

# Inclusive Growth Index - Methodology

Nour Barnat, Diana Camerini, Petra Kynclova, Cloe Schaub

2023-09-26



# Contents

<b>About the project</b>	<b>5</b>
Other technical information . . . . .	5
<b>1 Introduction</b>	<b>7</b>
Current edition . . . . .	7
Previous editions . . . . .	9
To do list . . . . .	11
<b>2 Data collection and processing</b>	<b>13</b>
2.1 Data collection . . . . .	13
<b>3 Data interpolation, extrapolation and imputation</b>	<b>41</b>
3.1 Data interpolation . . . . .	42
3.2 Data extrapolation . . . . .	44
3.3 Data imputation . . . . .	44
<b>4 Data transformation</b>	<b>47</b>
4.1 By indicator . . . . .	48
4.2 By transformation . . . . .	50
<b>5 Principal Component Analysis</b>	<b>53</b>
<b>6 Weights</b>	<b>55</b>
6.1 How many principal components to keep? . . . . .	56
6.2 Final IGI . . . . .	56



# About the project

This is a methodological guide for compilation of the Inclusive Growth Index (IGI). Each chapter contains detailed information about individual steps required for the IGI compilation and production.

The Inclusive Growth Index is a composite indicator of growth developed by UNCTAD.

The 2023 edition of the IGI is calculated by using updated data series, following the methodology from 2022. In the version developed in 2022 (see **SDG Pulse 2022**), it consists of **27 baseline indicators**, grouped in **4 pillars (economic growth, living conditions, equality and environment)**.

The SDG Pulse 2022 chapter is available at: <https://sdgpulse.unctad.org/inclusive-growth/>

To produce annual releases of the indicator, a number of steps are followed. These include:

- 1) Introduction
- 2) Data collection and processing
- 3) Data interpolation, extrapolation and imputation
- 4) Data transformation
- 5) Principal Component Analysis
- 6) Assigning weights to indicators
- 7) Index compilation

The underlying data, imputation used and the resulting Inclusive Growth Index for 2021 is examined in more detail: [https://bookdown.org/kynclova\\_petra/IGI\\_results/](https://bookdown.org/kynclova_petra/IGI_results/)

## Other technical information

Following R packages should be installed and loaded:

```
# Libraries  
library(tidyverse)  
library(readxl)
```

General guidance on construction of composite indicators by the JRC: <https://bluefoxr.github.io/COINrDoc/>

PCA for composite indices and questions related to PCA applied on longitudinal data:

- <https://stats.stackexchange.com/questions/158281/can-pca-be-applied-for-time-series-data?rq=1>
- <https://stats.stackexchange.com/questions/346521/how-to-weight-composites-based-on-pca-with-longitudinal-data>

# Chapter 1

## Introduction

The first edition of the IGI consisted of three pillars, namely economy, living conditions and equality [Barnat et al.] and was composed of 21 indicators, including one environmental indicator (carbon dioxide - CO<sub>2</sub> emissions under pillar 2). This new, expanded IGI includes more equality metrics addressing gender inequality more broadly, and includes a new separate pillar dedicated to environmental issues. These were highlighted as potential development areas of the original index, and can now be addressed also due to progress with data availability for countries.

### Current edition

#### **2023: Release before the SDG Summit 2023**

The current edition of the IGI is comprised of **four pillars** and **27 indicators**. Indicators were selected as for the SDG Pulse 2022, however updated for new available data series such as when constant prices were rebased (GDP, poverty etc.).

#### **Overview:**

- Number of countries: **129 economies**
- Number of pillars: **4 pillars**
- Number of indicators: **27 indicators**
- Years coverage:
  - Data collected, imputed and forecasted for the period 2000-2023

– Index compiled only for one year - **2021**

**Pillars:**

- Pillar 1 – **Economic growth**
- Pillar 2 - **Living conditions**
- Pillar 3 – **Equality**
- Pillar 4 - **Environment**

**Table with pillars and indicators**

*Note: Names of indicators are taken as from their original sources*

Pillar 1	Economic growth
Indicator 1.1	GDP per capita, PPP (constant 2017 international \$)
Indicator 1.2	Adjusted net national income per capita (constant 2015 US \$)
Indicator 1.3	Labour productivity - GDP per person employed (constant 2017 PPP USD)
Indicator 1.4	Employment to population ratio, 15+, total (%) (modeled ILO estimate)
Indicator 1.5	Electricity consumption/population (kWh per capita)
Indicator 1.6	Exports of goods and services (% of GDP)
Pillar 2	Living conditions
Indicator 2.1	Logistics performance index: Overall (1=low to 5=high)
Indicator 2.2	Fixed Internet broadband subscriptions per 100 people, units
Indicator 2.3	Under-five mortality rate (deaths per 1.000 live births)
Indicator 2.4	People using safely managed drinking water services (% of population)
Indicator 2.5	School enrollment, secondary (% gross)
Indicator 2.6	Coverage of essential health services
Indicator 2.7	Proportion of adults (15 years and older) with an account at a bank or other financial institution or with a mobile-money-service provider

Pillar 3	Equality
Indicator 3.1	Income concentration ratio (Gini index), units
Indicator 3.2	Poverty headcount ratio at 5.50 USD a day (2011 PPP) (% of population)
Indicator 3.3	School enrollment, secondary (gross), gender parity index (GPI)
Indicator 3.4	Ratio of female to male employment rate (modeled ILO estimate)
Indicator 3.5	Ratio of youth to adult employment rate (modeled ILO estimate)
Indicator 3.6	Gender parity in the number of seats held by women and men in national parliaments
Indicator 3.7	Ratio of female to male labour force participation rate (%) (ILO modeled estimate)
Indicator 3.8	Ratio of female age of first marriage to male age of first marriage
Indicator 3.9	Ratio of the share of wage and salaried workers in women's employment to men's employment
Indicator 3.10	Share of women's service employment to total employment, raised to the power of the inverse of the Palma ratio

  

Pillar 4	Environment
Indicator 4.1	CO <sub>2</sub> emissions (kg per PPP USD of GDP)
Indicator 4.2	Energy intensity level of primary energy (MJ/\$2017 PPP GDP)
Indicator 4.3	Efficiency of water use (water productivity)
Indicator 4.4	Terrestrial protected areas (% of total land area)

## Previous editions

### 2022: SDG Pulse

In the version developed in 2022 [UNCTAD, 2022], the IGI consists of **27 base-line indicators**, grouped in **4 pillars (economic growth, living conditions, equality and environment)**, building on the previous version of IGI [Barnat et al.].

The SDG Pulse 2022 chapter is available at: <https://sdgpulse.unctad.org/inclusive-growth/>

#### Overview:

- Number of countries: **96 countries**
- Number of pillars: **4 pillars**
- Number of indicators: **27 indicators**
- Years coverage: **2009 and 2020**

**Pillars:**

- Pillar 1 – **Economic growth**
- Pillar 2 - **Living conditions**
- Pillar 3 – **Equality**
- Pillar 4 - **Environment**

**2019: Inclusive growth of the Eurasian Economic Union Member States**

The first iteration of IGI consisted of three pillars, namely economy, living conditions and equality [UNCTAD and EEC, 2019]. It was composed of 21 indicators, including one environmental indicator (carbon dioxide - CO<sub>2</sub> emissions under pillar 2).

All countries with complete information for at least one pillar's indicators were included in the calculations. Consequently, analyses were based on data from 167, 131 and 90 countries for the first, the second and the third pillars, respectively. However, the final index was computed only for the 86 states with complete data available for all three pillars.

**Overview:**

- Number of countries: **86 countries**
- Number of pillars: **3 pillars**
- Number of indicators: **21 indicators**
- Years coverage: **2019**

**Pillars:**

- Pillar 1 – Economic, composed of GDP and national income per capita, power consumption, employment and trade
- Pillar 2 - Living conditions, composed of social and health conditions, logistics and finance, natural environment
- Pillar 3 – Inequality, composed of measures of inequality in labour participation, income, school enrolment and political participation.

## To do list

- Methodology review to compile the index for panel data
- Review of existing data sources to better fill data gaps
- Review of existing data series to include additional indicators
- Additional work on country metadata needed



## Chapter 2

# Data collection and processing

### 2.1 Data collection

Data are collected from various international sources to obtain the best country coverage. The starting year is 2000 until the most recent available data point.

Data are sourced from following databases:

- World Bank Open Data
- SDG Global Database
- ILOSTAT
- World Income Inequality Database
- IEA World Energy Statistics
- World Marriage Database
- World Population Prospects
- UNSD National Accounts Main Aggregates Database

Date of finalizing data collection: *30-08-2023*

#### **Pillar 1**

Pillar 1 includes a series of indicators on **economic growth**.

It includes 6 indicators, namely:

### 1.1: GDP per capita, PPP

**Data source:** World Bank

**Data series code:** “NY.GDP.PCAP.PP.KD”

**Indicator series name:** GDP per capita, PPP (constant 2017 international \$)

**Brief Description:** GDP per capita based on purchasing power parity (PPP). PPP GDP is gross domestic product converted to international dollars using purchasing power parity rates. An international dollar has the same purchasing power over GDP as the U.S. dollar has in the United States. GDP at purchaser’s prices is the sum of gross value added by all resident producers in the country plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. Data are in constant 2017 international dollars. It replaces the same series used in the past, with 2017 instead of 2011 as base year.

**Metadata:** <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/NY.GDP.PCAP.PP.KD>

**Data access:** <https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD>

**SDG linkages:** SDG target 8.1: Sustain per capita economic growth in accordance with national circumstances and, in particular, at least 7 per cent gross domestic product growth per annum in the least developed countries

+ SDG indicator 8.1.1: Annual growth rate of real GDP per capita

### 1.2: Adjusted net national income per capita

**Data source:** World Bank

**Data series code:** “NY.ADJ.NNTY.PC.KD”

**Indicator series name:** Adjusted net national income per capita (constant 2015 US\$)

**Brief Description:** The indicator is developed by World Bank staff estimates based on sources and methods in World Bank’s *“The Changing Wealth of Nations: Measuring Sustainable Development in the New Millennium”* (2011). The adjusted net national income complements gross national income (GNI) in assessing economic progress (Hamilton and Ley 2010) by providing a broader measure of national income that accounts for the depletion of natural resources. Adjusted net national income is calculated by subtracting from GNI a charge for the consumption of fixed capital (a calculation that yields net national income) and for the depletion of natural resources. The deduction for the depletion

of natural resources, which covers net forest depletion, energy depletion, and mineral depletion, reflects the decline in asset values associated with the extraction and harvesting of natural resources. This is analogous to depreciation of fixed assets. Growth rates of adjusted net national income are computed from constant price series deflated using the gross national expenditure (formerly domestic absorption) deflator.

**Metadata:** <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/NY.ADJ.NNTY.PC.KD>

**Data access:** <https://data.worldbank.org/indicator/NY.ADJ.NNTY.PC.KD>

**SDG linkages:** SDG target 8.1: Sustain per capita economic growth in accordance with national circumstances and, in particular, at least 7 per cent gross domestic product growth per annum in the least developed countries

+ No specific indicator

### 1.3: Labour productivity

**Data source:** World Bank

**Data series code:** “SL.GDP.PCAP.EM.KD”

**Indicator series name:** Labour productivity - GDP per person employed (constant 2017 PPP USD)

**Brief Description:** Description/footnote: GDP per person employed is gross domestic product (GDP) divided by total employment in the economy. Purchasing power parity (PPP) GDP is GDP converted to 2017 constant international dollars using PPP rates. An international dollar has the same purchasing power over GDP that a U.S. dollar has in the United States. It replaces the same series used in the past, with 2017 instead of 2011 as base year. World Bank, World Development Indicators database. Estimates are based on employment, population, GDP, and PPP data obtained from International Labour Organization, United Nations Population Division, Eurostat, OECD, and World Bank.

**Metadata:** <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SL.GDP.PCAP.EM.KD>

**Data access:** <https://data.worldbank.org/indicator/SL.GDP.PCAP.EM.KD>

**SDG linkages:** SDG target 8.2: Achieve higher levels of economic productivity through diversification, technological upgrading and innovation, including through a focus on high-value added and labour-intensive sectors

+ SDG indicator 8.2.1: Annual growth rate of real GDP per employed person

#### 1.4: Employment to population ratio

**Data source:** World Bank

**Data series code:** “SL.EMP.TOTL.SP.ZS”

**Indicator series name:** Employment to population ratio, 15+, total (%) (modeled ILO estimate)

**Brief Description:** Employment to population ratio is the proportion of a country’s population that is employed. Employment is defined as persons of working age who, during a short reference period, were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period (i.e. who worked in a job for at least one hour) or not at work due to temporary absence from a job, or to working-time arrangements. Ages 15 and older are generally considered the working-age population.

The employment-to-population ratio indicates how efficiently an economy provides jobs for people who want to work. A high ratio means that a large proportion of the population is employed. But a lower employment-to-population ratio can be seen as a positive sign, especially for young people, if an increase in their education causes it.

**Metadata:** <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SL.EMP.TOTL.SP.ZS>

**Data access:** <https://data.worldbank.org/indicator/SL.EMP.TOTL.SP.ZS>

**SDG linkages:** SDG 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

+ No specific indicator

#### 1.5: Electricity consumption

**Data source:** International Energy Agency (IEA)

**Data series code:** “ELEPOP”

**Indicator series name:** Electricity consumption/population (kWh per capita)

**Brief Description:** Electricity consumption equals domestic supply less losses. Electric power consumption per capita (kWh ) is the production of power plants and combined heat and power plants less transmission, distribution, and transformation losses and own use by heat and power plants, divided by midyear population. Energy data are compiled by the International Energy Agency (IEA). IEA data for economies that are not members of the Organisation for Economic Co-operation and Development (OECD) are based on national energy data adjusted to conform to annual questionnaires completed by OECD member governments. Electricity consumption is equivalent to production less

power plants' own use and transmission, distribution, and transformation losses less exports plus imports. It includes consumption by auxiliary stations, losses in transformers that are considered integral parts of those stations, and electricity produced by pumping installations. Where data are available, it covers electricity generated by primary sources of energy - coal, oil, gas, nuclear, hydro, geothermal, wind, tide and wave, and combustible renewables.

**Metadata:**

- General information: <https://www.iea.org/data-and-statistics/data-product/world-energy-balances>
- Database documentation: [https://iea.blob.core.windows.net/assets/08e3af5c-a438-4e50-b44d-1419cb7ff309/WORLDBAL\\_Documentation.pdf](https://iea.blob.core.windows.net/assets/08e3af5c-a438-4e50-b44d-1419cb7ff309/WORLDBAL_Documentation.pdf)
- More information: <https://www.iea.org/data-and-statistics/data-product/electricity-information#electricityheat-supply-and-consumption-oecd-and-selected-countries>

**Data access:**

- The download requires a subscription to their data portal (not free of charge):  
<https://doi.org/10.1787/enestats-data-en>
- Full access for UNCTAD staff: <https://go.openathens.net/redirector/un.org?url=https://www.oecd-ilibrary.org/>

**SDG linkages:** SDG 7: Ensure access to affordable, reliable, sustainable and modern energy for all

+ No specific indicator

## 1.6: Export of goods and services

**Data source:** World Bank

**Data series code:** "NE.EXP.GNFS.ZS"

**Indicator series name:** Export of goods and services (% of GDP)

**Brief Description:** Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial,

information, business, personal, and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments. Data are originally sourced from the World Bank national accounts data, and OECD National Accounts data files. Some of the data are the results of national estimates, some other come from actual records.

**Metadata:** <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/NE.EXP.GNFS.ZS>

**Data access:** <https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS>

**SDG linkages:** SDG target 17.11: Significantly increase the exports of developing countries, in particular with a view to doubling the least developed countries' share of global exports by 2020

+ SDG indicator 17.11.1: Developing countries' and least developed countries' share of

## Pillar 2

The second pillar is dedicated to **living conditions** indicators.

It includes 7 indicators, namely:

### 2.1: Logistics Performance Index

**Data source:** World Bank

**Data series code:** "LP.LPI.OVRL.XQ"

**Indicator series name:** Logistics performance index: Overall (1=low to 5=high)

**Brief Description:** The Logistics Performance Index overall score reflects perceptions of a country's logistics based on the efficiency of customs clearance process, quality of trade- and transport-related infrastructure, ease of arranging competitively priced shipments, quality of logistics services, ability to track and trace consignments, and frequency with which shipments reach the consignee within the scheduled time. The index ranges from 1 to 5, with a higher score representing better performance. Data are from the Logistics Performance Index survey conducted by the World Bank in partnership with academic and international institutions and private companies and individuals engaged in international logistics. The 2023 LPI survey was conducted from September 6 to November 5, 2022. It provided 4,090 country assessments by 652 logistics professionals in 115 countries in all World Bank regions. Respondents evaluate eight countries on six core dimensions on a scale from 1 (worst) to 5 (best). The eight countries are chosen based on the most important export and import markets of the respondent's country, random selection, and, for landlocked

countries, neighboring countries that connect them with international markets. Scores for the six areas are averaged across all respondents and aggregated to a single score using principal components analysis. Details of the survey methodology and index construction methodology are included in Appendix 5 of the 2023 LPI report available at: <https://lpi.worldbank.org/report>.

**Metadata:** <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/LP.LPI.OVRL.XQ>

**Data access:** <https://data.worldbank.org/indicator/LP.LPI.OVRL.XQ>

**SDG linkages:**

SDG target 9.1: Develop quality, reliable, sustainable and resilient infrastructure, including regional and transborder infrastructure, to support economic development and human well-being, with a focus on affordable and equitable access for all

+ SDG indicator 9.1.2: Passenger and freight volumes, by mode of transport

SDG target 17.11: Significantly increase the exports of developing countries, in particular with a view to doubling the least developed countries' share of global exports by 2020

+ SDG indicator 17.11.1: Developing countries' and least developed countries' share of global exports

*This could be extended for more SDG linkages.*

## 2.2: Fixed Internet broadband subscriptions

**Data source:** Global SDG Indicators Database

**Data series code:** "IT\_NET\_BBND"

**Indicator series name:** Fixed broadband subscriptions per 100 inhabitants, by speed (per 100 inhabitants)

**Brief Description:** Fixed broadband subscriptions refer to subscriptions to high-speed access to the public Internet (a TCP/IP connection), at downstream speeds equal to, or greater than, 256 kbit/s. It includes cable modem, DSL, fibre-to-the-home/building, other fixed -broadband subscriptions, satellite broadband and terrestrial fixed wireless broadband. This total is measured irrespective of the method of payment. It excludes subscriptions that have access to data communications (including the Internet) via mobile-cellular networks.

**Metadata:** <https://unstats.un.org/sdgs/metadata/files/Metadata-17-06-01.pdf>

**Data access:** <https://unstats.un.org/sdgs/dataportal/database>

**Disaggregation:** None

**SDG linkages:**

SDG target 17.6: Enhance North-South, South-South and triangular regional and international cooperation on and access to science, technology and innovation and enhance knowledge-sharing on mutually agreed terms, including through improved coordination among existing mechanisms, in particular at the United Nations level, and through a global technology facilitation mechanism

+ SDG indicator 17.6.1: Fixed Internet broadband subscriptions per 100 inhabitants, by

### 2.3: Child mortality

**Data source:** Global SDG Indicators Database

**Data series code:** "SH\_DYN\_MORT"

**Indicator series name:** Under-five mortality rate (deaths per 1,000 live births)

**Brief Description:** The under-five mortality rate is the probability of a child born in a specific year or period dying before reaching the age of 5 years, if subject to age-specific mortality rates of that period, expressed as deaths per 1000 live births. The under-five mortality rate as defined here is, strictly speaking, not a rate (i.e. the number of deaths divided by the number of population at risk during a certain period of time), but a probability of death derived from a life table and expressed as a rate per 1000 live births.

**Metadata:** <https://unstats.un.org/sdgs/metadata/files/Metadata-03-02-01.pdf>

**Data access:** <https://unstats.un.org/sdgs/dataportal/database>

**Disaggregation:** Sex: Both sexes

+ X.Sex.=="BOTHSEX"

**SDG linkages:**

SDG target 3.2: By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births

+ SDG indicator 3.2.1: Under-5 mortality rate

**2.4: Safe water****Data source:** Global SDG Indicators Database**Data series code:** “SGH\_H2O\_SAFE”**Indicator series name:** Proportion of population using safely managed drinking water services, by urban/rural (%)**Brief Description:** The proportion of the population using safely managed drinking water services is defined as the proportion of population using an improved drinking water source which is accessible on premises, available when needed and free from faecal and priority chemical contamination. ‘Improved’ drinking water sources include: piped supplies, boreholes and tubewells, protected dug wells, protected springs, rainwater, water kiosks, and packaged and delivered water. WHO/UNICEF Joint Monitoring Program for Water Supply, Sanitation and Hygiene has established international standards for classification of drinking water facilities and service levels to benchmark and compare progress across countries (see washdata.org).**Metadata:** <https://unstats.un.org/sdgs/metadata/files/Metadata-06-01-01.pdf>**Data access:** <https://unstats.un.org/sdgs/dataportal/database>**Disaggregation:** Location: All areas (for some countries only urban/rural are available, such as China)

+ X.Location.=="ALLAREA"

**SDG linkages:**

SDG target 6.1: By 2030, achieve universal and equitable access to safe and affordable drinking water for all

+ SDG indicator 6.1.1: Proportion of population using safely managed drinking water services

**2.5: School enrollment (secondary)****Data source:** World Bank**Data series code:** “SE.SEC.ENRR”**Indicator series name:** School enrollment, secondary (% gross)**Brief Description:** Net enrollment rate is the ratio of children of official school age who are enrolled in school to the population of the corresponding official school age. Secondary education completes the provision of basic education that began at the primary level, and aims at laying the foundations for lifelong

learning and human development, by offering more subject- or skill-oriented instruction using more specialized teachers.

**Metadata:** <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SE.SEC.ENRR>

**Data access:** <https://data.worldbank.org/indicator/SE.SEC.ENRR>

**SDG linkages:**

SDG 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

+ No specific indicator

## 2.6: Universal health coverage

**Data source:** Global SDG Indicators Database

**Data series code:** “SH\_ACS\_UNHC”

**Indicator series name:** Universal health coverage (UHC) service coverage index

**Brief Description:** The access to universal health is an indicator developed by the World Health Organisation (WHO). Coverage of essential health services is defined as the average coverage of essential services based on tracer interventions that include reproductive, maternal, newborn and child health, infectious diseases, non-communicable diseases and service capacity and access, among the general and the most disadvantaged population.

**Metadata:** <https://unstats.un.org/sdgs/metadata/files/Metadata-03-08-01.pdf>

**Data access:** <https://unstats.un.org/sdgs/dataportal/database>

**Disaggregation:** None

**SDG linkages:**

SDG target 3.8: Achieve universal health coverage, including financial risk protection, access to quality essential health-care services and access to safe, effective, quality and affordable essential medicines and vaccines for all

+ SDG indicator 3.8.1: : Coverage of essential health services

## 2.7: Access to financial services

**Data source:** Global SDG Indicators Database

**Data series code:** “FB\_BNK\_ACCSS”

**Indicator series name:** Proportion of adults (15 years and older) with an account at a financial institution or mobile-money-service provider, by sex (% of adults aged 15 years and older)

**Brief Description:** The percentage of adults (ages 15+) who report having an account (by themselves or together with someone else) at a bank or another type of financial institution or personally using a mobile money service in the past 12 months.

**Metadata:** <https://unstats.un.org/sdgs/metadata/files/Metadata-08-10-02.pdf>

**Data access:** <https://unstats.un.org/sdgs/dataportal/database>

**Disaggregation:** Sex: Both sexes, Age: 15 years old and over, Location: All areas, Education level: Total or no breakdown by education level, Quantile: Total (national average) or no breakdown

```
+ X.Sex.=="BOTHSEX"
+ X.Age.=="15+"
+ X.Location.=="ALLAREA"
+ X.Education.level.=="_T"
+ X.Quantile.=="_T"
```

#### SDG linkages:

SDG target 8.10: Strengthen the capacity of domestic financial institutions to encourage and expand access to banking, insurance and financial services for all

+ SDG indicator 8.10.2: Proportion of adults (15 years and older) with an account at a bank or other financial institution

## Pillar 3

The third pillar is dedicated to **equality** indicators.

It includes 10 indicators, namely:

### 3.1: Gini index

**Data source:** World Bank

**Data series code:** "SI.POV.GINI"

**Indicator series name:** Gini index

**Brief Description:** Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy deviates from a perfectly equal distribution. A

Lorenz curve plots the cumulative percentages of total income received against the cumulative number of recipients, starting with the poorest individual or household. The Gini index measures the area between the Lorenz curve and a hypothetical line of absolute equality, expressed as a percentage of the maximum area under the line. Thus a Gini index of 0 represents perfect equality, while an index of 100 implies perfect inequality.

**Metadata:** <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SI.POV.GINI>

**Data access:** <https://data.worldbank.org/indicator/SI.POV.GINI>

**SDG linkages:**

SDG 10: Reduce inequality within and among countries

+ No specific indicator

### 3.2: Poverty headcount ratio

**Data source:** World Bank

**Data series code:** “SI.POV.LMIC”

**Indicator series name:** Poverty headcount ratio at \$3.65 a day (2017 PPP) (% of population)

**Brief Description:** Poverty headcount ratio at \$3.65 a day is the percentage of the population living on less than \$3.65 a day at 2017 international prices. Poverty measures based on international poverty lines attempt to hold the real value of the poverty line constant across countries, as is done when making comparisons over time. The welfare of people living in different countries can be measured on a common scale by adjusting for differences in the purchasing power of currencies. The commonly used \$1 a day standard, measured in 1985 international prices and adjusted to local currency using purchasing power parities (PPPs), was chosen for World Development Report 1990 because it was typical of the poverty lines in low-income countries at the time. As differences in the cost of living across the world evolve, the international poverty line has to be periodically updated using new PPP price data to reflect these changes. The last change was in September 2022, when we adopted \$2.15 as the international poverty line using the 2017 PPP. The \$3.65 poverty line is derived from typical national poverty lines in countries classified as Lower Middle Income. The \$6.85 poverty line is derived from typical national poverty lines in countries classified as Upper Middle Income.

**Metadata:** <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SI.POV.LMIC>

**Data access:** <https://data.worldbank.org/indicator/SI.POV.LMIC>

**SDG linkages:**

SDG target 1.1: By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than \$1.25 a day

+ SDG indicator 1.1.1: Proportion of the population living below the international poverty line b

**3.3: School enrolment, secondary (gross), gender parity index (GPI)**

**Data source:** World Bank

**Data series code:** “SE.ENR.SECO.FM.ZS”

**Indicator series name:** School enrolment, secondary (gross), gender parity index (GPI)

**Brief Description:** Gender parity index for gross enrollment ratio in secondary education is the ratio of girls to boys enrolled at secondary level in public and private schools. The Gender Parity Index (GPI) indicates parity between girls and boys. A GPI of less than 1 suggests girls are more disadvantaged than boys in learning opportunities and a GPI of greater than 1 suggests the other way around. Eliminating gender disparities in education would help increase the status and capabilities of women.

This indicator is calculated by dividing female gross enrollment ratio in secondary education by male gross enrollment ratio in secondary education. Data on education are collected by the UNESCO Institute for Statistics from official responses to its annual education survey. All the data are mapped to the International Standard Classification of Education (ISCED) to ensure the comparability of education programs at the international level. The current version was formally adopted by UNESCO Member States in 2011. The reference years reflect the school year for which the data are presented. In some countries the school year spans two calendar years (for example, from September 2010 to June 2011); in these cases the reference year refers to the year in which the school year ended (2011 in the example).

**Metadata:** <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SE.ENR.SECO.FM.ZS>

**Data access:** <https://data.worldbank.org/indicator/SE.ENR.SECO.FM.ZS>

**SDG linkages:**

SDG 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

+ No specific indicator

**3.4: Ratio of female to male employment-to-population ratio****Data source:** World Bank**Data series code:** "SL.EMP.TOTL.SP.ZS"

+ Female: "SL.EMP.TOTL.SP.FE.ZS", Male: "SL.EMP.TOTL.SP.MA.ZS"

**Indicator series name:** Ratio of female to male employment-to-population ratio (%) (modeled ILO estimate)

+ Employment to population ratio, 15+, female (%) (modeled ILO estimate)

+ Employment to population ratio, 15+, male (%) (modeled ILO estimate)

**Brief Description:** Employment to population ratio is the proportion of a country's population that is employed. Employment is defined as persons of working age who, during a short reference period, were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period (i.e. who worked in a job for at least one hour) or not at work due to temporary absence from a job, or to working-time arrangements. Ages 15 and older are generally considered the working-age population. The employment-to-population ratio indicates how efficiently an economy provides jobs for people who want to work. A high ratio means that a large proportion of the population is employed. But a lower employment-to-population ratio can be seen as a positive sign, especially for young people, if an increase in their education causes it.

Ratio of female to male employment-to-population ratio is calculated by dividing the employment-to-population ratio among women, by the corresponding rate for men.

**Metadata:**

<https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SL.EMP.TOTL.SP.FE.ZS> <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SL.EMP.TOTL.SP.MA.ZS>

**Data access:**

<https://data.worldbank.org/indicator/SL.EMP.TOTL.SP.FE.ZS> <https://data.worldbank.org/indicator/SL.EMP.TOTL.SP.MA.ZS>

**SDG linkages:**

SDG 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

+ No specific indicator

**3.5: Ratio of youth to adult employment-to-population ratio**

**Data source:** ILOSTAT

**Data series code:** "EMP\_2WAP"

+ Female: "AGE\_YTHADULT\_Y15-24", Male: "AGE\_YTHADULT\_YGE25"

**Indicator series name:** Ratio of youth to adult employment-to-population ratio (modeled ILO estimate)

+ Employment to population ratio, ages 15-24, total (%) (modeled ILO estimate)

+ Employment to population ratio, 25+, total (%) (modeled ILO estimate)

**Brief Description:** Employment to population ratio is the proportion of a country's population that is employed. Employment is defined as persons of working age who, during a short reference period, were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period (i.e. who worked in a job for at least one hour) or not at work due to temporary absence from a job, or to working-time arrangements. Ages 15 and older are generally considered the working-age population. The employment-to-population ratio indicates how efficiently an economy provides jobs for people who want to work. A high ratio means that a large proportion of the population is employed. But a lower employment-to-population ratio can be seen as a positive sign, especially for young people, if an increase in their education causes it.

Ratio of youth to adult employment-to-population ratio is calculated by dividing the employment-to-population ratio among young people (15-24 years old), by the corresponding rate for adults (25 years and older).

**Metadata:** <https://ilostat.ilo.org/resources/concepts-and-definitions/description-labour-force-statistics/>

**Data access:**

[https://www.ilo.org/shinyapps/bulkexplorer59/?lang=en&id=EMP\\_2WAP\\_SEX\\_AGE\\_RT\\_A](https://www.ilo.org/shinyapps/bulkexplorer59/?lang=en&id=EMP_2WAP_SEX_AGE_RT_A)

**SDG linkages:**

SDG 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

+ No specific indicator

### 3.6: Gender parity in the number of seats held by women and men in national parliaments

**Data source:** Global SDG Indicators Database

**Data series code:** "SG\_GEN\_PARL\_PAR"

+ SDG series code: "SG\_GEN\_PARL"

**Indicator series name:** Proportion of seats held by women in national parliaments (% of total number of seats)

**Brief Description:** The proportion of seats held by women in national parliaments, currently as of 1 January of reporting year, is currently measured as the number of seats held by women members in single or lower chambers of national parliaments, expressed as a percentage of all occupied seats. National parliaments can be bicameral or unicameral. This indicator covers the single chamber in unicameral parliaments and the lower chamber in bicameral parliaments. It does not cover the upper chamber of bicameral parliaments. Seats are usually won by members in general parliamentary elections. Seats refer to the number of parliamentary mandates or the number of members of parliament.

Gender parity in the number of seats held by women and men in national parliaments is calculated as a ratio between the number of seats held by women and men.

**Metadata:** <https://unstats.un.org/sdgs/metadata/files/Metadata-05-05-01a.pdf>

**Data access:** <https://unstats.un.org/sdgs/dataportal/database>

**Disaggregation:** None

#### SDG linkages:

SDG target 5.5: Ensure women's full and effective participation and equal opportunities for leadership at all levels of decision-making in political, economic and public life

+ SDG indicator 5.5.1: Proportion of seats held by women in (a) national parliaments and

### 3.7: Ratio of female to male labour force participation rate

**Data source:** World Bank

**Data series code:** "SL.TLF.CACT.FM.ZS"

**Indicator series name:** Ratio of female to male labour force participation rate (%) (modeled ILO estimate)

**Brief Description:** Labor force participation rate is the proportion of the population ages 15 and older that is economically active: all people who supply labor for the production of goods and services during a specified period. Ratio of female to male labor force participation rate is calculated by dividing female labor force participation rate by male labor force participation rate and multiplying by 100.

The labor force is the supply of labor available for producing goods and services in an economy. It includes people who are currently employed and people who are unemployed but seeking work as well as first-time job-seekers. Not everyone who works is included, however. Unpaid workers, family workers, and students are often omitted, and some countries do not count members of the armed forces. Labor force size tends to vary during the year as seasonal workers enter and leave.

**Metadata:**

<https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SL.TLF.CACT.FM.ZS>

**Data access:**

<https://data.worldbank.org/indicator/SL.TLF.CACT.FM.ZS>

**SDG linkages:**

SDG 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

+ No specific indicator

### 3.8: Ratio of female age of first marriage to male age of first marriage

**Data source:** UNDESA Population Division - World Marriage Database

**Data series code:** “AFMR”

**Indicator series name:** Ratio of female age of first marriage to male age of first marriage

**Brief Description:** The singulate mean age at marriage (SMAM) is the mean age at first marriage among persons who ever marry by a certain age limit, usually before the age of 50 years. It measures the average number of years lived as single or “never married” by a hypothetical cohort of individuals for which the proportions never married at each age are the same as those observed at a moment in time for a given population.

Compilation issues: Additional sources of data include survey programmes, such as the Demographic and Health Surveys (DHS), Multiple Indicator Cluster Surveys (MICS), Reproductive Health Surveys (RHS), Pan-Arab Project for Child

Development Surveys (PAPCHILD), Pan-Arab Project for Family Health Survey (PAPFAM) as well as national surveys. Tabulations for several countries are based on census microdata samples provided by National Statistical Offices to the Integrated Public Use Microdata Series, International (Minnesota Population Center, 2019)<sup>6</sup> and survey microdata provided by the DHS and MICS. In order to maximize the availability of data on consensual unions and a wider range of age groups, two series of data on marital status are presented from the Demographic and Health Surveys and the Multiple Indicator Cluster Surveys: marital status data generated from the individual questionnaire for men and women of reproductive age that include consensual unions as a separate data point, and marital status data generated from the household questionnaire that pertain to age groups up to 75 years and over. Data generated from the household questionnaire are differentiated by designating them as such in the Data Source field.

In the case of DHS data, as available, two values are presented from the same survey: 1) The survey raw data (data as they were collected, without any structural changes). 2) Survey data made available via STATcompiler, a Web-based tool provided by DHS that allows users to build customized tables for DHS countries based on hundreds of indicators. The original country raw data is converted into a standardized format allowing easy comparison among countries or different DHS phases in the same country. The values from both sources may differ slightly due to the conversion of the raw data into the standardized format provided by the STATcompiler.

*Note:* Next release is expected in the first half of 2024.

The recommended priority on data sources is: DHS -> MICS -> Census

**Metadata:**

[https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesapd\\_2019\\_world\\_marriage\\_metadata.pdf](https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesapd_2019_world_marriage_metadata.pdf)

**Data access:**

<https://population.un.org/MarriageData/Index.html#/home>

**SDG linkages:**

SDG target 5.3: Eliminate all harmful practices, such as child, early and forced marriage and female genital mutilation

+ SDG indicator 5.3.1: Proportion of women aged 20–24 years who were married or in a

**3.9: Ratio of the share of wage and salaried workers in women’s employment to men’s employment**

**Data source:** World Bank

**Data series code:** “FMEMP”

+ Female: "SL.EMP.WORK.FE.ZS", Male: "SL.EMP.WORK.MA.ZS"

**Indicator series name:** Ratio of female to male employment-to-population ratio (%) (modeled ILO estimate)

+ Wage and salaried workers, female (% of female employment) (modeled ILO estimate)

+ Wage and salaried workers, male (% of female employment) (modeled ILO estimate)

**Brief Description:** Wage and salaried workers (employees) are those workers who hold the type of jobs defined as “paid employment jobs,” where the incumbents hold explicit (written or oral) or implicit employment contracts that give them a basic remuneration that is not directly dependent upon the revenue of the unit for which they work.

The methodology to estimate global and regional wage trends was developed by the ILO for the previous editions of the Global Wage Report (GWR) in collaboration between technical departments and the Department of Statistics, following four peer reviews conducted by five independent experts. The appendix of the GWR describes the methodology adopted as a result of this process.

Global estimates on wages are not in the ILOEST database or other database on ILOSTAT.

Ratio of the share of wage and salaried workers in women’s employment to men’s employment is calculated by dividing the the share of wage and salaried workers among women, by the corresponding share for men.

**Metadata:**

<https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SL.EMP.WORK.FE.ZS>    <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SL.EMP.WORK.MA.ZS>

**Data access:**

<https://data.worldbank.org/indicator/SL.EMP.WORK.FE.ZS>    <https://data.worldbank.org/indicator/SL.EMP.WORK.MA.ZS>

**SDG linkages:**

SDG 10: Reduce inequality within and among countries

+ No specific indicator

**3.10: Share of women’s service employment to total employment, raised to the power of the inverse of the Palma ratio**

**Data sources:** World Bank, World Income Inequality Database (WIID)

**Data series code:** “MCARE”

+ Employment in services: "SL.SRV.EMPL.FE.ZS"  
 + female labor force: "SL.TLF.TOTL.FE.ZS"  
 + Palma ratio: "PALMA"

**Indicator series name:** Share of women's service employment to total employment, raised to the power of the inverse of the Palma ratio

+ Employment in services, female (% of female employment) (modeled ILO estimate)  
 + Labor force, female (% of total labor force)  
 + Palma ratio

**Brief Description:** Employment is defined as persons of working age who were engaged in any activity to produce goods or provide services for pay or profit, whether at work during the reference period or not at work due to temporary absence from a job, or to working-time arrangement. The services sector consists of wholesale and retail trade and restaurants and hotels; transport, storage, and communications; financing, insurance, real estate, and business services; and community, social, and personal services, in accordance with divisions 6-9 (ISIC 2) or categories G-Q (ISIC 3) or categories G-U (ISIC 4).

Female labor force as a percentage of the total show the extent to which women are act.

The Palma ratio is the ratio of the richest 10% of the population's share of gross national income divided by the poorest 40% share (Palma, 2014).

The last indicator represents the extent and quality of the market care sector. Because women's service sector work tends to be concentrated in the caring professions, we use women's services employment as a share of total employment (men plus women) to proxy for the extent of the market care sector. To get at the question of quality, we effectively discount this measure by the extent of income inequality in the economy (by raising it to the power of the inverse of the income inequality measure) on the argument that the more inequality, the lower the quality (and pay) of care sector work. The so-called 'Palma' ratio, which is the share of income going to the richest 10% of the population divided by the share of income going to the poorest 40% of the population, is used for income inequality (Palma, 2014). The higher the value of this variable, the more gender egalitarian (GE) the distribution of social reproduction.

*Note:* Using the Palma ratio emphasizes the importance of what is happening in the tails as opposed to characterizing income distribution overall.

*Note:* What is GE?

Both women and men contribute time and money to social reproduction, either directly or indirectly through taxes and charitable contributions. How they split these responsibilities is correlated with whether the high- or low-road regime

prevails. The more that women and men share the time and financial costs of care, the more likely that increases in female labour force participation and market output will increase the production of human capacities by more than it will cut into profits—and therefore, the more likely the high-road case will be. This is the gender egalitarian (GE) case.

**Metadata:**

<https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SL.SRV.EMPL.FE.ZS> <https://databank.worldbank.org/metadataglossary/world-development-indicators/series/SL.TLF.TOTL.FE.ZS>  
<https://www.wider.unu.edu/sites/default/files/WIID/PDF/WIID-User-Guide-30JUN2022.pdf>

**Data access:**

<https://data.worldbank.org/indicator/SL.SRV.EMPL.FE.ZS> <https://data.worldbank.org/indicator/SL.TLF.TOTL.FE.ZS> <http://www.wider.unu.edu/project/world-income-inequality-database-wiid> <http://www4.wider.unu.edu/>

**SDG linkages:**

SDG 10: Reduce inequality within and among countries SDG 5: Achieve gender equality and empower all women and girls

+ No specific indicator

## Pillar 4

The fourth pillar is dedicated to **environmental** indicators.

It includes 4 indicators, namely:

### 4.1: Carbon dioxide (CO<sub>2</sub>) emissions per value added

**Data source:** Global SDG Indicators Database

**Data series code:** “EN\_ATM\_CO2GDP”

**Indicator series name:** Carbon dioxide emissions per unit of GDP PPP (kilogrammes of CO<sub>2</sub> per constant 2017 United States dollars)

**Brief Description:** The indicator CO<sub>2</sub> emissions per unit of value added represents the amount of emissions from fuel combustion produced by an economic activity, per unit of economic output. When computed for the whole economy, it combines effects of the average carbon intensity of the energy mix (linked to the shares of the various fossil fuels in the total); of the structure of an economy (linked to the relative weight of more or less energy-intensive sectors); of the average efficiency in the use of energy.

CO<sub>2</sub> emissions per unit of value added is an indicator computed as ratio between CO<sub>2</sub> emissions from fuel combustion and the value added of associated economic activities. The total intensity of the economy is defined as the ratio of total CO<sub>2</sub> emissions from fuel combustion and per unit of GDP.

**Metadata:** <https://unstats.un.org/sdgs/metadata/files/Metadata-09-04-01.pdf>

**Data access:** <https://unstats.un.org/sdgs/dataportal/database>

**Disaggregation:** None

**SDG linkages:**

SDG target 9.4: By 2030, upgrade infrastructure and retrofit industries to make them sustainable, with increased resource-use efficiency and greater adoption of clean and environmentally sound technologies and industrial processes, with all countries taking action in accordance with their respective capabilities

+ SDG indicator 9.4.1: CO<sub>2</sub> emission per unit of value added

#### 4.2: Energy intensity

**Data source:** Global SDG Indicators Database

**Data series code:** “EG\_EGY\_PRIM”

**Indicator series name:** Energy intensity level of primary energy (Megajoules per constant 2017 purchasing power parity GDP)

**Brief Description:** Energy intensity is an indication of how much energy is used to produce one unit of economic output. It is an inverse proxy of the efficiency with which an economy is able to use energy to produce economic output. A lower ratio indicates that less energy is used to produce one unit of output, so decreasing trends indicate progress.

Energy intensity is only an imperfect proxy for energy efficiency. It can be affected by a number of factors, such as climate, structure of the economy, nature of economic activities etc. that are not necessarily linked to pure efficiency. For better assessment of energy efficiency progress, more disaggregated data are needed, such as those at the sectoral and end-use level.

This indicator is based on the development of comprehensive energy statistics across supply and demand for all energy sources – statistics used to produce the energy balance. Once the energy balance is developed, the indicator can be obtained by dividing total energy supply over GDP.

**Metadata:** <https://unstats.un.org/sdgs/metadata/files/Metadata-09-04-01.pdf>

**Data access:** <https://unstats.un.org/sdgs/dataportal/database>

**Disaggregation:** None

**SDG linkages:**

SDG target 7.3: By 2030, double the global rate of improvement in energy efficiency

+ SDG indicator 7.3.1: Energy intensity measured in terms of primary energy and GDP

#### 4.3: Water use efficiency

**Data source:** Global SDG Indicators Database

**Data series code:** "ER\_H2O\_WUEYST"

**Indicator series name:** Change in Water Use Efficiency over time (United States dollars per cubic meter)

**Brief Description:** The indicator "Change in Water Use Efficiency over Time" tracks the value added in US dollars per volume of water used in cubic meters, by a given economic activity over time. It considers water use by all economic activities, with a focus on agriculture, industry and the service sector. The indicator allows countries to assess to what extent their economic growth depends on the use of their water resources. Regional differences in climate and water availability must be considered in the interpretation of this indicator, in particular for agriculture.

The rationale behind this indicator consists in providing information on the efficiency of the economic and social usage of water resources, i.e., value added generated by the use of water in the main sectors of the economy, and distribution network losses.

**Metadata:** <https://unstats.un.org/sdgs/metadata/files/metadata-06-04-01.pdf>

**Data access:** <https://unstats.un.org/sdgs/dataportal/database>

**Disaggregation:** Activity: Total

+ X.Activity.=="TOTAL"

**SDG linkages:**

SDG target 6.4: By 2030, substantially increase water-use efficiency across all sectors and ensure sustainable withdrawals and supply of freshwater to address water scarcity and substantially reduce the number of people suffering from water scarcity

+ SDG indicator 6.4.1: Change in water-use efficiency over time

#### 4.4: Terrestrial protected areas (% total land area)

**Data source:** Global SDG Indicators Database

**Data series code:** “ER\_PTD\_TERR”

**Indicator series name:** Terrestrial biodiversity area as % total protected areas

**Brief Description:** The indicator “Proportion of important sites for terrestrial and freshwater biodiversity that are covered by protected areas, by ecosystem type” shows temporal trends in the mean percentage of each important site for terrestrial and freshwater biodiversity (i.e., those that contribute significantly to the global persistence of biodiversity) that is covered by designated protected areas and Other Effective Area-based Conservation Measures (OECMs). Protected areas, as defined by the IUCN, are clearly defined geographical spaces, recognized, dedicated and managed, through legal or other effective means, to achieve the long-term conservation of nature with associated ecosystem services and cultural values.

This indicator is calculated from data derived from a spatial overlap between digital polygons for protected areas from the World Database on Protected Areas (UNEP-WCMC & IUCN 2020), digital polygons for Other Effective Area-based Conservation Measures from the World Database on OECMs and digital polygons for terrestrial and freshwater Key Biodiversity Areas (from the World Database of Key Biodiversity Areas, including Important Bird and Biodiversity Areas, Alliance for Zero Extinction sites, and other Key Biodiversity Areas).

**Metadata:** <https://unstats.un.org/sdgs/metadata/files/Metadata-15-01-02.pdf>

**Data access:** <https://unstats.un.org/sdgs/dataportal/database>

**Disaggregation:** Average proportion of Freshwater Key Biodiversity Areas (KBAs) covered by protected areas [15.1.2] with code ER\_PTD\_TERR

#### SDG linkages:

SDG target 15.1: By 2020, ensure the conservation, restoration and sustainable use of terrestrial and inland freshwater ecosystems and their services, in particular forests, wetlands, mountains and drylands, in line with obligations under international agreements

+ SDG indicator 15.1.2: Proportion of important sites for terrestrial and freshwater b

#### Other indicators used

Following data series are downloaded for the process of IGI compilation:

- Overview indicators: Population, GDP

- Indicators used as proxies for interpolation:
- Indicators to fill missing data series: Gini index, CO<sub>2</sub> emissions per unit of GDP

### Population

**Data source:** UN DESA Population Division - World Population Prospects

**Data series code:** “POP”

**Indicator series name:** Population

**Metadata:** <https://population.un.org/wpp/>

**Data access:** <https://population.un.org/wpp/>

### GDP

**Data source:** National Accounts - Analysis of Main Aggregates (AMA)

**Data series code:** “GDP\_USD2015”

**Indicator series name:** GDP

**Brief description:** GDP at constant 2015 US dollars

**Metadata:** <https://unstats.un.org/unsd/snaama/>

**Data access:** <https://unstats.un.org/unsd/snaama/>

### Electricity net consumption (EIA)

**Data source:** U.S. Energy Information Administration (EIA)

**Data series code:** None - data used to fill the gaps

**Indicator series name:** Electricity net consumption (billion kWh)

**Brief description:** The database includes values on electricity net consumption (billion kWh). Per capita values can be calculated by using population from WPP.

Total electric power consumption = total net electricity generation + electricity imports - electricity exports - electricity transmission and distribution losses.

Data are reported as net consumption, not gross consumption. Net consumption excludes the energy consumed by the generating units.

**Metadata:** <https://www.eia.gov/electricity/>

**Data access:** <https://www.eia.gov/international/data/world/electricity/electricity-consumption?pd=2&p=0000002&u=0&f=A&v=mapbubble&a=->



**Brief description:** Gaps in CO<sub>2</sub> emissions per unit of GDP can be filled by using data from the EDGAR. These data show a very high correlation with the principal data source.

In this report fossil CO<sub>2</sub> emissions include sources from fossil fuel use (combustion, flaring), industrial processes (cement, steel, chemicals and urea) and product use. Please note that in the analysis presented in this report no short cycle carbon CO<sub>2</sub> is included in any sector.

EDGARv7.0 includes IEA CO<sub>2</sub> emissions (IEA Energy Balances, 2020) for 1990-2019 from fossil fuel consumption and they are extended (FT approach) up to 2021 using BP statistics keeping the same sectoral breakdown. Updates up to 2021 for cement, lime, ammonia and ferroalloys production are based on USGS statistics, urea production and consumption are based on IFA statistics, associated gas used from flaring from GGFR/NOAA, steel production from world steel and cement clinker production from UNFCCC (2022).

CO<sub>2</sub> emissions and removals from the LULUCF sector by EU27 countries and world macro-regions covering the time period 1990-2021 are presented. The LULUCF sector includes: Forest Land (Rossi et al., in prep.), Deforestation, Drainage of Organic Soils, Other Land use and their conversion (Grassi et al., 2022).

LULUCF emissions for the year 2021 are a copy of the previous year values. Compared to previous LULUCF versions, wildfire emissions are also included using the GWIS estimates (Global Wildfire Information System).

**Metadata:** [https://edgar.jrc.ec.europa.eu/report\\_2022](https://edgar.jrc.ec.europa.eu/report_2022)

**Data access:** [https://edgar.jrc.ec.europa.eu/booklet/EDGARv7.0\\_FT2021\\_fossil\\_CO2\\_booklet\\_2022.xlsx](https://edgar.jrc.ec.europa.eu/booklet/EDGARv7.0_FT2021_fossil_CO2_booklet_2022.xlsx)

*Note:* New report and data series released on 8 September 2023. Data links and names can change.

### **Water productivity, total (constant 2015 US\$ GDP per cubic meter of total freshwater withdrawal)**

**Data source:** World Bank

**Data series code:** “ER.GDP.FWTL.M3.KD”

**Indicator series name:** Water productivity, total (constant 2015 US\$ GDP per cubic meter of total freshwater withdrawal)

**Brief description:** Water productivity is an indication only of the efficiency by which each country uses its water resources. Given the different economic structure of each country, these indicators should be used carefully, taking into account a country’s sectorial activities and natural resource endowments. GDP data are from World Bank’s national accounts files. Water withdrawals can

exceed 100 percent of total renewable resources where extraction from non-renewable aquifers or desalination plants is considerable or where water reuse is significant. Withdrawals for agriculture and industry are total withdrawals for irrigation and livestock production and for direct industrial use (including for cooling thermoelectric plants).

The indicator “Efficiency of water use (water productivity)” is unavailable for some countries. However, a similar indicator “Water productivity, total (constant 2015 US\$ GDP per cubic meter of total freshwater withdrawal)” is available for almost all countries. Thus, missing data for the preferred indicator were populated using a regression model based on the highly correlated (0.815) variable on water productivity as the auxiliary variable.

**Metadata:** [https://databank.worldbank.org/metadataglossary/sustainable-development-goals-\(sdgs\)/series/ER.GDP.FWTL.M3.KD](https://databank.worldbank.org/metadataglossary/sustainable-development-goals-(sdgs)/series/ER.GDP.FWTL.M3.KD)

**Data access:** <https://data.worldbank.org/indicator/ER.GDP.FWTL.M3.KD>

## Chapter 3

# Data interpolation, extrapolation and imputation

As a general rule of thumb, the higher a country's level of development, the more complete the data available. The PCA requirement for complete data is therefore only fulfilled for a limited number of relevant variables or countries. If all incomplete variables and country data led to exclusion, it would result in a very limited measure of inclusive growth and, most likely, a bias towards developed countries. It would also introduce complications in the application of PCA, as this requires a relatively large sample to produce stable results. Consequently, imputation is required to maximize the inclusion of the available source indicators and economies.

Former editions of the IGI aimed to construct the index for one year only. Data for indicators were collected and data gaps filled by various methods (linear interpolation, using external data as proxies, imputation). To minimize data loss, all countries with complete information for at least one pillar's indicators were included in the calculations. Consequently, analyses were based on data from 167, 131 and 90 countries for the first, the second and the third pillars, respectively. However, the final index was computed only for the 86 states with complete data available for all three pillars.

Since 2023, data are collected for all years starting 2000 until the most recent available data point. Obtaining full panel data set for all years requires following steps:

- Data interpolation
- Data extrapolation
- Data imputation

## 3.1 Data interpolation

### Filling data gaps from external sources

There are indicators highly correlated with another indicator or data from alternative sources. Using such proxies could help to fill data gaps especially in cases where no data point is available for the entire series. In that case, missing data are populated by a regression model based on the auxiliary variable. There are two indicators where this approach is considered.

- **IGI 1.5: Electric power consumption, kWh/person**

Gaps in electric power consumption per capita can be filled by using data from the U.S. Energy Information Administration website. The database includes values on electricity net consumption (billion kWh). Per capita values can be calculated by using population from WPP. These data show a very high correlation (0.9971) with the principal data source.

- **IGI 2.4: People using safely managed drinking water services (% of population)**

The indicator “People using safely managed drinking water services” is seen as a more relevant indicator for medium-to-high income countries, but this variable is unavailable for some countries. However, a similar indicator “People using at least basic drinking water services” is available for almost all countries. Thus, missing data for the preferred indicator were populated using a regression model based on the highly correlated (0.8213) variable on basic water services as the auxiliary variable.

- **IGI 3.1: Income concentration ratio (Gini index), units**

Gaps in Gini index can be filled by using data from the World Income Inequality Database.

- **IGI 4.1: CO<sub>2</sub> emissions (kg per PPP USD of GDP)**

Gaps in CO<sub>2</sub> emissions per unit of GDP can be filled by using data from the EDGAR. These data show a very high correlation (0.9971) with the principal data source.

- **IGI 4.3: Efficiency of water use (water productivity)**

The indicator “Efficiency of water use (water productivity)” is unavailable for some countries. However, a similar indicator “Water productivity, total (constant 2015 US\$ GDP per cubic meter of total freshwater withdrawal)” is available for almost all countries. Thus, missing data for the preferred indicator were populated using a regression model based on the highly correlated (0.815) variable on water productivity as the auxiliary variable.

## Remaining issues

Although data availability improved significantly by filling gaps from external sources, there are still some indicators identified as having lower data coverage.

- **IGI 2.1: Logistics Performance Index**
  - The index is available for 139 economies. There are gaps because the index is not applied for all years but the overall coverage remains good.
- **IGI 2.7: Access to financial services**
  - The indicator *Proportion of adults (15 years and older) with an account at a financial institution or mobile-money-service provider, by sex (% of adults aged 15 years and older)* is drawn from Gallup survey data covering almost 128,000 people in 123 economies. According to metadata, four rounds of data collection were completed, for years: 2011, 2014, 2017, and 2022. However, there are only two years available in the SDG Global Database - 2017 and 2021.
- **IGI 3.8: Ratio of female age of first marriage to male age of first marriage**
  - World Marriage data include data from various data sources, often including duplicates. Manual cleaning was conducted to prepare the final data set. However, data gaps still persist. For new IGI editions, a better approach on decision making has to be developed to reduce the data processing burden by manual selection.
  - It is not expected that the average age of first marriage would significantly change over the time. Data gaps for countries with no data can be filled by calculating regional averages, for instance for M49 sub-regions.

In addition, poverty indicators such as poverty headcount ratio, Gini index or Palma ratio are not available for some countries. For instance, data for Singapore, which is very high in the ranking, are fully imputed for the indicator IGI 3.2 Poverty headcount ratio.

In 2023, there are 12 countries with only one data series missing to be included in the index. These countries are potential candidates to be included in the index in future editions when data gaps are filled.

- FB\_BNK\_ACCSS (IGI 2.7) is missing for Fiji, Gabon, Timor-Leste
- MAR\_AGE\_MAL (IGI 3.8) is missing for Jamaica, North Macedonia
- SI.POV.LMIC (IGI 3.2) is missing for Cambodia, New Zealand, Singapore
- NE.EXP.GNFS.ZS (IGI 1.6) is missing for Liberia, Malawi
- NY.ADJ.NNTY.PC.KD (IGI 1.2) is missing for Malta
- LP.LPI.OVRL.XQ (IGI 2.1) is missing for Eswatini

### Linear interpolation

In this interpolation step, for each country, series with sufficient observations (more than 8 observations) and with gaps within the observations (with a maximum gap size of 5) are linearly interpolated. Series which do not satisfy these conditions are imputed at a later stage together with series with no observations at all.

## 3.2 Data extrapolation

Data extrapolation (or forecasting) refers to the prediction of the future values of a series after the last observed data, and until the last year for which we would like prediction (hereafter end year). If the end year is the current year, this extrapolation would generally be referred to as “nowcasting”. Forecasting values further than a couple of years ahead of available data is to be considered meaningless.

Since most series exhibit some trend, using double-exponential smoothing is likely to improve the extrapolations over simply repeating the last observed value. Moreover, this method needs fewer parameters to be estimated than ARIMA. Double-exponential smoothing is known to perform well in a wide range of forecasting tasks including for time series with relatively few observations, as is often our case. In view of the above, we found that the simplicity of the double-exponential smoothing and its better empirical performance were strong arguments for the change. For the reasons already discussed, it was also decided to perform extrapolation right after the interpolation step.

**Accounting for the COVID-19 pandemic:** Some indicators are thought to have been impacted by the COVID-19 pandemic. These indicators can be identified and extrapolated, similarly as for the Productive Capacities Index (PCI).

## 3.3 Data imputation

The next step includes data imputation by using non-parametric methods. The missing data are imputed by using the missForest imputation method

[Stekhoven and Buehlmann, 2012] which is frequently used in the case of mixed-type data. The imputation was conducted by using the R package `missForest` [Stekhoven, 2022].

Series which have too many missing values need to be imputed for the PCA to run: the threshold for imputation (as opposed to interpolation and extrapolation) is set to gaps of 5 or more observations (and can be changed in the settings); this of course includes cases where series have no observation at all.

### MissForest method

MissForest is a nonparametric imputation method that can accommodate almost any kind of data, and is provided in the package of the same name [Stekhoven, 2022]. It can cope with mixed-type of variables, nonlinear relations, complex interactions and high dimensionality. It only requires the observation (i.e. the rows of the data frame supplied to the function) to be pairwise independent. The algorithm is based on random forests [Breiman, 2001], which are powerful predictive models which, for the sake of brevity, can be compared to very flexible nonlinear regression models.

For each variable `missForest` fits a random forest on the observed part and then predicts the missing part. The algorithm continues to repeat these two steps until a stopping criterion is met or the user specified maximum of iterations is reached. For further details see [Stekhoven and Buehlmann, 2012].

In more detail: The decision tree starts with using the most discriminative variable to create new branches. Those branches are then split in the same logic until the decision tree has fine grained nodes. The average of the values in those nodes become the estimated value. The default for continuous variables is that the node only contains 1 observation. Missing values are thus estimated as an observed value with a similar value on all other chosen variables. This is just one decision tree, this estimate is averaged out over all 100 trees. And then this process is redone in the next iteration until estimates of missing values hardly change.

### Lessons learned

MissForest is a powerful imputation method, easy to be applied. Nevertheless, it reveals certain limitations when applied on country data with significant differences in performance. A good example is GDP, which could be very low for LDCs but very high for high-income economies. The imputed values could be biased towards those extreme outlying observations.

More detailed notes resulting from using the method on the PCI data and input values for the SDG costing exercise:

- **If countries are imputed separately**

- It should not be possible to impute completely missing data for a variable - there is nothing to train the model on.
- If there is one observed value - this will be the estimate for all missing values.
- The range of estimated missing values is the same as for the observed values.
- The positive is that the imputations stay in the range for the country. But, for example, if there is only data up until 2010 - there is not room for growth beyond the 2010 level.

- **If all countries of a group/region are imputed together.**

- There is more data to model on.
- The range of possible imputed values is the range of that variable for the group.
- Implicitly, if  $y$  is the imputed variable and  $x$  is the set of other variables - the relationship between  $y$  and each  $x$  is taken into account.
- One  $x$  will be *Year* and supposedly there is at least one observation in the class for each year. Trends over time will be accounted for and implicitly used for extrapolation. *Year* can come in at any point in the decision tree. In some trees it could be the first variable to split the tree, in others it will represent the time trend among observations with similar values on one or more  $x$ .
- One  $x$  is the country itself. If that is a discriminating variable (i.e. if the values of the country is distinct from other countries), it will be used in the model and the estimate will look more like the other values of the country. In the current setup, all observations from a country (2000 - 2022) are, a priori, treated as independent observations. Countries with fewer observations will influence the estimates less than others.

- **If all countries, regardless of group, are imputed together**

- The logic is extended. There is even more data to model on.
- The range of possible imputed values is now the same as the observed values of the variable in all countries.
- One  $x$  will be the group, in fact another  $x$  will be the split such as a group of developed/developing. So if the groups actually differs on the imputed variable, this will be accounted for.

## Chapter 4

# Data transformation

Some of the source indices had to be inverted to make the interpretation of the PCA results more straightforward. Values were inverted by deducting the original value from the theoretical maximum value of that variable, or the observed maximum value in the absence of a clear theoretical maximum. After this transformation, for all source indices, a higher value signified a better result. This approach greatly simplified the interpretation of results.

Inversion is especially relevant for the Equality pillar. For instance, the income concentration ratio (IGI 3.1), based on the Gini coefficient, ranges between 0 and 100, where 0 signifies total equality and 100 total inequality. The Gini was inverted (using the theoretical maximum of 100) so that, counter intuitively, a higher value signified less income inequality i.e. now 100 represents maximum equality. The poverty headcount ratio (IGI 3.2), which also ranges between 0 and 100 per cent of population, was also inverted (again using the theoretical maximum of 100). In the inverted form, a higher value signifies less poverty i.e. less people with an income below US\$5.50 per day. Indicators in the inequality pillar were also transformed into parity ratios, where necessary i.e. to compare differences between two groups of population.

For example, under-five mortality rate (deaths per 1,000 live births) (IGI 2.3) was inverted by using the observed maximum among the countries. In the inverted form, a higher value signified more live births. CO<sub>2</sub> emissions per unit of GDP (IGI 4.1) is another example, where in an inverted form, the higher value signified fewer CO<sub>2</sub> emissions per unit of GDP. The same is applied for energy intensity (IGI 4.2).

A symmetric transformation was undertaken for indicators of secondary school enrollment (IGI 3.3), ratio of female to male employment-to-population ratio (IGI 3.4), ratio of youth to adult employment-to-population ratio (IGI 3.5), gender parity in the number of seats held by women and men in national parliaments (IGI 3.6), ratio of female to male labour force participation rate (IGI

3.7), ratio of female age of first marriage to male age of first marriage (IGI 3.8) and ratio of the share of wage and salaried workers in women's to men's employment (IGI 3.9). After the transformation, the same rate for female and male (or youth and adults) equals one - the best possible value. The proportion of seats held by women in national parliaments (percentage of total number of seats) was transformed so that a 50-50 parity in Parliament (the optimal solution) equates to the highest possible value (value = 1) and all other distributions or solutions are less than one.

## 4.1 By indicator

- **IGI 2.3 (SH\_DYN\_MORT): Under-five mortality rate (deaths per 1,000 live births)**
  - The indicator was inverted by using the observed maximum among the countries.
- **IGI 3.1 (SI.POV.GINI): Gini index**
  - For instance, the income concentration ratio, based on the Gini coefficient, ranges between 0 and 100, where 0 signifies total equality and 100 total inequality. The Gini was inverted (using the theoretical maximum of 100) so that, counter intuitively, a higher value signified less income inequality i.e. now 100 represents maximum equality.
- **IGI 3.2 (SI.POV.LMIC): Poverty headcount ratio at \$3.65 a day (2017 PPP) (% of population)**
  - The poverty headcount ratio, which also ranges between 0 and 100 per cent of population, was also inverted (again using the theoretical maximum of 100). In the inverted form, a higher value signifies less poverty i.e. less people with an income below US\$5.50 per day.
- **IGI 3.3 (SE.ENR.SECO.FM.ZS): School enrolment, secondary (gross), gender parity index (GPI)**
  - A symmetric transformation was undertaken.
  - After the transformation, the same rate for female and male equals one - the best possible value.
- **IGI 3.4 (SL.EMP.TOTL.SP): Ratio of female to male employment-to-population ratio (%) (modeled ILO estimate)**
  - First, a ratio is calculated by using indicator disaggregated by sex.
  - A symmetric transformation was undertaken.

- After the transformation, the same rate for female and male equals one - the best possible value.
  - **IGI 3.5 (EMP\_2WAP): Ratio of youth to adult employment-to-population ratio (modeled ILO estimate)**
    - First, a ratio is calculated by using indicator disaggregated by sex.
    - A symmetric transformation was undertaken.
    - After the transformation, the same rate for youth and adult equals one - the best possible value.
  - **IGI 3.6 (SG\_GEN\_PARL\_PAR): Gender parity in the number of seats held by women and men in national parliaments**
    - The proportion of seats held by women in national parliaments (percentage of total number of seats) was transformed so that a 50-50 parity in Parliament (the optimal solution) equates to the highest possible value (value = 1) and all other distributions or solutions are less than one.
    - Using the indicator “Proportion of seats held by women in national parliaments (% of total number of seats)” we calculate the remaining share of men.
    - A symmetric transformation was undertaken.
    - After the transformation, the same rate for female and male equals one - the best possible value.
- $$SG\_GEN\_PARL\_PAR = \frac{SG\_GEN\_PARL}{100 - (SG\_GEN\_PARL)}$$
- **IGI 3.7 (SL.TLF.CACT.FM.ZS): Ratio of female to male labour force participation rate (%) (modeled ILO estimate)**
    - A symmetric transformation was undertaken.
    - After the transformation, the same rate for female and male equals one - the best possible value.
  - **IGI 3.8 (AFMR): Ratio of female age of first marriage to male age of first marriage**
    - A symmetric transformation was undertaken.
    - After the transformation, the same rate for female and male equals one - the best possible value.
  - **IGI 3.9 (FMEMP): Ratio of the share of wage and salaried workers in women’s to men’s employment**
    - A symmetric transformation was undertaken.

- After the transformation, the same rate for female and male equals one - the best possible value.
- **IGI 3.10 (MCARE): Employment in services, female (% of female employment) (modeled ILO estimate), raised to the power of the inverse of the Palma ratio**
  - The indicator is calculated as ratio of “Employment in services, female (% of female employment) (modeled ILO estimate)” to “Labor force, female (% of total labor force)” multiplied by the inversed “Palma ratio”

$$MCARE = \left( SL.SRV.EMPL.FE.ZS * SL.TLF.TOTL.FE.ZS \right)^{\left( \frac{1}{PALMA} \right)}$$

- **IGI 4.1 (EN\_ATM\_CO2GDP): Carbon dioxide emissions per unit of GDP PPP (kilogrammes of CO2 per constant 2017 United States dollars)**
  - CO2 emissions per unit of GDP is another example, where in an inverted form, the higher value signified fewer CO2 emissions per unit of GDP.
- **IGI 4.2 (EG\_EGY\_PRIM): Energy intensity level of primary energy (megajoules per constant 2017 purchasing power parity GDP)**
  - Energy intensity is another example, where in an inverted form, the higher value signified less energy per unit of GDP.

## 4.2 By transformation

- **Special calculations**
  - Ratio: 3.4, 3.5, 3.8, 3.9

$$ratio = female/male$$

- Gender parity: 3.6

$$SG\_GEN\_PARL\_PAR = \frac{SG\_GEN\_PARL}{100 - (SG\_GEN\_PARL)}$$

- Adjusted for the inversed Palma ratio: 3.10

$$MCARE = \left( SL.SRV.EMPL.FE.ZS * SL.TLF.TOTL.FE.ZS \right)^{\left( \frac{1}{PALMA} \right)}$$

- **Inverse transformation with 100**

- Used for indicators: 3.1, 3.2

$$inv(value) = 100 - value$$

- **Inverse transformation with max**

- Used for indicators: 2.3, 4.1, 4.2

$$inv(value) = max(value) - value$$

- **Inverse symmetric transformation**

- Used for indicators: 3.4, 3.5, 3.6, 3.8, 3.9

$$inv(value) = 1 - abs(value - 1)$$

- Used for indicators: 3.3, 3.7

$$inv(value) = 1 - abs(value/100 - 1)$$



## Chapter 5

# Principal Component Analysis

Compiling composite indicators requires the use of multivariate analysis in order to investigate the overall structure of the indicators, assess the suitability of the data and guide methodological choices e.g. weighting and aggregation of the indicator components. Composite indicators calculated in an arbitrary manner where little attention is given to the interrelationships between the source variables may lead to misleading results that are unhelpful for policy purposes.

Principal component analysis (PCA) is one of the most widely used techniques for multivariate analysis. First introduced by Pearson [Pearson, 1901] and developed independently by Hotelling [Hotelling, 1933], PCA can be used to reveal interrelationships among a set of variables. This is done by transforming potentially correlated variables into a set of uncorrelated variables using their covariance matrix or its standardized form, the correlation matrix. This enables the variability within the underlying information contained in  $N$  variables to be identified. As such, PCA can be used to emphasize patterns among multivariate data.

Through an orthogonal linear transformation, PCA calculates the projection of the original data into a new set of  $N$  coordinates, known as principal components. This new space has some interesting characteristics, including that its coordinates are mutually orthogonal and that they are ordered in decreasing order according to the amount of information contained from the original variables. Therefore, the first principal component (*PC1*) accounts for the largest amount of the total variability in the set of  $N$  original variables. The second vector (*PC2*), orthogonal to the first, accounts for the largest amount of the remaining variability in the original variables. Each succeeding PC is linearly uncorrelated to the others and accounts for the largest amount of the remaining variability [Jolliffe, 2002]. By selecting the first  $N$  principal components, the

number of dimensions to be included in an analysis can be reduced (from  $N$  to  $n$ ) while retaining as much of the information in the original variables as possible, a process called dimensionality reduction. The ranking of the principal components in order of their significance (based on the proportion of total variability that they capture) is denoted by the eigenvalues associated with each PC.

The principal components associated with all variables identified as relevant for measuring inclusive growth are calculated. By retaining only those principal components whose eigenvalue is greater than 1 and explained variance greater than 10%, a smaller number of independent indices of inclusive growth can be generated.

Before undertaking PCA, it was necessary to convert the original variables into standard comparable units as different scales could affect the application of the method. Therefore, each variable was standardized to have a mean of zero and a standard deviation of one. The PCA was then applied to the completed, standardized data. The results presented here correspond to PCA output following an orthogonal rotation (varimax). The rotation increases the specificity of each component, leading to a simpler structure and easier interpretation of the results.

## Chapter 6

# Weights

Weights can have a significant effect on the overall composite indicator and country rankings. Attributing weights to source indicators or pillars of a composite index can be done in several ways. They can be set to equality and such an unweighted index would imply that each pillar has equal importance for measuring inclusive growth. Weights can also be assigned based on expert assessment, policy priorities or theoretical factors. Finally, they can be determined using statistical techniques. Irrespective of which method is used, weights reflect value judgements. Opinions on which approach is the best vary. It should also be noted that even ‘unweighted’ indicators are implicitly weighted.

An equal weight could be assigned to each source indicator i.e. to each of the 27 indicators. The risk is that some source indicators are correlated and may be driven by common factors or trends. Combining correlated variables may potentially introduce an element of ‘double counting’ into the overall index. The equal weights approach also implies that all variables have the same importance to the index, and that a strong statistical, empirical and theoretical basis is missing. Equal weights can be tempting where knowledge regarding the causal relationships between source indicators is insufficient or consensus on the theoretical approach or policy priorities is lacking. Furthermore, as the source indicators are grouped into pillars that are subsequently aggregated, applying equal weights to the indicators implies an unequal weighting of the pillars as each pillar contains a different number of indicators.

For the purposes of this study, pillar weights were determined by quantifying the interconnections of source indicators using PCA. The advantage of this approach is that risk of subjective bias associated with experts’ views or other non-statistical methods is reduced. PCA is one of the most widely recognized statistical techniques used to calculate index component weights. As described above, this methodology transforms correlated source indicators to form new variables referred to as principal components, which account for decreasing shares of the

original variance of data. Each principal component was attributed a weight corresponding to the share of variance explained.

## 6.1 How many principal components to keep?

Based on the OECD Handbook on Constructing Composite Indicators [Nardo et al., 2005], the following rule is applied.

The code is automated to retain number of components based on following criteria:

- SS loadings > 1 &
- Proportion Var > .1

The SS loadings (sum of squares loadings) represent the amount of variance explained by each principal component, and the Proportion Var is:

$$\frac{SS\_loadings}{sum(SS\_loadings)}$$

## 6.2 Final IGI

The year of interest **2021** is chosen to calculate individual weights based on the PCA. At this stage, only IGI for 2021 is produced.

The next goal is to calculate the IGI for all years (2000-2021) by using the 2021 weights. This step has not been implemented yet.

### Recoding of countries

There are five countries which were recoded to correspond with the UNCTAD-stat target economies.

- France: 250 -> 251 (France including French Guiana, Guadeloupe, Martinique, Mayotte)
- Norway: 578 -> 579 (Including Svalbard and Jan Mayen)
- Switzerland: 756 -> 757 (Switzerland including Liechtenstein)
- United Kingdom: 826 -> 926 (United Kingdom including Channel Islands and Isle of Man)
- United States of America: 840 -> 842 (United States of America including Puerto Rico)

Data are collected from various international data sources and country coverage. Detailed metadata should be provided disaggregated by country, indicator and data source.

# Bibliography

- Nour Barnat, Steve MacFeely, Fernando Cantu, Anu Peltola, Anastasia Khazhgerieva, Andrey Pantelev, and Nikolay Ryabtsev. Compiling an inclusive growth index: Methodological challenges, considerations and conclusions. *Statistical Journal of the IAOS*, (Preprint):1–21.
- Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.
- Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:498–520, 1933. URL <https://api.semanticscholar.org/CorpusID:144828484>.
- Ian Jolliffe. *Principal component analysis*. Springer Verlag, New York, 2002.
- Michela Nardo, Michaela Saisana, Andrea Saltelli, Stefano Tarantola, Anders Hoffman, and Enrico Giovannini. Handbook on constructing composite indicators. 2005. doi: <https://doi.org/https://doi.org/10.1787/533411815016>. URL <https://www.oecd-ilibrary.org/content/paper/533411815016>.
- Karl Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.
- Daniel J. Stekhoven. *missForest: Nonparametric Missing Value Imputation using Random Forest*, 2022. R package version 1.5.
- Daniel J. Stekhoven and Peter Buehlmann. Missforest - non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1):112–118, 2012.
- UNCTAD. Sdg pulse 2022: Stark contrasts in inclusive growth – progress towards equal opportunities needed everywhere. Available at <https://sdgpulse.unctad.org/inclusive-growth/>, 2022.
- UNCTAD and EEC. Inclusive growth of the eurasian economic union member states: Assessments and opportunities. Available at [http://www.eurasiancommission.org/ru/act/integr\\_i\\_makroec/dep\\_makroec\\_pol/Documents/Inclusive\\_growth\\_in\\_EAEU\\_Member.pdf](http://www.eurasiancommission.org/ru/act/integr_i_makroec/dep_makroec_pol/Documents/Inclusive_growth_in_EAEU_Member.pdf), 2019.