness reference.[27] The Groenveld and Meeden method for calculating skewness is

$$skew(X) = \frac{(\mu - \nu)}{E(|X - \nu|)}$$

with μ as the mean and ν the median.[28] In the context of index methodology, X would be the sample, so the expected value operator in the denominator would be calculated as

$$E(|X - \nu|) = \frac{1}{n}\sum_{i=1}^{n}|x_i - \nu|$$

which can be regarded as the average of the absolute value difference between all of the scores of an indicator and the median of the sample with n being the number of points in the data and x_i are the data points of the indicator.

Kurtosis is another measure of variability within a sample or a distribution. It "provides a measure of outliers (i.e., the presence of "heavy tails") in a distribution,"[29] or in our consideration a sample in the form of indicator scores. There are also a number of different methods to calculate kurtosis and interpretations of its result. There is no reference to which kurtosis method the GII uses. The commonly used method is

[25] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. "The Global Innovation Index 2016: Winning with Global Innovation". Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.
[26] Weisstein, Eric W. "Skewness". MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/Skewness.html, accessed on June 15, 2017.
[27] Groeneveld, R. A. and G. Meeden. "Measuring Skewness and Kurtosis". The Statistician 33: 391–99, 1984.
[28] Wikipedia contributors. "Skewness". Wikipedia, The Free Encyclopedia, 5 Jun. 2017, accessed on 15 Jun. 2017.
[29] Wikipedia contributors. "Kurtosis". Wikipedia, The Free Encyclopedia, 3 Feb. 2017, accessed on 15 Jun. 2017.

$$kurt(X) = \frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2}$$

with the expected value operator calculated as before.[30]

The GII considers an indicator to be in need of treatment if its skewness is greater than 2, or its kurtosis is greater than 3.5. The treatment depends if there are five outliers or more. Up to 5 outliers, the data is winsorized, which means the highest value is replaced by the next highest value. That process continues until the skewness or kurtosis is within the acceptable limits. If there are more than 5 outliers, the data is subjected to a logarithmic transformation of the form

$$\ln\left(\frac{(\max(X)-1)(x-\min(X))}{\max(X)-\min(X)}+1\right)$$

for indicators where a higher number is a desirable outcome. Alternatively, indicators where a lower number is more desirable the transformation is

$$\ln\left(\frac{(\max(X)-1)(\max(X)-x)}{\max(X)-\min(X)}+1\right)$$

where the logarithmic function brings values closer together. However, the skewness and kurtosis of one of the treated indicators with more than 5 outliers of the GII was found to increase with the logarithmic transformation. That was due to the logarithmic function decreasing with increasing magnitude from 0 to 1, so it was treated with winsorization instead.[31]

It is not clear what the criterion is for the consideration of outlier data points of an indicator. A guess would be that with an indicator that falls outside the bounds of skewness or kurtosis, the farthest points from the mean or median are considered outliers. If so, those points would be the ones to get winsorized until the data falls within acceptable bounds.

The SUII employs the use of a Chauvanet's criterion type bounds to detect outliers.[32] Chauvanet's criterion "finds a probability band, centered on the mean of a normal distribution, that should reasonably contain all n samples of a data set."[33] The SUII applies it by considering all data points of an indicator that falls beyond 2 standard deviations from the mean to be outliers. Outliers in the data

[30] Ibid.

[31] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. "The Global Innovation Index 2016: Winning with Global Innovation". Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.

[32] Hollanders, Hugo, Nordine Es-Sadki, and Minna Kanerva. "European Innovation Scoreboard 2016 Methodology Report". Mastricht University (Maastricht Economic and Social Research Institute on Innovation and Technology – MERIT), 2016.

[33] Wikipedia contributors. "Chauvenet's criterion". Wikipedia, The Free Encyclopedia, 1 Jan. 2017, accessed on 15 Jun. 2017.

are replaced with the maximum or minimum values across all countries and years that fall within the acceptable bounds.[34]

Treatment of outliers in the SUII does not deal with the question of how skewed the data is. The SUII considers indicator data to be skewed if its skewness is greater than 1. Such indicators are transformed by taking the square root of its data, which brought down the skewness of considered indicators to be within acceptable bounds.[35] It is unclear which method for calculating skewness is used in the SUII. It can be assumed that the most common definition is used which is

$$skew(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{3/2}}$$

with the variables taking the same definition as before, in addition to σ being the standard deviation and $\bar{x}$ is the average.[36] It is also unclear whether the square root transformation is carried before or after the outliers are replaced.

The IDI deals with indicators with skewness and outliers selectively. It was judged that the indicator of Internet bandwidth per Internet user has a variability that is too high. It was dealt with first by performing a logarithmic transformation by taking the logarithmic values of the indicator, and then enforcing a cutoff of two standard deviations from the mean. That implies that data points falling outside of the two standard deviations bounds were replaced by the value of the bound. Other indicators were treated by placing a cutoff value of two standard deviations from the mean without a transformation. Other indicators had caps introduced through conceptual reasoning.[37]

HDI and IHDI also had conceptual limits put for indicators. Life expectancy has a minimum of 20 years and a maximum of 85 years because these are judged to be limits life expectancy. Expected years of schooling has a limit of 0 and 18 years because societies can subsist with no formal education and 18 years are the common period for a master's degree. Mean years of schooling is limited between 0 and 15 years, with the upper limit being the projected maximum year of education by 2025. Standard of living measured in GNI per capita in USD has a limit of 100 and 75,000. The logarithm of GNI per capita is then considered to dampen effect of income on living conditions. The lower limit is set due to effects on living conditions being immeasurable for societies based on income under that limit. The higher limit is set because "there is virtually no gain in human development and well-being from income per-capita above" that limit[38, 19].

The report of the NRI does not mention dealing with outliers for indicators. But the report states on one of its graphs that outliers for the graph are considered the data points that fall outside of the 25th

[34] Hollanders, Hugo, Nordine Es-Sadki, and Minna Kanerva. "European Innovation Scoreboard 2016 Methodology Report". Mastricht University (Maastricht Economic and Social Research Institute on Innovation and Technology – MERIT), 2016.
[35] Ibid.
[36] Wikipedia contributors. "Skewness". Wikipedia, The Free Encyclopedia, 5 Jun. 2017, accessed on 15 Jun. 2017.
[37] "Measuring the Information Society Report 2016". ITU, 2016.
[38] "Human Development Report 2016: Human Development for Everyone: Technical Notes". UNDP, 2016.

to 75th percentile range.[39] The DTF manages outliers similarly with placing a cutoff of the 95th percentile or the 99th percentile depending on the distribution of the indicators. The tax rate indicator is limited to the 15% percentile of the data, bearing in mind that a lower tax rate is considered a better measurement with a higher score.[40]

GCI indicators are based on the Executive Opinion Survey (EOS). Therefore, most of the treatment of outliers is done at the survey level and goes into forming the criterion for qualification of data, which will be considered later in the paper. For indicators that are not from the EOS, outliers are treated but it is not clear how. However, skewness is treated for imports as percentage of GDP indicator by taking its logarithm; without a clarification of the criteria for it to be considered.[41]

Some indices that are based on surveys by depending on a published one or carrying their own do not need to deal with outliers because surveys usually publish results on a scale. Similarly, most indices do not detect and treat outliers for indices that are bound by their design like percentages indicators. SPI for example has capped indicators by design, except the greenhouse emissions indicator, which was capped for six countries because they skewed the data.[42] Other indices do not deal with outliers because of the nature of their calculation like the CG and MPI. Although that does not theoretically remove the possibility of skewness, having the data collected on a scale practically diminishes high variability of the data.

B. Normalization

After the data has been treated for skewness and outliers, and limits placed for conceptual purposes, data for all indicators are transformed on the same scale by normalization. Although the methods are similar, the variation arises in the choice for the limits of the normalization.

Normalization methods are critical in the formulation of indices as they enable the combination of different aspects into a single measurement of a phenomenon. The parameters used in each methodology are arguably more important than the methodology itself. Some indices choose maximum and minimum values independent or slightly dependent on the reporting year of the indicators in order to provide greater comparability.

Almost all indices use the same idea for normalization of their indices. It largely depends on the scale chosen to represent the findings. For indices with a scale from 0 to 100 like the GII, SII, HCI, and SPI, the transformation is as follows for indices where higher values are desirable,

$$\left(\frac{x - \min(X)}{\max(X) - \min(X)}\right) 100$$

[39] Baller, Silja, Soumitra Dutta and Bruno Lanvin. "Insight Report: The Global Information Technology Report 2016: Innovating in the Digital Economy". Johnson Cornell University, World Economic Forum, and INSEAD, 2016.
[40] "Doing Business 2017: Equal Opportunity for All". World Bank, 2017.
[41] Schwab, Klaus and Xavier Sala-i-Martin. "Insight Report: The Global Competitiveness Report 2016-2017". World Economic Forum, 2016.
[42] Stern, Scott, Amy Wares, and Tamar Hellman. "Social Progress Index 2016 Methodological Report". The Social Progress Imperative, 2016.

and is as follows where lower values are desirable

$$\left(\frac{\max(X) - x}{\max(X) - \min\,(X)}\right)100$$

making higher values achieve lower scores. HCI sets the maximum and minimum values in the normalization calculation through what is perceived as "logical" limits.[43]

SUII, DTF, and HDI are scaled on 0 to 1, which uses the same method as above but without multiplication by 100. The SUII uses the maximum and minimum values of the whole period the index was carried out across all countries.[44] IDI is also scaled on 0 to 1 but it normalizes differently with its indicators being a ratio to the maximum value.[45]

DTF maximum and minimum values are taken over all the data since the index started in 2005 and are set constant for 5 years. DTF total tax rate indicator is treated differently than other indicators to reduce bias for countries that are able to apply no or relatively low levels of taxes. Taking into account that a higher tax rate is considered undesirable, the indicator is normalized as

$$\left(\frac{\max(X) - \mathrm{x}}{\max(X) - \min\,(X)}\right)^{0.8}$$

where the 0.8 exponent decreases variability.[46]

NRI and GCI are scaled on 1 to 7 and are normalized identically with

$$6\left(\frac{x - \min\,(X)}{\max(X) - \min\,(X)}\right) + 1$$

for indicators where higher values are desirable, and

$$6\left(\frac{\min(X) - \mathrm{x}}{\max(X) - \min\,(X)}\right) + 7$$

for scores where lower variables are desirable.

[43] "Insight Report: The Human Capital Report 2016: Technical Notes". Mercer and World Economic Forum, 2016.
[44] Hollanders, Hugo, Nordine Es-Sadki, and Minna Kanerva. "European Innovation Scoreboard 2016 Methodology Report". Mastricht University (Maastricht Economic and Social Research Institute on Innovation and Technology – MERIT), 2016.
[45] "Measuring the Information Society Report 2016". ITU, 2016.
[46] "Doing Business 2017: Equal Opportunity for All". World Bank, 2017.

The GCI treats the inflation indicator differently as it considers a rate of inflation between 0.5% and 2.9% to be optimal with a score of 7. Outside of these bounds the score then decreases.[47] There are a number of methods that can be used to apply that effect. It is unclear which method the GCI used.

CPI uses a slightly different normalization method to scale on 0 to 100. It uses the z-score transformation calculated by

$$\frac{x-\mu}{\sigma}\,100$$

with the variables taking the same definitions as before. CPI uses the mean and standard deviation of the 2012 data for its calculations instead of changing them every year. This is done in order to facilitate comparison over time. Theoretically it is possible for this method to give a value greater than 100 if the difference between a score of a country in an indicator and the 2012 mean for that indicator is greater than one standard deviation. That is a very possible occurrence considering that outlier criteria in other indices are considered to be beyond two or three standard deviations. However, CPI caps the scores at 100.[48]

BTI indicators are from a survey carried with scores given from 1 to 10, except for the GNI per capita at USD PPP and the UN Education Index. Both of these indicators are normalized similarly to the NRI normalization method. The Management Index is scaled using the Level of Difficulty (LD) criteria. LD itself includes as one of its indicators the average of the stateness and rule of law criteria. The score for LD is normalized by

$$LD_N = \left(\frac{LD-1}{10-1}\right)0.25+1$$

to be on a scale of 1 to 1.25. The Management Index (MI) is calculated by

$$MI = \frac{10}{12.5}LD_N\,MC$$

where MC is the management criteria score of the average of its indicators.[49]

The GCI also uses its indicators in the process to normalize other indicators. The normalized scores for business impact of Tuberculosis and HIV/AIDS are calculated by first considering the ratio of incidence rate in a country to the highest incidence rate (ρ). The EOS score (1 – 7) of impact (EOS_I) is then included as the business impact of each disease as $\frac{1}{\rho}EOS_I$. That final calculation is the one

[47] Schwab, Klaus and Xavier Sala-i-Martin. "Insight Report: The Global Competitiveness Report 2016-2017". World Economic Forum, 2016.
[48] "Corruption Perception Index 2016: Technical Methodology Note". Transparency International, 2016.
[49] "Bertelsmann Stifung Transformation Index BTI 2016: Methodology". BTI, 2016.

normalized for the business impact of Tuberculosis and HIV/AIDS score. GCI also uses its indicators in the calculation of domestic market size and size of foreign market indicators.[50]

BTI also sets limits for indicators depending on its classification of countries as autocracies or democracies. There are 7 indicators where if a country falls short of a threshold on one of them it would be considered an autocracy. There are six indicators that have an upper limit for autocracies, and a lower limit for democracies.[51]

C. IHDI Atkinson Inequality Measure

As mentioned previously, the IHDI uses Atkinson inequality type measures to account for inequality in life expectancy, education, and standard of living. The use of the measures decreases the HDI scores as the measured inequality increases. The Atkinson method detects inequality using the ratio between the geometric average and the arithmetic average.[52] That is because the geometric average is less than its corresponding arithmetic average, and is affected more significantly by relatively small values. The two are equal only when all the numbers being averaged are equal.

The Atkinson measure for life expectancy (A_H) measures inequality across the different age groups. The data used for that are from abridged life tables.[53] The tables are from the UNDESA World Population Prospects database, and it provides data separated in age groups of

… a set of values showing the mortality experience of a hypothetical group of infants born at the same time and subject throughout their lifetime to the specific mortality rates of a given period.[54]

From these tables, the age of death (T_g) at each age group (g), and the proportion of survivors in each age group (w_g) can be calculated. The Atkinson measure for the Health Index is then calculated as

$$A_H = 1 - \frac{\prod_{g=1}^{n} [T_g]^{w_g}}{\sum_{g=1}^{n} w_g T_g}$$

with n being the total number of age groups in the table, and the weights used are specified to add up to 1.[55]

The Atkinson measure for the Education Index (A_E) is calculated similarly as

[50] Schwab, Klaus and Xavier Sala-i-Martin. “Insight Report: The Global Competitiveness Report 2016-2017”. World Economic Forum, 2016.
[51] “Bertelsmann Stifung Transformation Index BTI 2016: Methodology”. BTI, 2016.
[52] “Human Development Report 2016: Human Development for Everyone: Technical Notes”. UNDP, 2016.
[53] Ibid.
[54] “World Population Prospects - Population Division”. UNDESA, 2015. https://esa.un.org/unpd/wpp/Download/Standard/Mortality, accessed on June 16, 2017.
[55] Kovacevic, M. “Measurement of Inequality in Human Development—A Review”. Human Development Research Paper, UNDP-HDRO, 2010.

$$A_E = 1 - \frac{\prod_{i=0}^{n}[x_i + 1]^{w_i}}{1 + \sum_{i=0}^{n} w_i x_i}$$

with w_i being the proportion of the population that had x_i years of schooling. The addition of one in the denominator and numerator is included to deal with the mathematical problem of a possibility of zero in the denominator.[56]

The Atkinson measure for standard of living (A_L) is calculated similarly. The data used for its calculation is household disposable income or consumption per capita, depending on the survey used for a country. The highest and lowest 0.5 percentiles are truncated from the data, and the actual USD numbers are used in contrast to their logarithm in the calculation of the sub-index itself. This is done so the measure would detect the full extent of inequality. In some cases, that required data is not available, so the asset index matching methodology was used.[57] The methodology is an "approach to simulate household income based on an asset index and publicly available macroeconomic data."[58]

IV. Data Considerations

The availability of data is one of the major challenges for index construction. Global indices try to measure international phenomena, while the data for their indicators are collected on a very small local scale of individuals and households. That causes inconsistency in the data that is collected in each country, especially missing data for different indicators for different countries. Although indices that use data from expert surveys face a weaker challenge with data consistency, they still need to consider a method for capturing it.

There are two main considerations when handling the inconsistency of missing data. The first consideration is cases where the lack of data disqualifies a country from being included in the index. The second is dealing with missing values of countries that will be included.

A. Qualification Criteria

The criteria for countries to be included usually consider the availability of data for countries. The criteria for data usually depend on the sources and how recent it was published. Sources are required to be reputable, transparent, and periodic. Although indices publish their reports with a yearly reference, GII 2016 for example, the data used is often of previous years, and the time of data is not necessarily the same for all indicators and countries. Table 2 lists the available criteria for including a country that reviewed indices published in their reports. Table 3 shows the earliest year of data used.

Table 2: Criteria for including a country

	Index	Criteria

[56] Ibid.
[57] "Human Development Report 2016: Human Development for Everyone: Technical Notes". UNDP, 2016.
[58] Harttgen, K., and S. Vollmer. "Using an Asset Index to Simulate Household Income". Economic Letters 121(2): 257–262, 2013.

1	GII	• Minimum of 60% available indicators in each sub-index. • Two sub-pillar scores available for each pillar.
2	NRI	• Less than 5 indicators missing, which means more than 90% of the indicators available. • EOS has to be carried out in the country. • One example of a sub-pillar showing an n/a for Affordability in Argentina, it had 2 out of the 3 indicators missing.
3	CPI	• At least 3 indicators available.
4	BTI	Conceptual criteria: • Countries that are yet to achieve full democracy and market economy (all OECD countries by 1989 are excluded. • Countries with more than 2 million population (with the exception of seven countries).
5	HCI	• For a country to be included it has to have minimum of 65% of indicators within each pillar. • For an indicator to be included, it has to be available for at least 50% of the countries.
6	SPI	• Countries with one or more indicators missing, in three or more components are excluded. • For an indicator to be included, it has to be available for at least 95% of the countries.
7	HDI	• Missing a maximum of one indicator.
8	MPI	• All indicators must be from a single survey.

SUII and SII do not face the challenge of missing data, since their region of interest is Europe, which has established data collection institutions. Despite its large coverage, IDI indicators are chosen with availability of data collected by ITU taken into account, so there are no criteria for inclusion of a country.[59] DTF and ACI do not have criteria for qualification by countries as they carry out their own survey. However, the DTF carries its surveys in the largest cities of countries, but enhances that for large countries with surveys carried in the two largest cities.[60] BTI carries its own surveys as well, but it states criteria for inclusion on the type of countries to be included and not the availability of the data.[61] The SPI and HCI include qualification criteria on countries and indicators. It is unclear which criterion is applied first.

[59] "Measuring the Information Society Report 2016". ITU, 2016.
[60] "Doing Business 2017: Equal Opportunity for All". World Bank, 2017.
[61] "Bertelsmann Stifung Transformation Index BTI 2016: Methodology". BTI, 2016.

Table 3: Earliest data used by indices

	Index	Earliest Year Used
1	GII 2016	2006
2	SUII 2016	2006
3	SII 2016	2013
4	NRI 2016	2013
5	IDI 2016	2014
6	CG 2016	2016
7	DTF 2017	2015
8	GCI 2015-2016	2008
9	CPI 2016	2014
10	BTI 2016[62]	2013
11	ACI 2017	2016
12	HCI 2016[63]	2013
13	SPI 2016	2011
14	HDI 2016[64]	2011
15	IHDI 2016	2004
16	MPI 2016[64]	2005

The years stated in Table 3 should be interpreted as the latest estimate of the oldest year of data used. This is because indices mention the publication year of the source where the data is taken from, however that source is likely to have used data from earlier years. Note also that the years mentioned are not representative of the data used, most of the data used by indices is from one to three years prior to publication.

B. Executive Opinion Survey Criteria and Imputation

The EOS is one of the main surveys used for indices of different fields. The EOS questions are answered on a scale of 1 to 7, which dictates the scales of some of the indices that use it. One of the indices that are based on the EOS is the GCI. The EOS criteria for filtering data are included in the GCI report. The EOS filtering system tests for the viability of the surveys, the relation of a survey to the rest of the surveys within a country, and the variability of the latest round of surveys in relation to previous rounds.

The EOS filtering criteria is different in nature than that of the indices because it is at the survey level. EOS starts by excluding surveys with 80% answers with the same scores because it demonstrates lack

[62] Published every two years.
[63] All information about the HCI is from the Technical Notes, except the earliest year of data used is from the User's Guide.
[64] All information about the HDI and MPI is from the Technical Notes, except the earliest year of data used is from the Human Development Report.

of focus while answering. All surveys that are missing more than 50% of their answers are also excluded.[65]

EOS then performs a type of multivariate statistical test called a Mahalanobis distant method test. The test is described in the GCI report as

This test estimates the probability that an individual survey in a specific country "belongs" to the sample of that country by comparing the pattern of answers of that survey against the average pattern of answers in the country sample.[66]

EOS carries the tests with the answers of 52 core questions of the survey. The limitation on that test is that the number of surveys in a country being tested has to be greater than the number of questions in the survey. If the probability that a survey does not belong to the group of surveys of a country is greater than 99.9%, then the survey is discarded.[67]

EOS follows to tests the questions of the qualified surveys. If the z-score of a question is greater than 3 compared to the same question in the same country, the answer for that question is discarded as well.[68] EOS also places lower and upper limits on the average score of 66 questions to test for the variability across time. The limits are

$$L = Q1 - 1.5\, IQR$$

and

$$U = Q3 + 1.5\, IQR$$

where L is the lower limit, U is the upper limit, $Q1$ is the 25th percentile, $Q3$ is the 75th percentile, and IQR is the inter quartile range defined as $IQR = Q3 - Q1$. These values are calculated for the average of the 66 questions for a country in a given year compared with the average of these questions for previous years. Countries that are outside of these bounds are considered outliers and further investigated, which can lead to the removal of a country from a given year results.[69]

The discarded and missing values for countries that qualify to be included are then, imputed by replacing them with the scores of the previous year.[70]

---

[65] Schwab, Klaus and Xavier Sala-i-Martin. "Insight Report: The Global Competitiveness Report 2016-2017". World Economic Forum, 2016.
[66] Ibid.
[67] Ibid.
[68] Note that it is the same z-score that is used to normalize indicators in the CPI.
[69] Schwab, Klaus and Xavier Sala-i-Martin. "Insight Report: The Global Competitiveness Report 2016-2017". World Economic Forum, 2016.
[70] Ibid.

C. Imputation

Imputing data is the method that each index uses to deal with the missing data in its indicators for those countries that qualify to be included. The general assumption made about missing data is that it is missing completely at random (MCAR). This assumes that "missing values do not depend on the variable of interest or on any other observed variable in the data set."[71]

GII, NRI, and GCI, do not replace data. Similarly, HCI does not replace missing indicators, however it replaces missing values for child labor and literacy rate by 0.5% and 95% respectively for developed countries. This is because some developed countries stopped publishing data about child labor and literacy rate.[72]

SUII replaces missing data with the data of the latest year available; if no data is available the indicator is not considered.[73]

IDI first attempts to replace missing data by calculating a growth rate for the missing indicator of similar countries. If the previous year data of the indicator is also missing, IDI uses the hot-deck method for imputing data.[74] The hot-deck method uses a collection of similar indicators and countries to estimate the missing indicator with the most similar case.[75] The similarity of indicators is based on what indicators are usually highly correlated. Country similarity is judged on geography, income, and other classifications depending on the indicator to be estimated.[76]

CPI replaces missing data for the purposes of calculating the mean and standard deviation used in the normalization procedure. The replaced data is not included in the calculation of the country scores. CPI uses regression estimate the missing values. CPI indicators are 13 similar indicators about corruption from different indices. It regresses each indicator against all other indicators that include at least 50% of the countries to estimate the missing values.[77]

SPI also uses regression to estimate missing indicators. The regression is done on the indicators of the component that includes the missing indicator. However, some indicators are missing from groups of countries that are judged to be missing not at random.[78] An example of these indicators is tolerance of homosexuality in Middle Eastern countries. In such cases, qualitative estimates are applied. They

---

[71] "Handbook on Constructing Composite Indicators: Methodology and User Guide". JRC European Commission and OECD, 2008.
[72] "Insight Report: The Human Capital Report 2016: Technical Notes". Mercer and World Economic Forum, 2016.
[73] Hollanders, Hugo, Nordine Es-Sadki, and Minna Kanerva. "European Innovation Scoreboard 2016 Methodology Report". Mastricht University (Maastricht Economic and Social Research Institute on Innovation and Technology – MERIT), 2016.
[74] "Measuring the Information Society Report 2016". ITU, 2016.
[75] Andridge, Rebecca R. "A Review of Hot Deck Imputation for Survey Non-response". International Statistical Review, U.S. National Library of Medicine, Apr. 2010. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3130338, accessed on 16 Jun. 2017.
[76] "Measuring the Information Society Report 2016". ITU, 2016.
[77] "Corruption Perception Index 2016: Technical Methodology Note". Transparency International, 2016.
[78] This is equivalent to not assuming MACR.

are also made for incidents where regression is judged to give irrational values.[79] HDI estimates missing data using cross-country regression, where the regression is carries on similar countries.[80]

The effect of replacing or not replacing missing indicators has a notable effect on the calculation of indices. Calculating the data with omission of missing data has a direct effect on the weights of averaging, even if simple averages are taken. However, with the replacement of data there is always the possibility that the estimation is significantly different than it would have been if the indicator value were available. Furthermore, the MACR assumption should be regarded with caution as the possible reasons for the unavailability of data are ignored.

V. Weights

The manner of assigning weights to indicators and different levels of aggregation in the calculation of indices is arguably one of the most dominating issues in index design. This is because it is the most direct way of affecting the scores of a composite index.

The main issue regarding weights is the issue of compensability. It is concerned with the notion that indicators, which are meant to measure different aspects, build up together so they compensate one another. The question arises as to what extent does a deficiency in one indicator can actually be compensated by another indicator considering the phenomenon an index is attempting to reflect. This issue arises in arithmetic aggregation methods like weighted and simple averaging, which are by far the most common methods due to their simplicity. Compensability introduces the importance of weights. Aggregation using the geometric mean does not have that issue as low scores scale all other scores lower without compensation, and in turn show greater improvement with increases in lower scores than arithmetic averaging.[81]

Beyond the conceptual consideration of importance in assigning weights, there is also the statistical one. Weights should deal with the possibility of double counting between indicators, as it is possible to have the same aspect measured by different indicators. However, that has to strike a balance with the indicators being related enough to belong to the phenomenon that is claimed to be measured. There is also the idea of statistical importance, which is different than conceptual importance in that it considers the indicators responsible for the greatest variation in the index as the most crucial.[82]

The allocation of different weights to indicators is done implicitly through taking simple averages at different levels of aggregation or explicitly by aggregating using a weighted average. If a sample is

[79] Stern, Scott, Amy Wares, and Tamar Hellman. "Social Progress Index 2016 Methodological Report". The Social Progress Imperative, 2016.
[80] "What Is an "imputed" Indicator – and for Which Countries Were These Imputed Statistics Used?". Human Development Reports, HDRO FAQs. http://hdr.undp.org/en/content/what-%E2%80%9Cimputed%E2%80%9D-indicator-%E2%80%93-and-which-countries-were-these-imputed-statistics-used, accessed on 16 Jun. 2017.
[81] "Handbook on Constructing Composite Indicators: Methodology and User Guide". JRC European Commission and OECD, 2008.
[82] "Handbook on Constructing Composite Indicators: Methodology and User Guide". JRC European Commission and OECD, 2008.

taken within each pillar and then the same is applied to the pillars to calculate the overall index, different weights are assigned implicitly.

Indicators that are grouped in pillars with less number of indicators will have more weights than indicators that are grouped in pillars with more number of indicators. Which seems to be an arbitrary assigning of weights from a conceptual point of view. If the number of indicators in each pillar is accounted for to have equal weights for all indicators in the index, then the pillars with more number of indicators will have greater weights than pillars with less number of indicators, which also seems arbitrary. In order to have equal weights for indicators and pillars in their contribution to the overall index, either simple average is taken across all indicators at once; or each level of aggregation has to have the same number of components. The first decreases the conceptual design of the index, while the latter also would put an arbitrary limit on index design. In general, different weights are assigned to indicators whether it is explicitly done or not.

A. Assignment of Weights

This section discusses indices that assign different weights explicitly. GII assigns a weight of 1 or 0.5 to indicators within their sub-pillars, and to sub-pillars within their pillars. The weights are based on an analysis of the Pearson correlation ratio.[83]

The Pearson correlation ratio gives a measure of the importance of a component based on the reduction to the variance of a grouping if that component was fixed.[84] The variance due to a single component on a grouping is directly related to how correlated it is to other components in the grouping. When components of a grouping are highly correlated, a variance in one of them causes slight variance in the grouping as the highly correlated group varies together. However, that behavior causes high correlation between each component of the highly correlated ones and the group. This gives a false appearance of high correlation between each of the highly correlated components and the group. The Pearson correlation ratio detects that behavior and assigns lower weights to highly correlated components to reflect their actual importance in the calculation.[85]

The Pearson correlation ratio decreases as correlation between components increases. As there is always some correlation between components of a group of indices, the weights are assigned to be the ratio of the Pearson correlation ratio to the sum of the Pearson correlation ratios of the group. The GII uses 1 and 0.5 for simplicity as that method most likely give a different weight to each component be it indicator in a sub-pillar or sub-pillar in a pillar. The threshold that transforms Pearson correlation ratio weights to the values of 1 and 0.5 is not mentioned in the GII report.

This analysis assigned 0.5 weights to 36 indicators and two sub-pillars. The two sub-pillars are creative goods and services, and creation of online content. The indicators that were assigned 0.5

[83] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. "The Global Innovation Index 2016: Winning with Global Innovation". Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.
[84] Grouping in this GII context is the sub-pillar or pillar with the components being the indicator or the sub-pillar respectively.
[85] Paruolo, P., M. Saisana, and A. Saltelli. "Ratings and Rankings: Voodoo or Science?" Journal of the Royal Statistical Society A 176 (3): 609–34, 2013.

weights are sometimes obvious ones like regulatory quality and rule of law in the regulatory environment sub-pillar.[86] It is possible for half weights to be assigned to less obvious indicators, typically shown as assigning one indicator with half weight within a sub-pillar. This implies that it is highly correlated to the group, but not to individual indicators in it. As evident, the Pearson correlation ratio method also employs the weighting to ensure a degree of distinction between averaged components of a group.

The NRI also assigns 0.5 weights to some of its indicators when averaging within a pillar.[87] Although the report does not specify the method of that allocation, it does report the indicators in pairs; showing which indicators within the pillar were judged to be highly correlated. That could have been done using the Pearson correlation ratio method, or another statistical method, or even qualitatively. Examples of the half-weighted indicators in the NRI in pairs are:[88]

- Efficiency of legal systems in settling disputes, and efficiency of legal system in challenging regulations.
- Numbers of procedures to enforce a contract, and number of days to enforce a contract.
- Number of days to start a business, and number of procedures to start a business.
- ICT use for business-to business transactions, and business-to-consumer Internet use.

Although IDI assigns weights to its sub-indices of 40% to ICT Access, 40% to ICT Use, and 20% to ICT Skills. ICT Skills was given a lower weight because it measured through schooling, secondary gross enrolment ratio, and tertiary gross enrolment ratio. The lower weight was given because these indicators were judged to be proxy indicators to ICT Skills.[89]

SII assigns weights to categories according to “consultations with analysts and experts.”[90] The indicators within categories are aggregated as a simple average. The weights assigned to categories are:

- Policy and Institutional Framework 44.4%
- Financing 22.2%
- Entrepreneurship 15%
- Society 18.3%

DTF aggregates mostly in simple averages, except for the getting credit pillar. The normalization is done after summing the two indicators, strength of legal rights index and depth of credit information index. The indicators are on a scale of 0-12 and 0-8 respectively. For countries where the DTF carries out surveys in two cities, each city is given a weight according to its population.[91]

[86] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. “The Global Innovation Index 2016: Winning with Global Innovation”. Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.
[87] Baller, Silja, Soumitra Dutta and Bruno Lanvin. “Insight Report: The Global Information Technology Report 2016: Innovating in the Digital Economy”. Johnson Cornell University, World Economic Forum, and INSEAD, 2016.
[88] Ibid.
[89] “Measuring the Information Society Report 2016”. ITU, 2016.
[90] “Old problems new solutions: Measuring the capacity for social innovation across the world: Social Innovation Index 2016”. The Economist Intelligence Unit, 2016.
[91] “Doing Business 2017: Equal Opportunity for All”. World Bank, 2017.

GCI assigns different weights to pillars according to the stage of development an economy is in. The stage of development is determined by the GDP per capita and the percentage of exports of mineral goods to total exports. Countries that have a percentage higher than 70% are considered to have lower levels of economic development. If the percentage is lower, exports of mineral goods are not considered. Countries with 100% are assigned to stage 1. But countries with higher income for 5 years than the top 10 countries for patent cooperation treaty and patent application per capita indicators are allocated to stage 3. There are a number of possible methods to take percentage of mineral goods into account; it is unclear which one GCI uses. It is planned that this method of assigning weights could be disregarded as economic development has shown not be restricted to stages with the growth of technology.[92]

	STAGE OF DEVELOPMENT				
	Stage 1: Factor-driven	Transition from stage 1 to stage 2	Stage 2: Efficiency-driven	Transition from stage 2 to stage 3	Stage 3: Innovation-driven
GDP per capita (US$) thresholds*	<2,000	2,000–2,999	3,000–8,999	9,000–17,000	>17,000
Weight for basic requirements	60%	40–60%	40%	20–40%	20%
Weight for efficiency enhancers	35%	35–50%	50%	50%	50%
Weight for innovation and sophistication factors	5%	5–10%	10%	10–30%	30%

Figure 2: Pillar weights of GCI. Source: The Global Competitiveness Report 2016-2017, page 38.

At the level of indicators moving upward the GCI uses simple averages with a few exceptions. As the NRI, number of days to start a business, and number of procedures to start a business are given half weights. At the indicator level, four indicators are given half weight within their groupings because they appear in two different aggregations.[93]

- Intellectual property protection indicator appears in:
  - Property rights category, of public institutions sub-pillar, of the institutions pillar, of the Basic Requirements sub-index; and
  - R&D pillar, of Innovation and Sophistication Factors sub-index.
- Mobile telephone subscriptions, and fixed telephone lines indicators appear in:
  - Electricity and telephony infrastructure sub-pillar, of infrastructure pillar, of Basic Requirements sub-index; and
  - ICT use sub-pillar, of technology readiness pillar, of Efficiency Enhancers sub-index.
- Reliance on professional management indicator appears in:
  - Business sophistication pillar, of Innovation and Business Sophistication sub-index; and
  - R&D pillar, of Innovation and Business Sophistication sub-index.

At the category level, simple averages are taken except for the domestic competition and foreign competition categories within the competition sub-pillar. Weights for these categories vary according to country equal to the portion of domestic (D) and foreign (F) competitions to total competition calculated through proxy variables as follows:

[92] Schwab, Klaus and Xavier Sala-i-Martin. "Insight Report: The Global Competitiveness Report 2016-2017". World Economic Forum, 2016.
[93] Ibid.

$$D = \frac{C + I + G + X}{C + I + G + X + M}$$

and

$$F = \frac{M}{C + I + G + X + M}$$

where C is consumption, I is investment, G is government spending, X is exports, and M imports. With their sum regarded as the total competition.[94]

At the sub-pillar level, the exceptions to simple averaging are competition and quality of demand conditions with 67% and 33%, public institutions and private institutions with 75% and 25%, and domestic market size and foreign market size with also 75% and 25%. There is no clear rationale for these weights.[95]

GCI indicators that are from EOS use a two-year weighted average of survey answers for each country. That is to dampen the effect of large variations in answers to survey questions, and is calculated as to prevent answers with larger sample sizes from dominating the score. The calculation is carried as

$$q^{2015-2016} = \frac{1}{2}(0.4\, q^{2015} + 0.6\, q^{2016}) + \frac{1}{2}\left(\frac{q^{2015}\, N^{2015}}{N^{2015} + N^{2016}} + \frac{q^{2016}\, N^{2016}}{N^{2015} + N^{2016}}\right)$$

where q is the question score for a given year and N is the number of respondents for question q.[96]

HCI carries out a simple average on the indicator level, then a weighted average at the age pillar level. The weight assigned to each age pillar is the global portion of those within the age bracket to the total population.[97]

SPI uses Principal Component Analysis (PCA) to assign weights to indicators within each component. Then uses simple averages at the component and dimension levels to calculate the overall index.[98]

PCA is type of multivariate analysis that produces results similar to the Pearson correlation ratio method. It separates components into groups of linear combination of its indicators.[99] The groups are ranked in order of maximum accounting for variation among the indicators, while the groups themselves being uncorrelated. The coefficients of the linear combination are depended on the variation each indicator is responsible for within each group, which is used to assign weights. PCA has

---

[94] Schwab, Klaus and Xavier Sala-i-Martin. "Insight Report: The Global Competitiveness Report 2016-2017". World Economic Forum, 2016.
[95] Ibid.
[96] Ibid.
[97] "Insight Report: The Human Capital Report 2016: Technical Notes". Mercer and World Economic Forum, 2016.
[98] Stern, Scott, Amy Wares, and Tamar Hellman. "Social Progress Index 2016 Methodological Report". The Social Progress Imperative, 2016.
[99] The common term used for PCA groups is components; it is not used here for clarity.

the added benefit of separating the indicators statistically on top of accounting for variation.[100] Prior to carrying out PCA though, SPI carries out a Kaiser-Meyer-Olkin test (KMO).[101] KMO measures how suited the data is for PCA.[102]

In the context of SPI, the first PCA group for each component ought to have much larger accounting for variation than the lower ranked groups; and each of the indicators in that group should contribute notably to variation within the group. That was the case except for few exceptions.

The health and wellness component was shown to contain two distinct statistical groups. The two groups are the life expectancy at 60 and premature deaths from non-communicable diseases indicators group, and obesity rate and suicide rate indicators group. That was treated by carrying a simple average of each of the two indicators separately, and then averaging the scores. If the PCA weights were directly used, the first two indicators would have had lower weights than the latter two. Another exception is that the weight given for freedom of religion in the personal freedom and choice component was just 0.05. This implies that it hardly has any role to play in its component, but the indicator was still kept because of conceptual significance.[103]

VI. Index Assessment

GII 2016 and IDI 2015 published audits carried by the Joint Research Centre of the European Commission (JRC). The audits examine coherence in terms of their structure, robustness in terms of their calculation methods, and statistical added value in terms of their presentation of new information. JRC assesses the characteristic of the indices to strike a balance between their structure being related, and their aggregation offering reliable distinct information. Other indices use similar methods in their modeling, however the GII and the IDI are the only ones that use these methods to assess the indices explicitly. Assessment is carried on data after treatment of outliers, imputation consideration, and normalization.

A. Assignment of Variances

The structure of the GII was tested by carrying out PCA on the sub-pillars of each pillar. The aim of the test is to see if the PCA of the sub-pillars in each pillar produces a component that accounts for a large variance of the sub-pillars, as this would imply that grouping the sub-pillars in one pillar is justifiable. The PCA test produced a component for each of the sub-pillars groupings that accounts for 60% to 84% of the variance. PCA is also carried for the five pillars of the Innovation Input sub-

---

[100] "Handbook on Constructing Composite Indicators: Methodology and User Guide". JRC European Commission and OECD, 2008.
[101] Stern, Scott, Amy Wares, and Tamar Hellman. "Social Progress Index 2016 Methodological Report". The Social Progress Imperative, 2016.
[102] "Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy." Statistics How To. http://www.statisticshowto.com/kaiser-meyer-olkin, accessed 30 Jun. 2017.
[103] Stern, Scott, Amy Wares, and Tamar Hellman. "Social Progress Index 2016 Methodological Report". The Social Progress Imperative, 2016.

index, where a single component was also shown to account for a large percent of the variance of the pillars with 76%. PCA also showed that each of the five pillars contribute similarly to the variance in that single component.[104]

JRC also uses PCA to test the structure of the ICT access and ICT use sub-indices of the IDI. PCA is carried on the indicators of each sub-index, with the principal component accounting for 78% and 86% of the variance for ICT access and ICT use respectively. PCA is also carried out at the sub-index level where the main grouping accounted for 92% of the variance in the sub-indices.[105]

JRC tests the weighting scheme for the IDI using the Pearson correlation ratio.[106] The test is carried at the indicator level for each sub-index, and at the sub-index level for the overall IDI. Especially of interest was the allocation of 40% weights to ICT use and ICT access sub-indices, and 20% for ICT skills sub-index. This was validated by ICT use and ICT access having a higher identical Pearson correlation ratio of 0.96, and ICT having a lower one of 0.83. At the sub-index level, equal weighting was justified by indicators having similar Pearson correlation ratios within their sub-index. The exception being for three indicators in ICT access, percentage of households with a computer and percentage of households with internet access having a high value of 0.93, and mobile cellular-telephone subscription per 100 inhabitants having a low value 0.57.[107]

The GII Innovation Input sub-index is tested also using Cronbach alpha.[108] Cronbach alpha (c-alpha) tests the reliability of aggregating a group of data by measuring to what extent they measure a single phenomenon.[109] This can be done at any level of aggregation, from indicators to sub-indices. C-alpha is calculated for the Innovation Input sub-index as

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} var(x_j)}{var(x_0)}\right)$$

where k is the number of pillars, $var(x_j)$ is the variance of a pillar score across countries, and $var(x_0)$ is the variance of the sum of pillar scores across countries.[110, 111] The c-alpha test for the Innovation Input sub-index was found to be 0.95, “well above the 0.70 threshold for a reliable aggregate.”[112]

---

[104] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. “The Global Innovation Index 2016: Winning with Global Innovation”. Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.
[105] “Measuring the Information Society Report 2015”. ITU, 2015.
[106] Refer to the Assignment of Weights section for a description of the Pearson correlation ratio (not to be confused with the Pearson correlation coefficient).
[107] Ibid.
[108] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. “The Global Innovation Index 2016: Winning with Global Innovation”. Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.
[109] “Handbook on Constructing Composite Indicators: Methodology and User Guide”. JRC European Commission and OECD, 2008.
[110] It is not clear whether the variance is taken of the pillar or the indicator scores, it is assumed here that it is of the pillar scores, but considering the indicator scores would also be a reliability test.
[111] Zaiontz, Charles. “Cronbach’s Alpha.” Real Statistics Using Excel, 2014. http://www.real-statistics.com/reliability/cronbachs-alpha, accessed on 29 Jun. 2017.
[112] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. “The Global Innovation Index 2016: Winning with Global Innovation”. Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.

JRC used c-alpha to test for the reliability of each of the IDI sub-indices at the indicator level. C-alpha values implied reliability with a ranged from 0.86 to 0.91. In addition, C-alpha trials where carried for each of the sub-indices while removing one of their indicators. This revealed an important role for the secondary gross enrolment ratio in ICT skills, where the c-alpha falls from 0.86 to 0.71 without it. The same reliability test was also carried at the sub-index level that showed a c-alpha of 0.95 for the overall IDI.[113]

B. Correlation Assessment

Another test for the GII is calculating the Pearson correlation coefficients between the sub-pillars and pillars, with the aim of seeing if the sub-pillars have a stronger connection to their own pillars than other ones. GII correlation test validates that each pillar has a strongest correlation to its pillar. The Innovation Output sub-index is also tested using correlation coefficients. The two pillars show that they are highly correlated to each other, and each pillar highly correlated to the Output sub-index. Furthermore, correlation coefficients are used to test the relation of the two sub-indices to each other, and to the GII; with 0.97 correlations for each with the GII, and 0.88 to each other.[114]

JRC uses correlation to make data checks on the IDI. Correlation is used to justify the use of the logarithmic transformation for the Internet bandwidth indicator, where the increase of correlation with other indicators after the transformation is taken to be a validation. Correlation is also used to test for the allocation of indicators to sub-indices with all the indicators showing highest correlation with their sub-index. "This outcome suggests that the indicators have been allocated to the most relevant ICT dimension."[115]

C. Statistical Added Value

The JRC carries a statistical added value test for the GII with a summary of its results shown in Figure 3. The test measures the difference for each country between the GII rank and the rank of each pillar. The distribution shown in Figure 3 is considered to show "the added value of the GII ranking, which helps to highlight other aspects of innovation that do not emerge directly by looking into the seven pillars separately."[116] The high correlation between the sub-pillars, pillars, sub-indices, and GII coupled with the considerable difference in ranks support this added value.[117]

[113] "Measuring the Information Society Report 2015". ITU, 2015.
[114] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. "The Global Innovation Index 2016: Winning with Global Innovation". Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.
[115] "Measuring the Information Society Report 2015". ITU, 2015.
[116] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. "The Global Innovation Index 2016: Winning with Global Innovation". Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.
[117] Ibid.

	Innovation Input Sub-Index					Innovation Output Sub-Index	
Rank differences (positions)	Institutions (%)	Human capital and research (%)	Infrastructure (%)	Market sophistication (%)	Business sophistication (%)	Knowledge and technology outputs (%)	Creative outputs (%)
More than 30	12.5	10.2	7.8	21.1	21.9	10.9	4.7
20–29	16.4	14.8	12.5	16.4	10.2	10.2	11.7
10–19	21.9	23.4	35.9	25.0	21.9	30.5	15.6
10 or more*	**50.8**	**48.4**	**56.3**	**62.5**	**53.9**	**51.6**	**32.0**
5–9	28.1	22.7	16.4	16.4	23.4	19.5	32.0
Less than 5	18.0	25.8	24.2	20.3	20.3	21.9	32.8
Same rank	3.1	3.1	3.1	0.8	2.3	7.0	3.1
Total†	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**
Pearson correlation coefficient with the GII	0.88	0.90	0.89	0.81	0.86	0.92	0.93

Source: Saisana, Domínguez-Torreiro, and Vertesy, European Commission Joint Research Centre, 2016.
* This column is the sum of the prior three rows.
† This column is the sum of all white rows.

Figure 3: Distribution of differences between pillars and GII rankings. Source: The Global Innovation Index 2016: Winning with Global Innovation, page 64.

JRC carries a similar statistical value-added test for the IDI, with the results shown in Figure 3. The numbers above the diagonal are the Spearman rank correlation coefficient. The numbers under the diagonal are the percentage of countries whose ranking is different by more than 10 positions between the column and row in comparison. The numbers shown are interpreted as the IDI having a high enough correlation within its structure to be coherent, with high enough difference in rankings between its components to present an added value while not being redundant.[118]

	IDI	ICT access	ICT use	ICT skills
IDI	-	0.984	0.984	0.903
ICT access	26%	-	0.953	0.857
ICT use	26%	50%	-	0.860
ICT skills	52%	62%	69%	-

Note: Numbers above the diagonal: Spearman rank correlation coefficients; numbers below the diagonal: percentage of countries (out of 167) that shift +10 positions or more between the rankings.

Figure 4: Distribution of difference between IDI and sub-indices rankings versus the Spearman rank correlation coefficients. Source: Measuring the Information Society Report 2015, page 200.

D. Monte Carlo Test of Robustness

JRC carries a robustness test for the GII to measure the uncertainty in its results. The test is carried using Monte Carlo simulations using the scenarios in Table 4. Each of the scenarios is calculated 1000 times to test the extent to which the rank of each country fluctuates as the calculation method changes. The results of the test are represented with the range of GII, Innovation Input, and Innovation Output ranks for each country. The range of ranks for countries was taken to be narrow enough for the GII to be robust, except for those whose range is more than 20 ranks.[119]

[118] "Measuring the Information Society Report 2015". ITU, 2015.
[119] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. "The Global Innovation Index 2016: Winning with Global Innovation". Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.

The test is carried at the pillar level for GII, with random weights given to pillars. The weights are chosen at random between 0.1 and 0.3 for Innovation Input pillars, and 0.4 to 0.6 to Innovation Output pillars. The imputation method used is the expected maximization imputation (EM).[120] EM is based on the maximum likelihood method, which "selects as estimates the values of the parameters that maximize the likelihood (the joint probability function or joint density function) of the observed sample."[121] The 1000 trials for each scenario consist of choosing a random weight within the bounds mentioned above for each trial. An index is then more robust than another if the calculations carried with different methods affect the rankings of the countries less.[122, 123]

Table 4: GII Monte Carlo simulations scenarios

	Imputation Method	Weights	Aggregation Method
1	EM	Random	Arithmetic
2	EM	Random	Geometric
3	No imputation	Random	Arithmetic
4	No imputation	Random	Geometric

JRC carries an identical test for the robustness of the IDI, with the only difference being the consideration of two instead of four scenarios. The scenarios are arithmetic and geometric aggregation, with a bound on random weights for sub-indices of 0.3 to 0.5 for ICT use and ICT access, and 0.15 to 0.25 for ICT skills. The bounds where chosen to be 25% lower and higher than the assigned value. Results for the IDI were also considered to show robustness, except for countries whose range is more than 15 ranks.[124]

VII. Measurement of Innovation

The measurement of innovation in the reviewed indices was carried out largely from a formal lens. That is in contrast to the research carried out by A2K4D and Open AIR, which focus on informal innovation in both the formal and the informal sectors. Inherent in those, for example, are various forms of unmeasured processes of human capital and skill development, informal acquisition, appropriation and sharing of knowledge, and the wide modes of collaboration that feed into the innovation process. These components of innovation are not considered because their traces are not recorded in official databases. In contrast, innovation was considered in the reviewed indices through mostly documented practices.

[120] Ibid.
[121] Wackerly, Dennis D., William Mendenhall, and Richard L. Scheaffer. "Mathematical statistics with applications", page 477. Seventh ed. Brooks/Cole, Cengage Learning, 2008.
[122] Note that a weighted geometric mean is calculated as in the Atkinson measure for the Health Index of the IHDI referred to the IHDI Atkinson Inequality Measure section.
[123] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. "The Global Innovation Index 2016: Winning with Global Innovation". Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.
[124] "Measuring the Information Society Report 2016". ITU, 2016.

GII carries out a very wide approach to the measurement of innovation through its two input and output sub-indices. It takes into account a range of aspects from regulatory environment to YouTube uploads. All of its indicators are based on documented phenomenon such as regulatory environment measured by indicators like cost of redundancy dismissal.[125] That is remotely related to innovation in the developing world where innovation is conventionally protected through social understandings. The IER as a ratio of Input to Output indices addresses that discrepancy slightly, but it still does not fulfill the measurement of informal innovation in developing countries.

The SUII takes a similar approach, but is more relevant to its context since it addresses innovation in Europe where formalization of innovation activities is mature. Indicators are focused on registered activity such as designs, patents, and publications; and definitions of classifications like what it is to be a knowledge-intensive service provider.

SII has a more flexible approach since it measures social innovation, although a lot of its indicators are based on documented aspects like ease of getting credits. It also has other more undocumented aspects like risk-taking mindset, citizen's attitude towards entrepreneurship, civil society engagement, and trust in society. These indicators are carried through surveys and not calculations of available data, which can be a valuable tool in informal innovation measurement. Other indicators like social innovation and research impact, and legal framework for social enterprises are determined by analysts at the Economist Intelligence Unit.[126] The method of expert seeking can also be of benefit to the A2K4D research. However, experts would be of societal understanding rather than legal or policy experts.

GCI measures innovation through one pillar consisting of 8 indicators. The indicators are:

- Capacity for innovation
- Quality of scientific research institutions
- Company spending on R&D
- University-industry collaboration in R&D
- Government procurement of advanced technology products
- Availability of scientists and engineers
- PCT patent applications
- Intellectual property protection

EOS measures all of them except for PCT patent applications.[127]

Although GCI has a method that can measure undocumented innovation through EOS, the questions regard formal avenues of innovation. That is except the government procurement question which expands areas of innovation to not include only business. The two main criticisms of the GCI

[125] Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. "The Global Innovation Index 2016: Winning with Global Innovation". Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.
[126] "Old problems new solutions: Measuring the capacity for social innovation across the world: Social Innovation Index 2016". The Economist Intelligence Unit, 2016.
[127] Schwab, Klaus and Xavier Sala-i-Martin. "Insight Report: The Global Competitiveness Report 2016-2017". World Economic Forum, 2016.

measurement of innovation is thus the type of questions asked, and the respondents targeted by the EOS. Its respondents are registered businesses from around the world, which limits the scope of innovation severely in the developing world.

GCI plans to address the first of these criticisms in future editions. Innovation will be considered as separate Innovation Ecosystem sub-index. It will consist of four pillars of technology adoption, market size, business dynamism, and innovation capacity. Technology adoption will include "more measures capturing non-ICT technologies." Market size is "rethought to capture market potential rewarding larger pools of ideas and economies of scale."[128] Business dynamism is "rethought to capture entrepreneurial spirit entry and bankruptcy regulation."[129] Innovation capacity "combines R&D with non-R&D factors (e.g., creativity, connectivity, business models)."[130] These changes go a long way from the current form to capturing previously unmeasured innovation. With target respondents that include the informal economy, the future EOS can be a significantly useful tool.

ACI measures the capacity to employ and exploit science, technology, and innovation (STI) in Africa. The index is largely based on a survey carried by the African Capacity Foundation of businesses in each country covered. The two sections of interest of the survey are: Institutional and regulatory framework for STI, and Innovation.[131]

The considerations for institutional and regulatory framework for STI are:

- Existence of a strategy for STI
- Capacity development is part of the strategy
- Country has indicators tracking R&D
- Country has body in charge of intellectual property protection
- Country joined regional economic communities for the promotion of STI

The considerations for innovation are:

- Capacity for innovation
- Qualified research institutes
- Company spending on R&D
- University – industry collaboration on R&D
- Government procurement of advance technology product
- Availability of scientists and engineers
- PCT patent application per million of people

That structure of the ACI is based heavily on conventional measurements of innovation. That is reflected on the priorities recommended by the ACI report as:[132]

---

[128] Ibid.
[129] Ibid.
[130] Ibid.
[131] "African Capacity Report 2017: Building Capacity in Science, Technology and Innovation for Africa's Transformation". African Capacity Building Foundation Knowledge and Learning Department, 2016.
[132] "African Capacity Report 2017: Building Capacity in Science, Technology and Innovation for Africa's Transformation". African Capacity Building Foundation Knowledge and Learning Department, 2016.

- Offering funding
- Improving investments in human resources
- Promoting exchange programs
- Sharing good practices
- Encouraging innovation in private firms

This list of recommendations reflects a desire by the African Capacity Foundation to formalize innovation to the conventional measurement, rather than attempt to more objectively measure it in Africa.

VIII. Conclusion

This paper attempted to explore the methodologies of 16 global indices with the aim of being a starting reference for the development of a new index that measures informal innovation. The indices cover a diversity of fields including innovation, ICT, economic environment, governance, and development. The mathematical techniques and considerations undertaken by these indices were explored in terms of structure, data preparation and manipulation, data consideration, and weights.

The index structures presented were in a variety of forms in terms of aggregation levels and methods. The vast majority of indices reviewed had at least three levels of aggregation and used different forms of arithmetic averaging. With the exception of the HDI and IHDI, which use the geometric mean at the last aggregation level.

Data preparation explored how the indices detected and dealt with skewness and outliers, different normalization techniques, and a concentration on the Atkinson family of inequality measures use in the IHDI. It was found that skewness and outliers are a core step in the calculation of indices. Despite them being a common consideration, there were a variety of techniques used. In exploring normalization within data preparation, choices of data limits were given the same attention by index methodologies as the methods used.

Data considerations were then reviewed in terms of how indices decide which information to use, and how they deal with missing indicators. The strictness of qualification criteria varied and did not seem to be related to the size of the index. EOS acceptable conditions for inclusion of surveys were also explored, which can help in the treatment of the A2K4D survey filtering.

Discussed in this paper was the different weighting methodologies used by indices. It was found that advanced mathematical techniques are needed to formalize the concepts of statistical importance. The mathematical concepts presented were the Pearson correlation ratio and PCA. Variation considerations were key in addressing the issue of compensability.

The different considerations and methods for assessing indices were then discussed in the context of GII and IDI JRC audits. The statistical techniques used to evaluate an index were shown to be a strong tool in verifying the conceptual structure of an index. The crucial question was about the extent to which the indicators and their higher levels coincide, so it is considered appropriate to aggregate them in a given structure, while being distinct enough for hidden information to reveal itself with aggregation.

Beyond the mathematical techniques, the conceptual design of measuring innovation in some of the indices used was assessed critically. Most of the indices that covered innovation did so in a conventional manner with documented information. The SUII and SPI had interesting measurements of innovation that can be adapted to the context of the developing world. GCI is a particular interesting example since the proposed changes widens the scope of innovation measurement to become closely in line with A2K4D aims.

Index methodology was found to be a rich field combining mathematics with the social sciences at the forefront of policy making. As indices are attempting to present their findings with increasing truthfulness, mathematical rigor is taking a significant role beside conceptual design.

Bibliography

"African Capacity Report 2017: Building Capacity in Science, Technology and Innovation for Africa's Transformation". African Capacity Building Foundation Knowledge and Learning Department, 2016.

Andridge, Rebecca R. "A Review of Hot Deck Imputation for Survey Non-response". International Statistical Review, U.S. National Library of Medicine, Apr. 2010.

Baller, Silja, Soumitra Dutta and Bruno Lanvin. "Insight Report: The Global Information Technology Report 2016: Innovating in the Digital Economy". Johnson Cornell University, World Economic Forum, and INSEAD, 2016.

"Bertelsmann Stifung Transformation Index BTI 2016: Methodology". BTI, 2016

"Corruption Perception Index 2016: Technical Methodology Note". Transparency International, 2016.

"Doing Business 2017: Equal Opportunity for All". World Bank, 2017.

Dutta, Soumitra, Bruni Lanvin, and Sacha Wunsch-Vincent. "The Global Innovation Index 2016: Winning with Global Innovation". Cornell University, INSEAD, and World Intellectual Property Organization (WIPO). 2016.

Groeneveld, R. A. and G. Meeden. "Measuring Skewness and Kurtosis". The Statistician 33: 391–99, 1984.

"Handbook on Constructing Composite Indicators: Methodology and User Guide". JRC European Commission and OECD, 2008.

Harttgen, K., and S. Vollmer. "Using an Asset Index to Simulate Household Income". Economic Letters 121(2): 257–262, 2013.

Hollanders, Hugo, Nordine Es-Sadki, and Minna Kanerva. "European Innovation Scoreboard 2016 Methodology Report". Mastricht University (Maastricht Economic and Social Research Institute on Innovation and Technology – MERIT), 2016.

"Human Development Report 2016: Human Development for Everyone". UNDP, 2016.

"Human Development Report 2016: Human Development for Everyone: Technical Notes". UNDP, 2016.

"Insight Report: The Human Capital Report 2016: Technical Notes". Mercer and World Economic Forum, 2016.

"Insight Report: The Human Capital Report 2016: User's Guide". Mercer and World Economic Forum, 2016.

"Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy." Statistics How To. http://www.statisticshowto.com/kaiser-meyer-olkin, accessed 30 Jun. 2017.

Kovacevic, M. "Measurement of Inequality in Human Development—A Review". Human Development Research Paper, UNDP-HDRO, 2010.

"Measuring the Information Society Report 2015". ITU, 2015.

"Measuring the Information Society Report 2016". ITU, 2016.

"Old problems new solutions: Measuring the capacity for social innovation across the world: Social Innovation Index 2016". The Economist Intelligence Unit, 2016.

Open AIR, "Open AIR: Metrics and Policies", www.openair.org.za.

Paruolo, P., M. Saisana, and A. Saltelli. "Ratings and Rankings: Voodoo or Science?" Journal of the Royal Statistical Society A 176 (3): 609–34, 2013.

Schwab, Klaus and Xavier Sala-i-Martin. "Insight Report: The Global Competitiveness Report 2016-2017". World Economic Forum, 2016.

"Seizing Opportunity Through License Compliance: BSA Global Software Survey May 2016". BSA The Software Alliance, 2016.

Stern, Scott, Amy Wares, and Tamar Hellman. "Social Progress Index 2016 Methodological Report". The Social Progress Imperative, 2016.

Wackerly, Dennis D., William Mendenhall, and Richard L. Scheaffer. "Mathematical statistics with applications", page 477. Seventh ed. Brooks/Cole, Cengage Learning, 2008.

Weisstein, Eric W. "Skewness". MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/Skewness.html, accessed on June 15, 2017.

"What Is an "imputed" Indicator – and for Which Countries Were These Imputed Statistics Used?". Human Development Reports, HDRO FAQs. http://hdr.undp.org/en/content/what-%E2%80%9Cimputed%E2%80%9D-indicator-%E2%80%93-and-which-countries-were-these-imputed-statistics-used, accessed on 16 Jun. 2017.

Wikipedia contributors. "Chauvenet's criterion". Wikipedia, The Free Encyclopedia, 1 Jan. 2017, accessed on 15 Jun. 2017.

Wikipedia contributors. "Kurtosis". Wikipedia, The Free Encyclopedia, 3 Feb. 2017, accessed on 15 Jun. 2017.

Wikipedia contributors. "Skewness". Wikipedia, The Free Encyclopedia, 5 Jun. 2017, accessed on 15 Jun. 2017.

"World Population Prospects - Population Division". UNDESA, 2015. https://esa.un.org/unpd/wpp/Download/Standard/Mortality, accessed on June 16, 2017.

Zaiontz, Charles. "Cronbach's Alpha." Real Statistics Using Excel, 2014. http://www.real-statistics.com/reliability/cronbachs-alpha, accessed on 29 Jun. 2017.

Recognizing Africa's role in the global knowledge economy.
www.openair.org.za

Open AIR
Open African Innovation Research (Open AIR) is a unique collaborative network of researchers investigating how intellectual property (IP) systems can be harnessed in open, participatory ways that have the potential to maximise knowledge access, innovation, and the sharing of benefits from innovation inclusively.

For more Information about Open AIR, please visit our website: www.openair.org.za
or contact one of our Program Managers:
ottawa@openair.org,za
capetown@openair.org.za

Open AIR is carried out with financial support from the International Development Research Centre, Canada, the Social Sciences and Humanities Research Council of Canada, UK Department for International Development, and Queen Elizabeth Scholars Program. More information about Open AIR's current and previous supporters can be found at http://www.openair.org.za/supporters/. The views expressed herein do not necessarily represent those of Open AIR's funders.