European Commission

JRC TECHNICAL REPORTS

JRC Statistical audit of the Equal Measures 2030 SDG Gender Index 2022

Caperna, G.
Kovacic, M.
Papadimitriou, E.

2022

Joint Research Centre

EUR 31016 EN

Report completed in March 2022.

Contact information
European Commission
Joint Research Centre
Directorate for Competences
Monitoring, Indicators and Impact Evaluation Unit
Competence Centre on Composite Indicators and Scoreboards

Competence Centre on Composite Indicators and Scoreboards
https://ec.europa.eu/knowledge4policy/composite-indicators

EU Science Hub
https://ec.europa.eu/jrc

JRC128800

EUR 31016 EN

PDF ISBN 978-92-76-49351-8 ISSN 1831-9424 doi:10.2760/993717

Luxembourg: Publications Office of the European Union, 2022

How to cite this report: Caperna G., Kovacic M., and Papadimitriou E., *JRC Statistical Audit of the Equal Measures 2030 SDG Gender Index 2022*, EUR 31016 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-49351-8, doi:10.2760/993717, JRC128800.

Contents

1 Introduction 2
2 Conceptual Framework 3
3 Data quality and availability 5
3.1 Management of missing data 5
3.2 Treatment of outliers 5
3.3 Normalisation 6
4 Statistical coherence 8
5 Impact of modelling assumptions on the Index results 19
6 Conclusions 25
List of tables 27
List of Figures 27

Abstract

The SDG Gender Index of Equal Measures 2030 is a multidimensional index that aligns to the normative framework of the Sustainable Development Goals (SDGs) and covers 56 gender issues in 14 of the 17 global goals in 144 countries across five regions of the world.

The European Commission's Competence Centre on Composite Indicators and Scoreboards (COIN) at the Joint Research Centre (JRC) in Ispra was invited by the developers to audit the SDG Gender Index for the second time, following the statistical audit performed in the first edition of the Index[1]. JRC-COIN aims to help ensure the transparency of the Index methodology and the reliability of its results. This JRC-COIN audit focuses on data quality, the statistical soundness of the multi-level structure of the Index, and the impact of key modelling assumptions on the results.

The analysis suggests that meaningful inferences can be drawn from the Index for the vast majority of countries. The SDG Gender Index (SDG-GI) is reliable and the framework has good statistical coherence. The Index's ranks are shown to be representative of a plurality of scenarios, and robust to some methodological changes and the pillar weights. Even though the SDG-GI has good statistical properties, JRC-COIN has made some suggestions for possible refinements.

[1] JRC Statistical audit of the Equal Measures 2030 SDG Gender Index, 2019

1 Introduction

Sound metrics and data are critical for turning the Sustainable Development Goals (SDGs) into practical tools for problem solving. Focusing on the gender perspective, it is important that advocates and decision-makers have the data they need and in the form they need them to guide their pursuit of the gender equality commitments in the Sustainable Development Goals.

The SDG Gender Index is developed by the Equal Measures 2030. The current is the second version of the Index with the first having been developed in 2019[2] after a pilot in six countries the year before (2018). The Index aims to help advocates to measure progress on the gender equality aspects of the SDGs and to use data, stories and evidence to hold policymakers accountable across countries. The 2022 EM2030 SDG Gender Index covers gender issues in 14 of the 17 global goals in 144 countries.

The statistical audit of this second version of the EM2030 SDG Gender Index was performed by the European Commission's Competence Centre on Composite Indicators and Scoreboards (COIN) at the Joint Research Centre (JRC) and was conducted upon invitation of the Index developers.

The SDG Gender Index framework is well constructed, and a lot of thought has clearly been put into it. However, conceptual and practical challenges are inevitable when trying to summarise with a single composite indicator the complexity of a multidimensional phenomenon. An analysis is needed to ensure and validate the statistical soundness of any composite index.

The analysis herein aims at shedding light on the transparency and reliability of the SDG Gender Index 2022 and thus enabling advocates and policymakers to derive more accurate and meaningful conclusions, and to potentially guide choices on priority setting and policy formulation.

In general, statistical soundness should be regarded as a necessary but insufficient condition for a sound index. This is because the correlations underpinning most of the statistical analyses carried out in this report need not necessarily represent the real influence of the individual indicators on the phenomenon being measured. The development of any index must therefore be nurtured by a dynamic, iterative dialogue between the principles of statistical and conceptual soundness.

The JRC assessment of the SDG Gender Index 2022 focuses on two main issues: the statistical coherence of the hierarchical structure of indicators and the impact of key modelling assumptions on the SDG Gender Index ranking.

It involves three steps: In the first step, the main descriptive statistics of the data are shown, and an initial data analysis is performed to detect missing values and potential outliers. In the next step, the statistical coherence is examined through a multilevel analysis of the correlations of the indicators and between the indicators and the Index. Finally, in the last step, the robustness of the Index and the impact of key modelling assumptions to the Index ranking are tested. In particular, the considered assumptions are the structure of the indicators' framework, the aggregation formula and the weighting scheme. The JRC analysis complements the reported country rankings for the SDG Gender Index with ranks intervals in order to better appreciate the robustness of these ranks to the computation methodology.

[2] "Harnessing the power of data for gender equality: Introducing the 2019 EM2030 SDG Gender Index"

2 Conceptual Framework

The conceptual framework of the SDG Gender Index (SDG-GI, henceforth) covers 14 out of the 17 Sustainable Development Goals (SDGs) signed on in 2015 by 193 UN member states, with the promise of contributing substantially to the realization of human rights. The authors aligned the Index framework with the SDGs since they were considered a good turning point for gender equality. The Index allows tracking progress in reaching gender equality by country, goal, and indicator as well as across time for most of the countries considered (135 out of 144).

Since the focus of the Index is on the gender aspect throughout the SDGs, the authors consider only those SDGs where the gender aspect is more evident and can be monitored, and arriving at a total of 14 out of 17 SDGs (**Table 1**). This choice is well justified, given its linkage with the 2030 Global Policy Agenda.

The Index draws on both official SDG indicators and complementary indicators, both gender-specific and not, but which may have a disproportionate effect on girls and women. The indicators look at the optimal targets and outcomes, as well as the enabling means (laws, policies, processes, and financing) that are needed to achieve gender equality as set against the Sustainable Development Goals.

The selection of indicators was based on five different criteria: (i) relevance to monitoring achievement of the SDGs; (ii) statistical adequacy (*i.e.*, they are valid and reliable measures); (iii) timeliness as they are up to date and published on a schedule; (iv) data quality since the data series represent the best available measure for a specific issue and derive from official national or international sources; and (v) coverage. Its design has been informed by consultations across the EM2030 partnership (including with national partners in the focus countries), the public, inputs from experts, and surveys with policymakers and gender advocates worldwide. For each of the 14 SDGs that are included in the EM2030 SDG Gender Index framework, three to five indicators were selected that capture the key gender dimensions of the goal, totalling 56 indicators.

Even though the present audit does not address the conceptual relevance of the indicators underpinning the framework, it is worth noting that the developers have used a parsimonious approach by selecting a rather balanced number of indicators across the SDGs.

Table 1: Conceptual framework of the SDG-GI

SDG	SDG ID	SDG Dimension	Indicator Name	Indicator Number
Poverty	SDG01	Social	Poverty	SB1.1
			Working women in poverty	SB1.2
			Women's land rights	SB1.3
			Women's perceptions of household income	SB1.4
Hunger	SDG02	Social	Undernourishment	SB2.1
			Food insecurity	SB2.2
			Anaemia	SB2.3
			Women's perceptions of food costs	SB2.4
Health	SDG03	Social	Maternal mortality	SB3.1
			Adolescent birth rate	SB3.2
			Family planning	SB3.3
			Women's perceptions of healthcare quality	SB3.4
Education	SDG04	Social	Over-age girls in primary school	SB4.1
			Expected years of schooling	SB4.2
			Not in education or employed	SB4.3
			Attained at least some secondary schooling	SB4.4
Equality	SDG05	Social	Early marriage	SB5.1
			Women's perceptions of informal support	SB5.2
			Legal grounds for abortion	SB5.3
			Women in parliament	SB5.4
			Women in ministerial roles	SB5.5
Water	SDG06	Environmental	Access to clean water	SB6.1
			Access to sanitation	SB6.2
			Women's perceptions of water quality	SB6.3
Energy	SDG07	Environmental	Access to electricity	SB7.1
			Use of clean fuels	SB7.2
			Women's perceptions of air quality	SB7.3
Workplace	SDG08	Socio-economic	Wage equality	SB8.1
			Women in vulnerable work	SB8.2
			Collective bargaining rights	SB8.3
			Laws on workplace equality	SB8.4
			Bank accounts	SB8.5
Industry	SDG09	Socio-economic	Use of digital banking	SB9.1
			Women's perceptions of road quality	SB9.2
			Women's access to internet	SB9.3
			Women in science and technology	SB9.4
Inequality	SDG10	Social	Income inequality (Palma)	SB10.1
			Personal autonomy and individual rights	SB10.2
			Migration treaty ratification	SB10.3
			Women can openly discuss politics	SB10.4
			Gender diversity laws	SB10.5
Cities	SDG11	Environmental	Women's perceptions of housing costs	SB11.1
			Co2 emissions	SB11.2
			Women's perceptions of public transport	SB11.3
			Share of slum population	SB11.4
Climate	SDG13	Environmental	Climate change leadership	SB13.1
			Women's perceptions of environmental policies	SB13.2
			Climate vulnerability	SB13.3
Peace and Justice	SDG16	Social	Women's access to justice	SB16.1
			Female victims of homicide rate	SB16.2
			Women's perceptions of public safety	SB16.3
			Government openness	SB16.4
Partnerships	SDG17	Socio-economic	Military expenditure	SB17.1
			Tax revenue	SB17.2
			Transparent national budgets	SB17.3
			Disaggregated statistics	SB17.4

Source: Developers of the Index and the European Commission's Joint Research Centre, 2022.

3 Data quality and availability

3.1 Management of missing data

The data used to construct the SDG Gender Index comes from diverse data sources, including nongovernmental organizations (NGOs), development agencies, civil society, and the private sector. When data were missing, imputation was performed using last year available or a secondary source with a similar definition of the indicator(s) but imputation based on a second source was only used in few cases. The imputation has been performed for both the baseline data (2015) and the most recent year available. Using the latest available data is common practice, especially when the data comes from national governments and is released with a delay with respect to the reference year.

JRC-COIN suggests to keep a high level of attention on indicators with a large number of missing values. Their role in a composite indicator may be unpredictable. An indicator with more than 20% of missing values could be a candidate for exclusion or - better yet – substitution, as soon as another one, that fits conceptually, becomes available. Nine indicators could be identified for having a relevant presence of missing values in this index.

The developers set thresholds for the inclusion of countries in the SDG-GI. Only countries with at least 13 goals out of 14 are included in the final composite indicator. In order to calculate a goal for a certain country there is the threshold of 75% data coverage. This means for instance, that if a country has two missing values in a goal with five indicators, it does not get a goal score for the specific goal. Thus, it is implied that the country has no data in that goal – even the three available data points are not getting used. As a consequence, this goal is excluded from the computation of the index – which means that its value is implicitly imputed as the geometric mean of the other goals for the specific country. If this happens for more than one goal, the country is excluded from the Index.

While this choice is completely legitimate and meaningful, JRC-COIN suggests considering an alternative approach to maximise the use of available data. That is, to maintain the developers' approach of not calculating the goal score when the aforementioned criteria are not fulfilled but at the same time use this subset of indicators - as a proxy of the goal – in the construction of the overall index.

It is correct to avoid the presentation of goal results if the internal coverage is poor, but in the computation of the overall index, a goal imputed on a poor subset of its indicators is closer to the real value than one imputed based on other goals. This modelling assumption is tested in Section 5, in the uncertainty analysis, and as a standalone sensitivity check.

3.2 Treatment of outliers

The audit also examined the presence of outliers that could potentially bias the effect of the indicators on the aggregates. JRC-COIN recommends an approach for outlier identification based on the values of skewness and kurtosis,[3] *i.e.*, when the variables simultaneously have an absolute skewness higher than 2.0 and a kurtosis higher than 3.5.

In order to treat outliers prior to normalization, the developers performed a modified winsorisation on seven indicators, namely SB2.2, SB3.1, SB4.1, SB11.2, SB16.2, SB17.1 and SB17.3. Instead of trimming the values, as suggested by the classic winsorisation approach, they defined a system meant to preserve the ratio between the values of the observations above the 97.5th percentile. It is not the aim of this document to describe the methodology in depth, but the rationale behind the modified winsorisation is clear. Nevertheless, JRC-COIN suggests explaining the methodology in a clear and transparent

[3] Groeneveld, R. A. and Meeden, G., 'Measuring Skewness and Kurtosis', *Journal of the Royal Statistical Society*, Series D, vol. 33, pp. 391–399, 1984.

way. This should increase the transparency of the methodology and avoid subjectivities. The effect of this treatment, along with some normalisation aspects, is tested in Section 5.

Table 2 offers summary statistics for the normalised indicators included in the SDG-GI. Only two indicators lie above the skewness and kurtosis threshold suggested by JRC-COIN as a rule of thumb (cells shaded in darker red in columns 9 and 10), namely SB11.2 ("Co2 emissions" – in Mt of Co2) and SB16.2 ("Female victims of homicide rate" measured as the incidence of female victims of intentional homicide per 100,000 population). The characteristics of the distribution of these two indicators is not surprising since the underlying phenomena are concentrated in a few countries and/or regions. For instance, the 14 biggest producers of Co2 generate more than 1700,00 Mt of Co2 every year, compared to an average of 88,4 Mt for the rest of the world. Similarly, 9 countries (out of 97 for which the indicator for female homicide is available) register, on average, 7.59 casualties per 100,000 population, compared to an average of 1.84 in the rest of the world. Considering the distribution of these two indicators after the modified winsorisation, JRC-COIN suggests testing other alternatives such as logarithmic transformation for the treatment of outliers in future editions of the Index.

3.3 Normalisation

The indicators are rescaled to a 0-100 scale, with higher values denoting better performances. This is a common and usually desired practice in the construction of composite indicators. The normalisation is done using all of the countries for which data are available. This practice is chosen to reflect more closely the global situation for each indicator. On the other hand, it implies that the computation of normalised values may also depend on countries that are not used in the Index.

The normalization formulas used are multiple and depend on the data. Most indicators are not normalised at all as they are already in the range of 0-100 or they are normalised with the MinMax formula and its counterpart for indicators with negative direction. Five indicators, measuring parity, are normalised to give a value of 100 to the countries with values close to the centre (47-53% for gender parity, and 0.97-1.03 for Inequality Palma Score). Some indicators with raw values ranging between 0 and 1 were simply rescaled to fit the range of 0-100. It is not the aim of this document to describe the single formula and goalpost used for each indicator.

As before, JRC-COIN suggests a detailed explanation of the methodology. The normalisation steps are intuitive and based on logical arguments; thus, a clear and transparent reporting of the choices made, will increase the transparency of the index. The effect of normalisation, together with the treatment of outliers, are tested in Section 5.

All rescaled variables are expressed as ascending variables (i.e. higher values denoting better performances). The rescaled data becomes easier to read and compare across all indicators in this way. Fixed boundaries (targets) are sometimes used instead of observed minimum and maximum values.

These are based on explicit/implicit SDG targets: as the developers point out, reaching the level of the highest performing country isn't enough if girls and women are still out of school or illiterate. However, the JRC recommends that the type and values of the target adopted for each indicator be made publicly disclosed and remain constant over editions to ensure that the results can be compared.

Table 2: Summary statistics of the indicators included in the SDG-GI

SDG ID	Indicator	Missing (N)	Missing (%)	Mean	Min	Max	Range	Skew	Kurtosis
SDG01	SB1.1	59	40.97	79.24	41.1	99.4	58.3	-1.15	1.01
	SB1.2	35	24.31	86.94	20.87	100	79.13	-1.62	1.73
	SB1.3	0	0	84.72	0	100	100	-1.3	0.44
	SB1.4	0	0	56.17	7	95	88	-0.12	-1.03
	SB2.1	10	6.94	90.54	53.4	97.5	44.1	-1.81	2.8
SDG02	SB2.2	43	29.86	87.97	28.38	99.82	71.44	-1.82	2.7
	SB2.3	2	1.39	73.66	38.2	93	54.8	-0.69	-0.35
	SB2.4	0	0	57.67	15	98	83	0.1	-1.32
SDG03	SB3.1	0	0	84.84	0.87	100	99.13	-1.87	3.09
	SB3.2	0	0	74.3	0	99.46	99.46	-0.99	0.45
	SB3.3	3	2.08	67.99	9.8	92.8	83	-0.74	-0.12
	SB3.4	0	0	59.07	20	96	76	0.08	-0.93
SDG04	SB4.1	26	18.06	91.5	45.95	100	54.05	-1.87	2.89
	SB4.2	1	0.69	54.36	11.59	100	88.41	-0.14	-0.29
	SB4.3	0	0	76.69	31.99	96.92	64.93	-0.8	0.55
	SB4.4	2	1.39	61.78	1.7	100	98.3	-0.37	-1.18
SDG05	SB5.1	4	2.78	87.93	38.1	100	61.9	-1.32	2.06
	SB5.2	1	0.69	80.17	37	98.1	61.1	-0.79	0.29
	SB5.3	1	0.69	61.54	0	100	100	-0.42	-1.41
	SB5.4	0	0	51.1	0.66	100	99.34	0.17	-0.51
	SB5.5	0	0	47.1	0	100	100	0.23	-0.97
SDG06	SB6.1	1	0.69	88.87	45.95	100	54.05	-1.4	0.73
	SB6.2	0	0	78.08	8.91	100	91.09	-1.16	-0.05
	SB6.3	0	0	71.07	28.01	99.01	71	-0.47	-0.12
SDG07	SB7.1	0	0	84.91	10.12	100	89.88	-1.56	1.05
	SB7.2	2	1.39	69.83	0.2	100	99.8	-0.82	-0.9
	SB7.3	0	0	74.51	38	95	57	-0.55	-0.2
SDG08	SB8.1	10	6.94	66.37	46.4	86	39.6	-0.17	-0.61
	SB8.2	0	0	88.44	32.71	99.99	67.28	-1.57	1.93
	SB8.3	41	28.47	81.09	0	100	100	-1.98	4.11
	SB8.4	0	0	73.87	0	100	100	-0.84	-0.12
	SB8.5	5	3.47	58	2.25	100	97.75	0.01	-1.28
SDG09	SB9.1	14	9.72	51.65	0.67	100	99.33	0.19	-1.34
	SB9.2	0	0	57.86	22	96	74	-0.1	-0.67
	SB9.3	32	22.22	65.26	1.49	100	98.51	-0.8	-0.66
	SB9.4	51	35.42	74.11	6.7	100	93.3	-0.82	0.35
SDG10	SB10.1	0	0	53.18	9.97	100	90.03	0.12	-1.14
	SB10.2	0	0	59.24	12.5	100	87.5	0	-1.14
	SB10.3	0	0	71.16	33.33	95.24	61.91	-0.62	-0.17
	SB10.4	1	0.69	71.69	13.58	98.9	85.32	-0.85	-0.28
	SB10.5	0	0	49.26	0	100	100	-0.19	-1.15
SDG11	SB11.1	0	0	67.22	33	98.01	65.01	0.01	-1.37
	SB11.2	4	2.78	88.48	0	100	100	-2.84	8.24
	SB11.3	0	0	58.12	11	94	83	-0.46	0.26
	SB11.4	41	28.47	66.37	11.6	100	88.4	-0.28	-0.8
SDG13	SB13.1	0	0	63.49	0	100	100	-0.69	-0.03
	SB13.2	1	0.69	51.76	12	92.01	80.01	0.29	-0.34
	SB13.3	0	0	58.19	32.35	75.09	42.74	-0.43	-0.62
SDG16	SB16.1	1	0.69	64.04	21.7	99.08	77.38	-0.13	-0.94
	SB16.2	47	32.64	82.85	15.71	98.69	82.98	-2.14	4.04
	SB16.3	2	1.39	56.52	8	95.01	87.01	0.05	-0.57
	SB16.4	1	0.69	46.26	6.92	86.5	79.58	0.34	-0.8
SDG17	SB17.1	2	1.39	79.2	8.25	100	91.75	-1.72	3.33
	SB17.2	8	5.56	48.14	2.83	99.76	96.93	0.16	-0.14
	SB17.3	37	25.69	18.66	0	100	100	1.56	1.26
	SB17.4	3	2.08	73.27	22.1	98.3	76.2	-0.89	1.42

Note: The cell with the percentage of missing values exceeding 10% are shaded in light red. The values of skewness and kurtosis exceeding the threshold are written in darker red.

Source: European Commission's Joint Research Centre, 2022.

4 Statistical coherence

The assessment of statistical coherence consists of a multi-level analysis of the correlations of variables, and a comparison of SDG Gender Index rankings with their constituent goals.

4.1 Correlation analysis

The statistical coherence of an index should be considered a necessary but insufficient condition for a sound index. Given that the statistical analysis is mostly based on correlations, the correspondence of every index to a real-world phenomenon needs to be critically addressed by developers and experts, because "correlations do not necessarily represent the real influence of the individual indicators on the phenomenon being measured" (OECD & JRC, 2008)[4]. This influence relies on the interplay between both conceptual and statistical soundness. The degree of coherence between the conceptual framework and the statistical structure of the data is an important factor for the reliability of an index.

Correlation analysis is used to assess the extent to which the observed data supports the conceptual framework. Within each level of the index, there should ideally be positive significant correlations (JRC-COIN suggest >0.30). This effectively ensures that the overall index scores adequately reflect the values of the underlying indicators.

The framework should avoid redundancy, which can be identified by very high correlations (>0.92). This is due to the fact that if two indicators are collinear, it may result in double counting (and thus over-weighting) of the same phenomenon.

Correlation analysis between indicators and aggregates

Table 3 shows the correlation coefficients between indicators within the same SDG. The majority of the correlations are significant and positive (>0.30). However, we highlight below a few cases that deserve some attention:

- Indicators SB7.3 ("Women's perceptions of air quality"), SB8.1 ("Wage equality"), SB11.2 ("Co2 emissions"), SB17.1 ("Military expenditure"), SB17.3 ("Transparent national budgets"), SB13.2 ("Women's perception of environmental policies") show shallow correlations with the other indicators in their respective SDGs. This could imply that these indicators do not fully cooperate with the others, which could result in a conflict in results and a reduction in the impact of the aggregate to which they belong in the following aggregations;

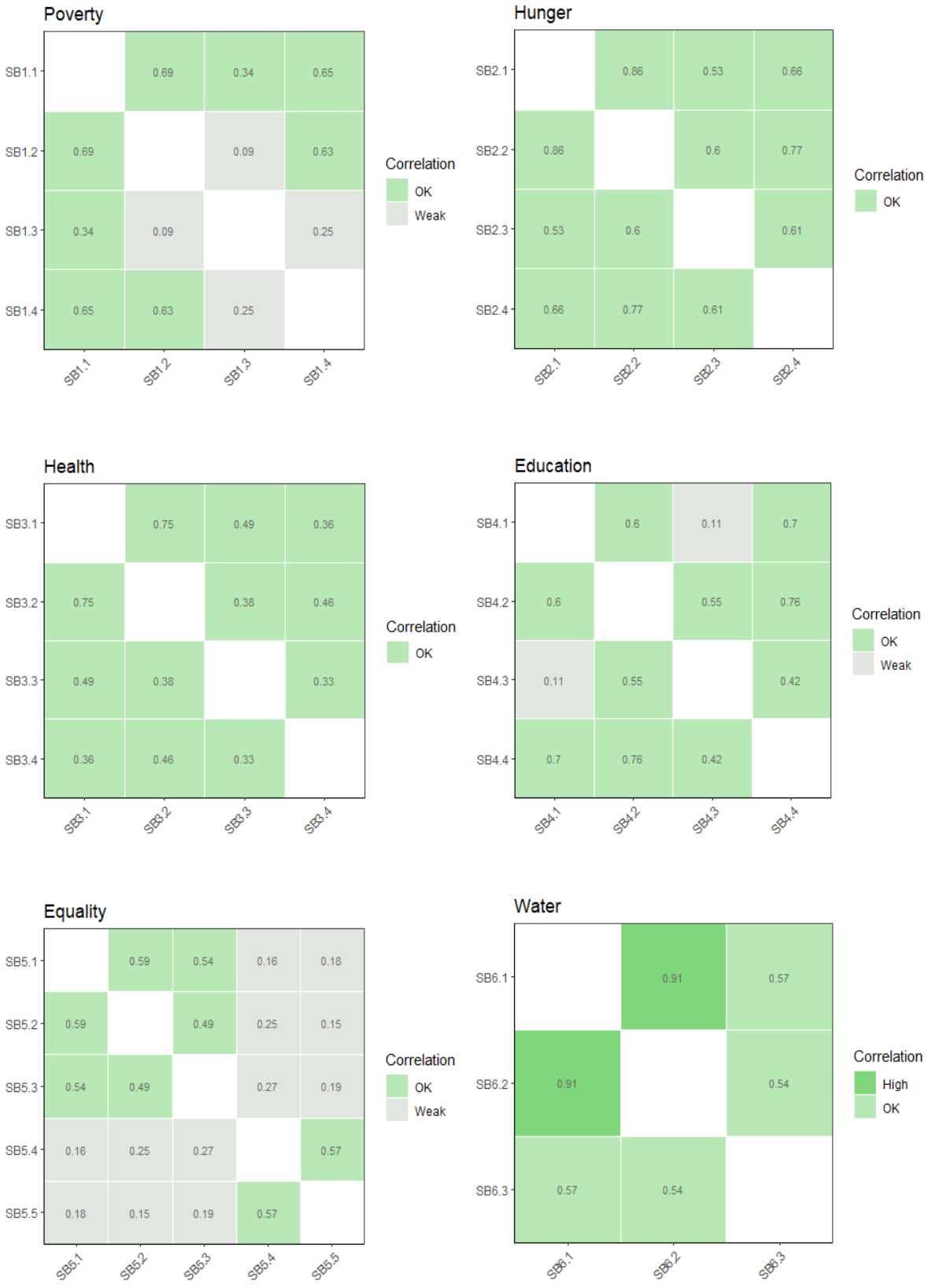
- As a result, the aforementioned indicators have low correlations with the Index (**Table 4**). Another indicator showing a low correlation with the Index, namely SB10.3 ("Migration treaty ratification"), however, results well correlated with the other indicators within its own SDG (except with the indicator SB10.1 – "Income inequality").

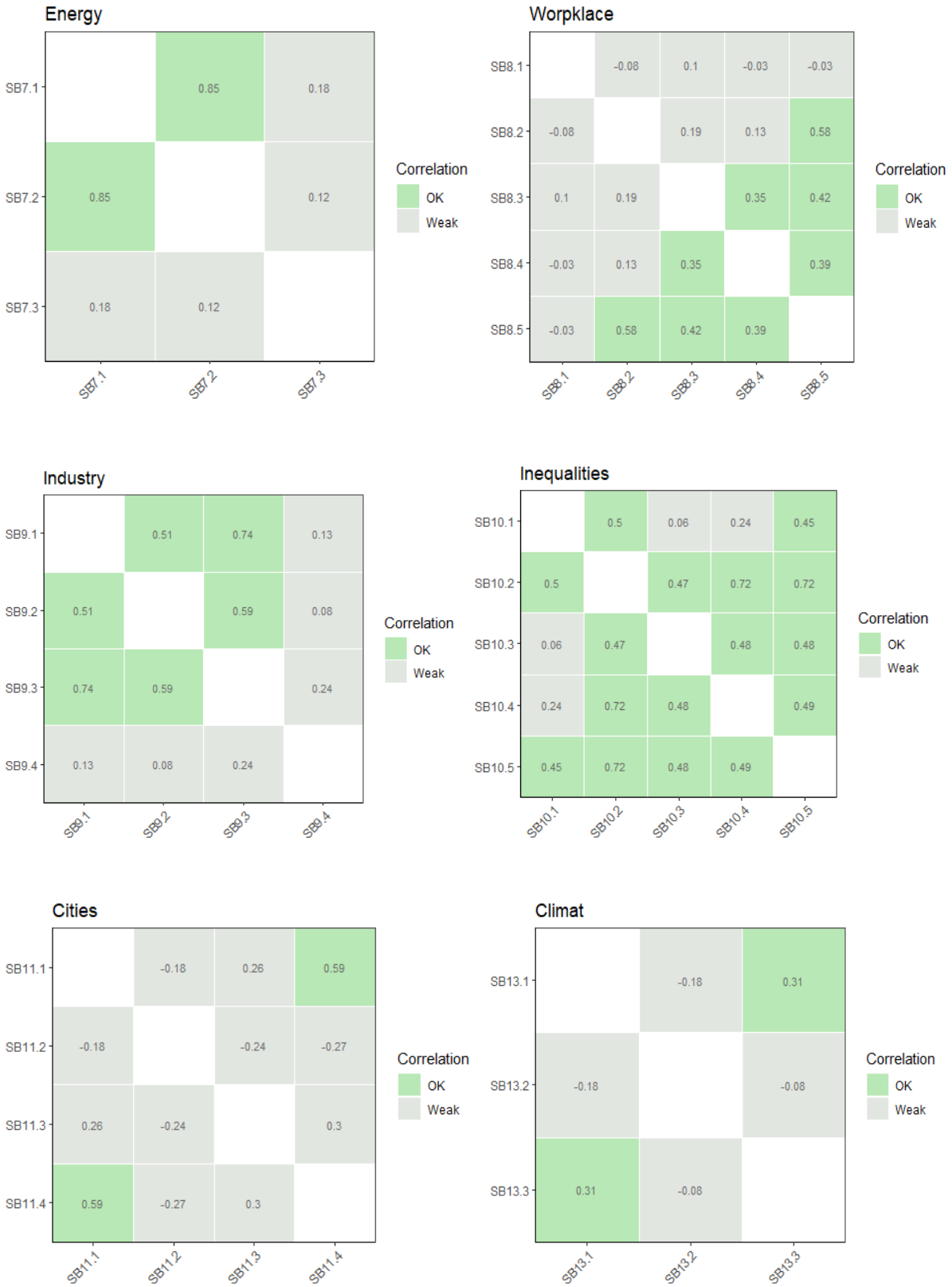
- Workplace, city, climate, and partnership goals have a poor internal correlation. This may make the value of the single goal less robust.

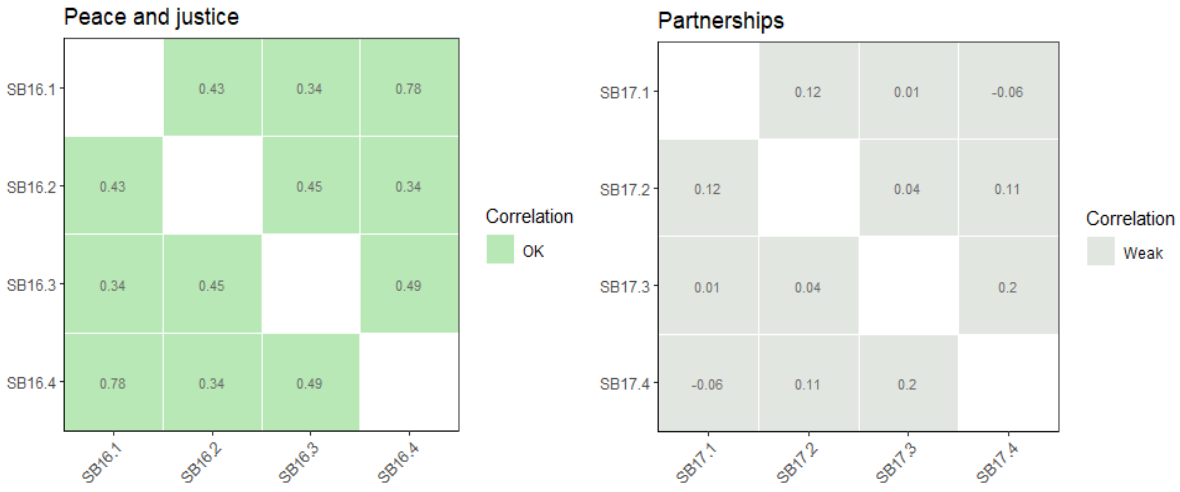
A suggestion would be to continue monitoring these specific indicators and their position in the framework for future index editions in order to check their behaviour and, if necessary, modify or substitute them. Particular attention is suggested to the indicators showing negative correlations.

[4] OECD/EC JRC (Organisation for Economic Co-operation and Development/European Commission, Joint Research Centre). 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. Paris: OECD.

Table 3: Correlations between indicators in the same SDG



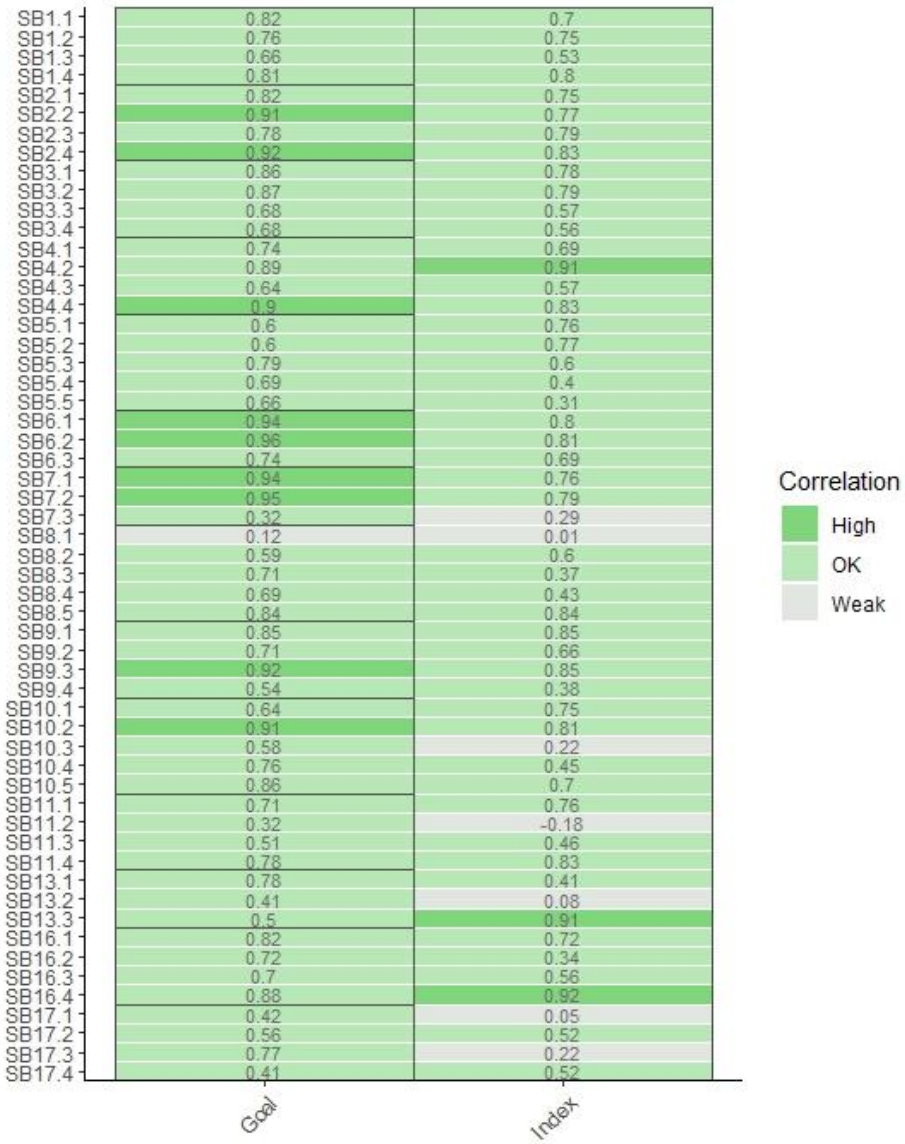

Energy
SB7.1
SB7.2
SB7.3
0.85
0.18
0.85
0.12
0.18
0.12
Correlation
OK
Weak
Worpklace
SB8.1
SB8.2
SB8.3
SB8.4
SB8.5
-0.08
0.1
-0.03
-0.03
-0.08
0.19
0.13
0.58
0.1
0.19
0.35
0.42
-0.03
0.13
0.35
0.39
-0.03
0.58
0.42
0.39
Correlation
OK
Weak
Industry
SB9.1
SB9.2
SB9.3
SB9.4
0.51
0.74
0.13
0.51
0.59
0.08
0.74
0.59
0.24
0.13
0.08
0.24
Correlation
OK
Weak
Inequalities
SB10.1
SB10.2
SB10.3
SB10.4
SB10.5
0.5
0.06
0.24
0.45
0.5
0.47
0.72
0.72
0.06
0.47
0.48
0.48
0.24
0.72
0.48
0.49
0.45
0.72
0.48
0.49
Correlation
OK
Weak
Cities
SB11.1
SB11.2
SB11.3
SB11.4
-0.18
0.26
0.59
-0.18
-0.24
-0.27
0.26
-0.24
0.3
0.59
-0.27
0.3
Correlation
OK
Weak
Climat
SB13.1
SB13.2
SB13.3
-0.18
0.31
-0.18
-0.08
0.31
-0.08
Correlation
OK
Weak



Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in light green. Correlations with low values (here <0.30) are written in grey. Correlations at risk of redundancy (here >0.91) are written in darker green. Correlations with meaningful (i.e., statistically significant) negative value (here -0.30) are highlighted in purple.

Source: European Commission's Joint Research Centre, 2022.

Table 4: Correlations between indicators and their aggregates



Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in light green. Correlations with low values (here <0.30) are written in grey. Correlations at risk of redundancy (here >0.91) are written in darker green. A negative although not statistically significant correlation (here -0.18) is written in white.

Source: European Commission's Joint Research Centre, 2022.

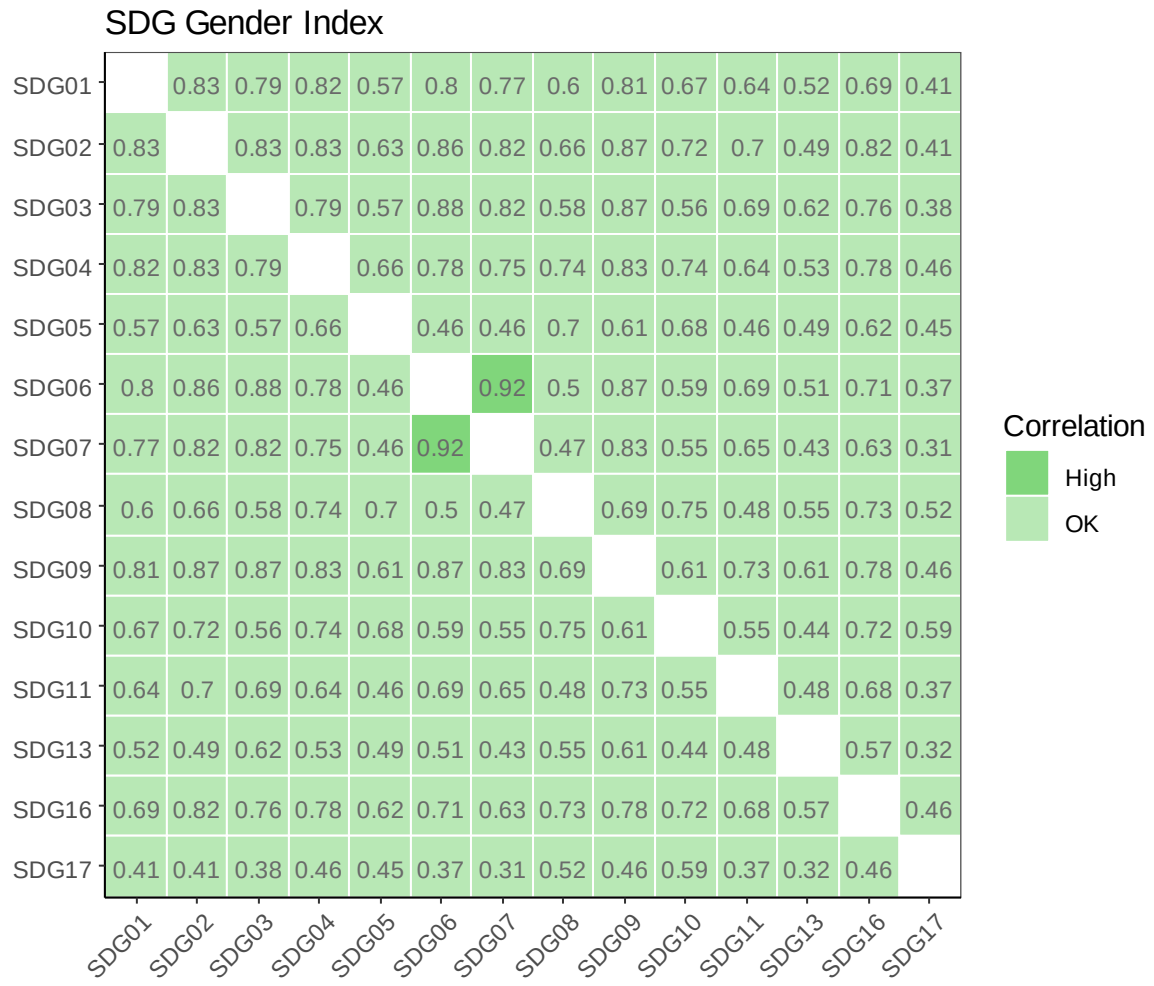
Correlation analysis between SDGs

Table 5 shows the correlation coefficients between SDGs. A few cases, which need specific attention, are identified in the points below:

- All correlations are significant and positive (>0.30). Only SDG17 ("Partnerships") shows marginal but not critical lack of correlation with multiple goals;

- SDG06 and SDG07 results are highly correlated, suggesting that there may be a risk of redundancy at the SDG level. This is mitigated at the index level (**Table 6**), with the two SDGs showing good positive correlations that do not exceed the 0.92 set threshold (0.87 and 0.82 respectively).

Table 5: Correlations between SDGs



Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in light green. Correlations with low values (here <=0.30) are written in grey. Correlations at risk of redundancy (here >=0.92) are written in darker green.

Source: European Commission's Joint Research Centre, 2022.

Correlation analysis between SDGs and Index

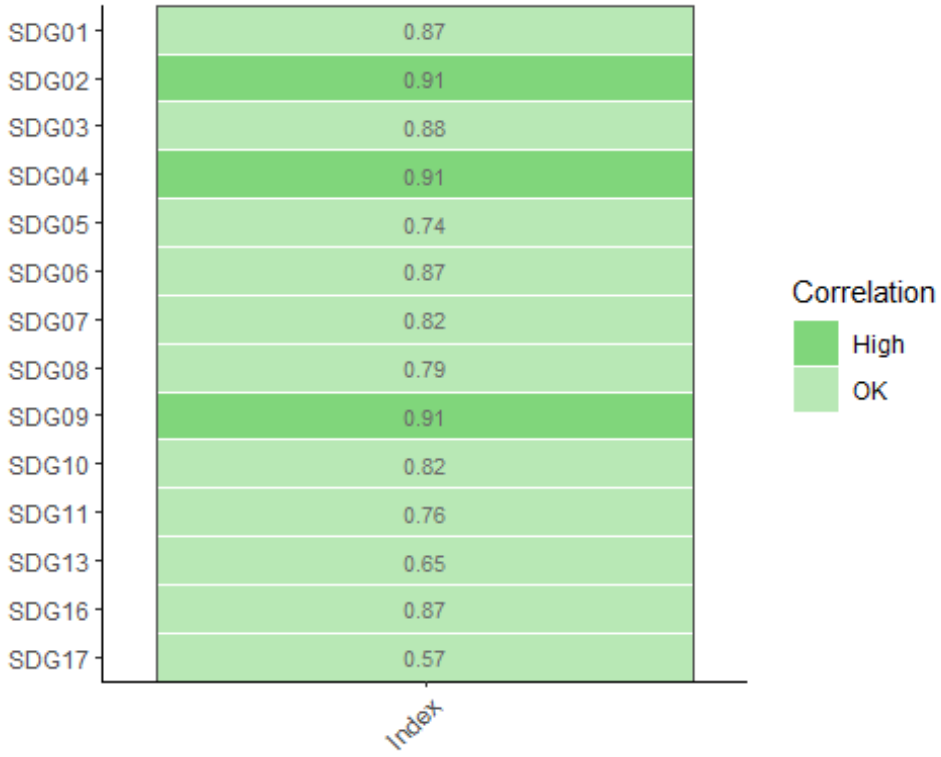
Table 6 shows the correlation between the aggregates (SDGs) and the Index. This is the most important level of aggregation because it represents the consistency of the general concept.

All the goals are well correlated with the Index. Only two goals (SDG13 and SDG17) show a relatively lower correlation with the Index (0.65 and 0.57 respectively) which was expected given the low correlation between the indicators within these two SDGs (**Table 3**).

Three goals (SDG02, SDG04 and SDG09) show a very high correlation with the Index (0.91). Some indicators within these SDGs were also highly correlated at the goal level (**Table 4**). While this is not a major concern, it should be taken into account in the future editions of the Index.

Other SDGs with very high correlations between their respective indicators (particularly SDG06 and SDG07), on the other hand, show good positive correlations at the index level (0.87 and 0.82 respectively).

Table 6: Correlations of SDGs with the Index



Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in light green. Correlations with low values (here <0.30) are written in grey. Correlations at risk of redundancy (here >=0.91) are written in darker green.

Source: European Commission's Joint Research Centre, 2022.

4.2 Principal components analysis of the SDG Gender Index

As a further step in the analysis of statistical coherence, principal components analysis (PCA), is used to confirm the presence of one single statistical dimension among the 14 goals that form the Index. Principal components analysis (PCA) explores the correlation of all the variables simultaneously, highlighting, if present, some common trends that describe a common concept among them. Technically, the expectation here is that there is only one principal component with an eigenvalue greater than 1, or explaining more than 70% of the variance. In practice, the achievement of these thresholds suggests the presence of a common, unidimensional phenomenon underlying the pillars/ goals.

In the previous (2019) edition of the Index, the aggregation structure was different; all indicators were directly aggregated to form the overall Index without the intermediate step of the 14 goals. The PCA analysis performed in the previous JRC audit[5], suggested the presence of several drivers among the indicators and the JRC recommendation for the future editions of the Index was to use the goals as an intermediate step towards the construction of the SDG Gender Index. Following the JRC suggestion, in the current version the developers used the goals as a further step of aggregation towards the overall Index.

[5] JRC Statistical audit of the Equal Measures 2030 SDG Gender Index, 2019, pages 10-12.

Principal components analysis at goals level shows the presence of two principal components (PC1 and PC2) with eigenvalues significantly higher than 1 (PC1 = 9.37, PC2 = 1.27) that explain about 76% of the total variation (**Table 7**). The first principal component (PC1) alone accounts for 66% of total variance explained, proving the fact that a large part of the variability of the goals' scores depends on a common concept.

Table 7: Eigenvalues and explained variance for the first ten principal components

PC	eigenvalue	percentage of variance	cumulative percentage of variance
PC1	**9.22**	**65.9**	**65.9**
PC2	**1.25**	**8.9**	**74.8**
PC3	0.71	5.1	79.8
PC4	0.58	4.2	84
PC5	0.44	3.1	87.1
PC6	0.36	2.5	89.7
PC7	0.32	2.3	92
PC8	0.26	1.9	93.8
PC9	0.22	1.6	95.4
PC10	0.17	1.2	96.6

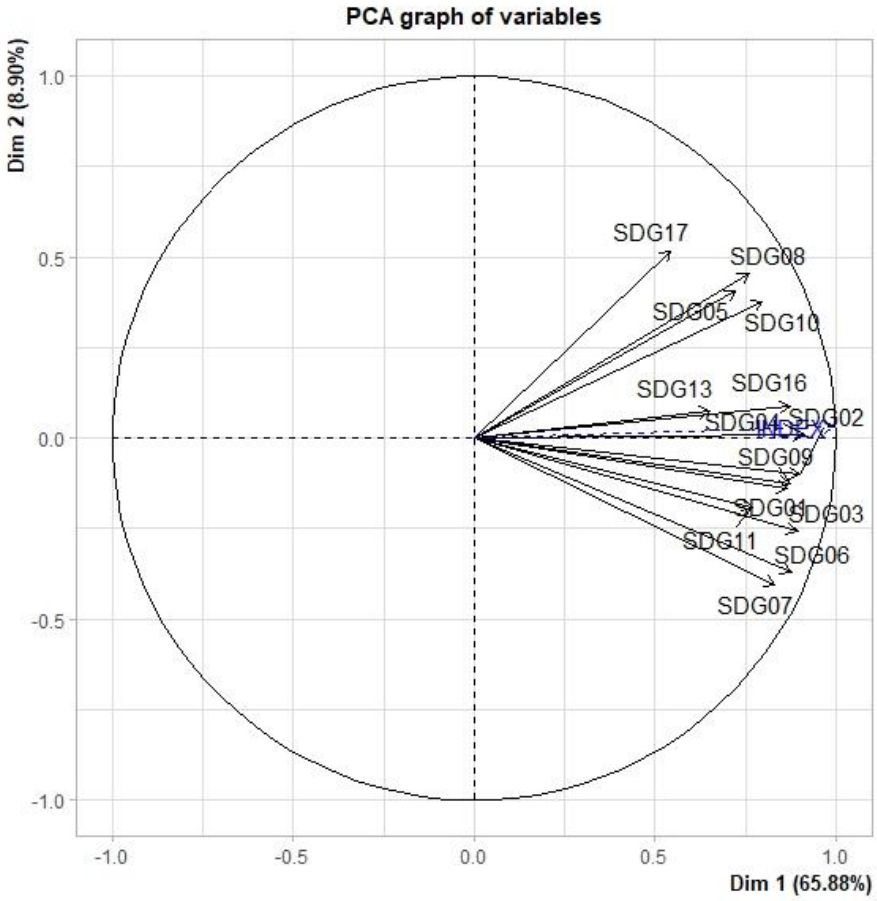
Source: European Commission's Joint Research Centre, 2022.

Figure 1 below, illustrates the projections of the goals onto the plane spanned by the first two principal components in a "factor map".

The correlation between each SDG goal and the principal component is given by the projection of the SDG vector onto the component axis. All goals correlate rather high with the first principal component: most correlations are above 0.70 apart from SDG13 and SDG17 that have lower but still good correlations (0.64 and 0.53 respectively). The second principal component is much less influential than the first and only accounts for 9% of the total variance.

The results of the PCA can overall be considered good, despite the possible weak suggestion of a bi-dimensional structure, since the first principal component accounts for a rather large part of the overall variance and the second one for a very small one. Many analyses have shown that sustainable development is multifaceted. Considering the variety and number of goals, the results show satisfying coherence, and the presence of an underlying single concept is clear.

Figure 1: Factor map of the 14 goals and comparison with the overall SDG-GI



Source: European Commission's Joint Research Centre, 2022.

Moreover, the results are in line with those obtained in the correlation analysis; all goals are correlating rather well with the overall index but not all of them have the same level of strong correlation.

4.3 Added value of the SDG Gender Index

Sometimes a high statistical association among the main components of an index can be due to the redundancy of information; thus, it is interesting to investigate whether the elements of the index contribute to the added value of the final score or they repeat similar information.

In the case of the SDG Gender Index, all 14 goals contribute to the overall score adding pieces of information, as the following **Table 8** confirms. At least 32% of the countries present in the Index show differences of more than 15 positions between the overall Index ranking and any individual goal rankings. In the cases of SDG17 and SDG13 this percentage is over 60% of the countries, confirming once again the results of the correlation analysis.

Table 8: Distribution of rank differences between pillar and SDG Gender Index rankings

SDG	More than 30 positions	16 - 30 positions	Over 15	6 -15 positions	5 and fewer positions	0 positions
SDG01	18.8	25	43.8	27.1	26.4	2.8
SDG02	9.7	22.9	32.6	37.5	25.7	4.2
SDG03	17.4	20.1	37.5	38.9	20.1	3.5
SDG04	8.3	25.7	34	42.4	22.9	0.7
SDG05	26.4	28.5	54.9	25.7	17.4	2.1
SDG06	11.8	24.3	36.1	37.5	26.4	0
SDG07	22.9	18.1	41	31.2	27.1	0.7
SDG08	17.4	22.2	39.6	34.7	22.2	3.5
SDG09	20.1	34.7	54.8	26.4	16	2.8
SDG10	19.4	20.1	39.5	30.6	25.7	4.2
SDG11	24.3	24.3	48.6	27.8	20.8	2.8
SDG13	39.6	22.2	61.8	20.8	16	1.4
SDG16	16.7	20.8	37.5	30.6	29.9	2.1
SDG17	42.4	25.7	68.1	20.1	9.7	2.1

Source: European Commission's Joint Research Centre, 2022.

4.4 Impact of the components of the SDG Gender Index

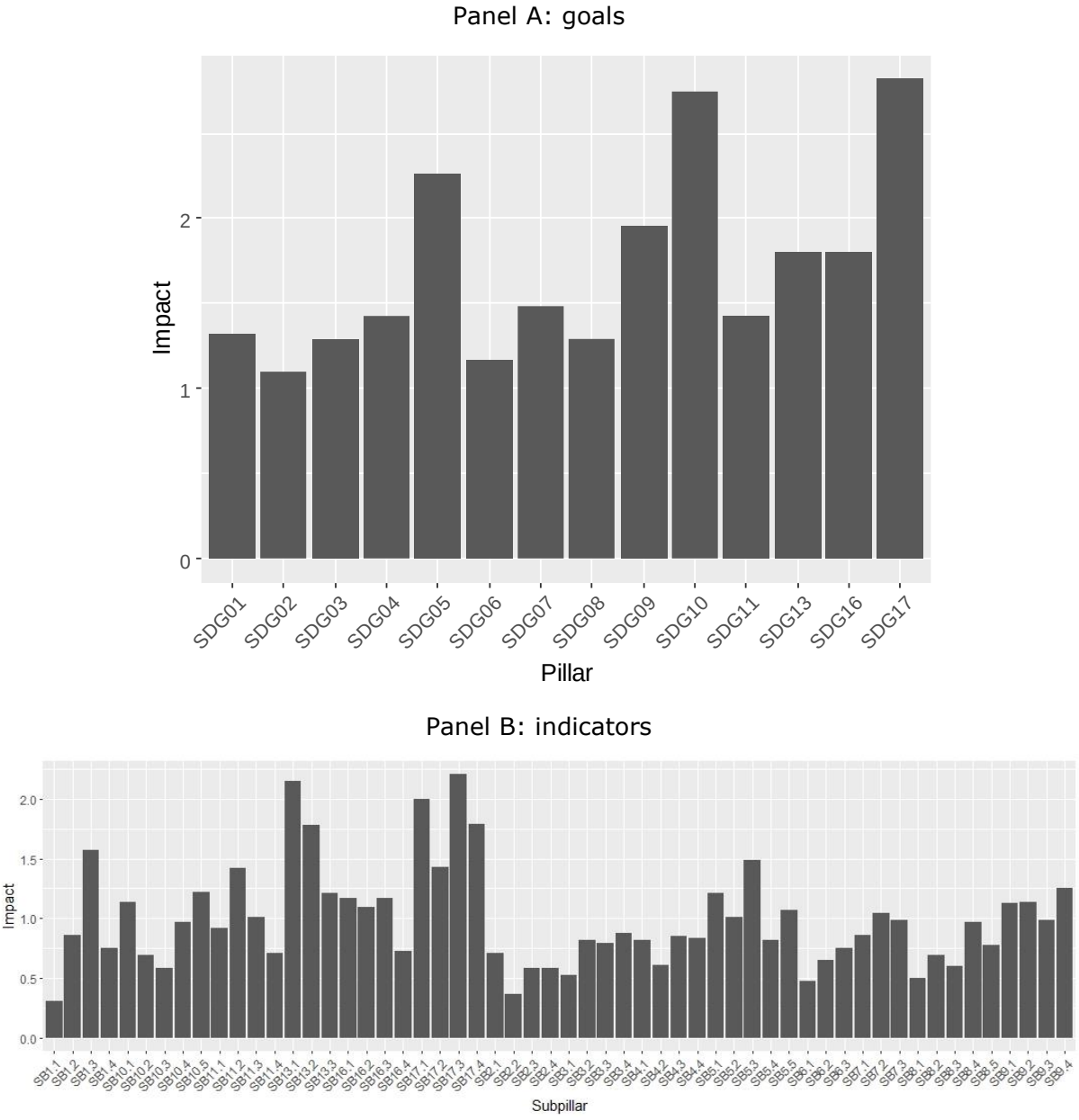
The study of the impact of the components (underlying indicator or aggregates) on the Index is conducted by observing alternative simulated rankings based on the elimination of one component at a time. One would typically expect to find some variability in rankings in such simulations. On the other hand, if no variability is present, the omitted component could be proven irrelevant, adding no significant valuable information to the Index. **Figure 2** outlines the average shifts in the SDG Gender Index country rankings when one element is omitted at a time.

Looking at what happens at goal level (**Figure 2,** panel A), the elimination of goals 17 ("Partnerships"), 10 ("Inequality") and 5 ("Gender Equality") have the greatest impact as they cause an average shift of the absolute rank of the countries of 2.81, 2.75 and 2.26 positions. With the elimination of goal 17, Kuwait and Saudi Arabia both gain 19 positions, while with the elimination of goal 5 Oman gains 19 positions. Similarly, when the goal 10 is omitted, Saudi Arabia and Qatar gain respectively 17 and 16 positions. This is a kind of result that may be useful to improve the understanding of the data inside the Index.

Among the single indicators, on the other hand, SB17.3 ("Transparent national budgets"), SB13.1 ("Climate change leadership"), and SB17.1 ("Military expenditure") have the most significant impacts on the rankings, with an average shift of the absolute rank of 2.20, 2.15 and 2.00 positions respectively (**Figure 2**, panel B). The elimination of one of these indicators would cause a relevant change in the rankings of countries.[6] The results of this analysis support the interpretation of the role of each element. There is no direct right or wrong when it comes to the impact of the single elements.

[6] Looking at the maximum rank shift observed when omitting an indicator, indicators 17.4 (Disaggregated statistics), 5.2 (Can count on the help), 13.1 (Climate change leadership) and 17.1(Military expenditure) show a significant impact on a country's rank causing changes of 18 positions (indicator 17.4) and 13 positions (indicators 5.2, 13.1 and 17.1) respectively.

Figure 2: Average shifts in SDG Gender Index country rankings when one element is omitted at a time (goals and indicators)



Source: European Commission's Joint Research Centre, 2022.

5 Impact of modelling assumptions on the Index results

A fundamental step in the statistical analysis of a composite indicator is to assess the effect of different modelling assumptions on the country rankings. Despite the efforts in the development process, there is an unavoidable subjectivity (or uncertainty) in the resulting choices. This subjectivity can be explored by comparing the results obtained under different – alternative – assumptions.

The literature on this topic[7] suggests assessing the robustness of the index by means of a Monte Carlo simulation and by applying a multi-modelling approach. This also assumes "error-free" data as possible errors have already been corrected in the preliminary stage of the index construction before the audit.

This Index analysed in this document, like most composite indicators, is the outcome of several choices. Among other things, these choices include: (i) the underlying theoretical framework; (ii) the indicators selected; (iii) the imputation of missing values; (iv) the weights assigned; and (v) the aggregation method. Some of these choices may be based on expert opinion or other consideration driven by statistical analysis or the need to ease communication or draw attention to specific issues.

This section aims to test the impact of varying some of these assumptions within a range of plausible alternatives in an uncertainty analysis. The objective is therefore to try to quantify the uncertainty in the ranks of SDG-GI, which can demonstrate the extent to which countries can be differentiated by their scores and ranks.

The modelling issues considered in the robustness assessment of the SDG-GI are (i) the aggregation formula; (ii) the exclusion/inclusion of goals according to data coverage; and (iii) the goals' weights. The following paragraphs deal with each of these in turn.

Aggregation formula. The developers of the SDG Gender Index opted for the geometric averaging of the fourteen goals, which implies some compensability, penalising all countries showing unbalanced performances. This approach can reward a country with generalised average results respect to countries with outstanding achievements in one goal accompanied by under-performing values in the others. To assess the impact of this choice, the JRC included in the analysis a comparison with the arithmetic mean, which allows, on the contrary, perfect compensability between outstanding performance and weak results. The comparison of the two aggregation approaches should be able to highlight countries with unbalanced profiles, since the geometric mean tends to penalise low values, especially in the presence of other values that are not so low (unbalanced profiles).

Exclusion of Goals. In order to investigate the stability of the rank intervals further, the analysis compare the original SDG-GI ranks with those that would have been obtained by including the goals with insufficient data coverage (i.e., those with less than 75% of available indicators) that were initially excluded from the calculation of the countries scores. To accomplish this, the missing indicators are replaced with the mean value of the remaining available indicators within the respective goal. This exercise allows us to determine how much the exclusion of goals with insufficient data coverage penalizes or rewards the countries' scores.

Weights. Monte Carlo simulation comprised 1 000 runs of different sets of weights for the 14 goals constituting the SDG-GI. The weights are the result of a random extraction based

[7] Saisana, M., B. D'Hombres, and A. Saltelli. 2011. 'Rickety Numbers: Volatility of University Rankings and Policy Implications'. *Research Policy*, 40: pp. 165–177.
Saisana, M., A. Saltelli, and S. Tarantola. 2005. 'Uncertainty and Sensitivity Analysis Techniques as Tools for the Analysis and Validation of Composite Indicators', *Journal of the Royal Statistical Society* A 168 (2): pp. 307–323.

on uniform continuous distributions centred in the reference values[8] plus or minus 20% of these values.

Four models were tested combining the different aggregation formulas and imputation methods, which resulted in a total of 4 000 runs of simulations (1 000 simulated sets of weights for each combination of aggregation and imputation). See **Table 9** for further details.

Table 9: Alternative assumptions considered in the robustness analysis

	Reference	**Alternative**
I. Aggregation formula	Geometric average	Arithmetic average
II. Goals with low coverage	Excluded	Included
III. Weighting system of goals	Fixed Equal weights 0.714	Varying up to 20% U[0.571;0.857]

Source: European Commission's Joint Research Centre, 2022.

The main results obtained from the robustness analysis are shown in **Figure 3**, with median ranks and 90% intervals computed across the 4 000 Monte Carlo simulations. Countries are ordered from worst to best according to their original SDG-GI rank, where each blue dot represents the median rank among the iterations for each country, and error bars represent the 90% interval across all simulations, i.e. from the 5th to the 95th percentile.

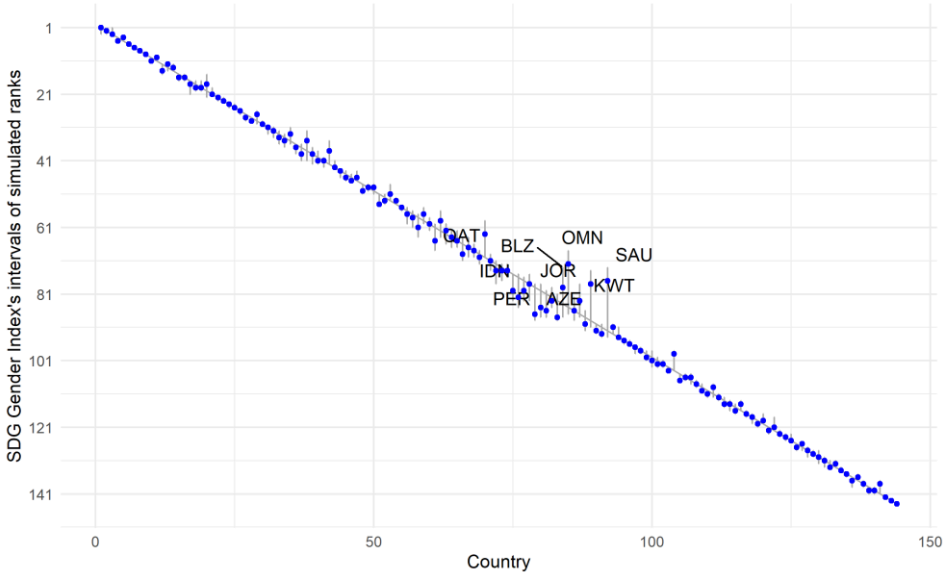
SDG-GI ranks are shown to be representative of a plurality of scenarios and robust to changes in the aggregation method, the exclusion/inclusion of goals, and the goals' weights for most of the countries. Suppose one considers the median rank across the simulated scenarios as being representative of these scenarios. In this case, the fact that the SDG-GI rank is close to the median rank (less than five positions away) for 92% of the countries suggests that the SDG-GI represents a suitable summary measure of the four scenarios tested. Furthermore, the reasonable narrow intervals for most of the countries' ranks (less than 10 positions for about 94% of countries) imply that the ranks are also, for most countries, robust to changes in the goals' weights and other modelling assumptions.

Qatar, Indonesia, Peru, Belize, Jordan, Azerbaijan, Kuwait, Oman, and Saudi Arabia are the only countries with simulated intervals greater than or equal to ten positions. This is most likely due to the effect of one of the methodological assumptions included in the uncertainty analysis. The following paragraphs delve into this topic.

Overall, changes in goal weights, goal exclusion/inclusion, and aggregation formula have little effect on country rankings in SDG-GI. These ranks are robust enough to allow for meaningful inferences for a vast majority of countries. For the sake of transparency and information, **Table 10** displays the country ranks along with the simulated intervals (central 90 percentiles observed among the 4 000 scenarios).

[8] Goals are all equally weighted in the original methodology (the weight is approximately 0.071

Figure 3: Robustness analysis on ranks, SDG-GI rank vs median rank and 90% intervals



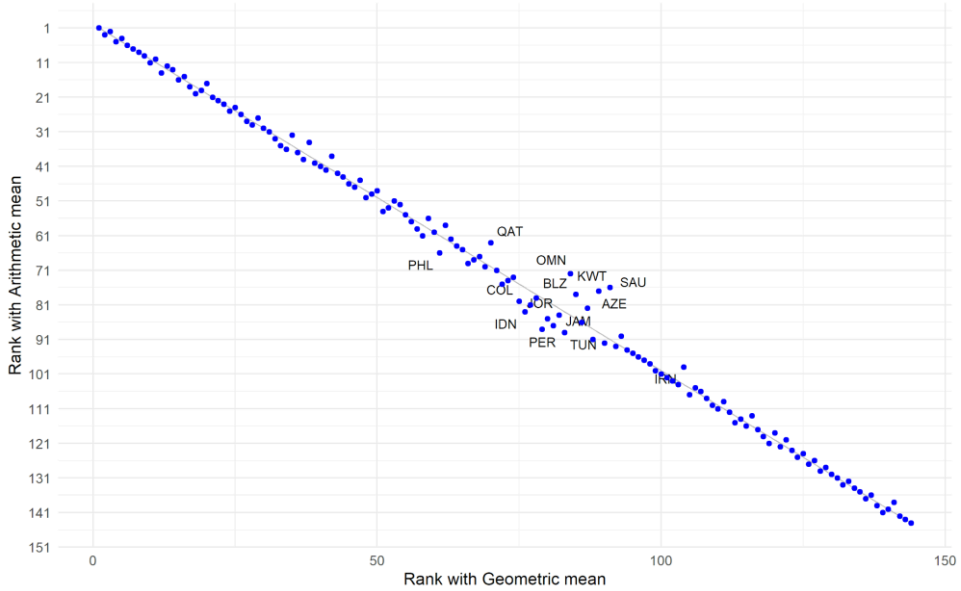
Source: European Commission's Joint Research Centre, 2022.
Note: Labelled countries show a shift of at least ten positions.

The uncertainty analysis is also complemented by a sensitivity exercise, in which the Index rankings are compared to the rankings resulting from specific changes in the modelling assumptions. In **Figure 4**, the ranks derived from SDG-GI are compared to the ranks obtained by changing the aggregation procedure from geometric to arithmetic mean. This comparison allows us to determine whether the variability in the rank intervals is due to the modelling assumptions that underpin the aggregation procedure. The countries above the grey diagonal line in the figure rise in rank position with the arithmetic mean. They are most likely penalized by the geometric mean because of their unbalanced profiles. When comparing the two alternative formulas, 13 countries show at least five positions of difference, 9 of which also have simulated intervals in **Figure 3** greater than – or equal to – ten positions. This could indicate that the aggregation formula has a substantial impact on the rankings. On one hand, it is relevant information for the interpretation of the results. On the other hand, it states the importance of the choice of aggregation formula, which needs to be clearly stated and motivated in the developers' methodological section.

Similarly, it is possible to compare the original ranks with the ranks that would have been obtained by including the goals with insufficient data coverage. This comparison allows us to investigate the rank intervals' stability further and propose an alternative way to use the goals' structure.

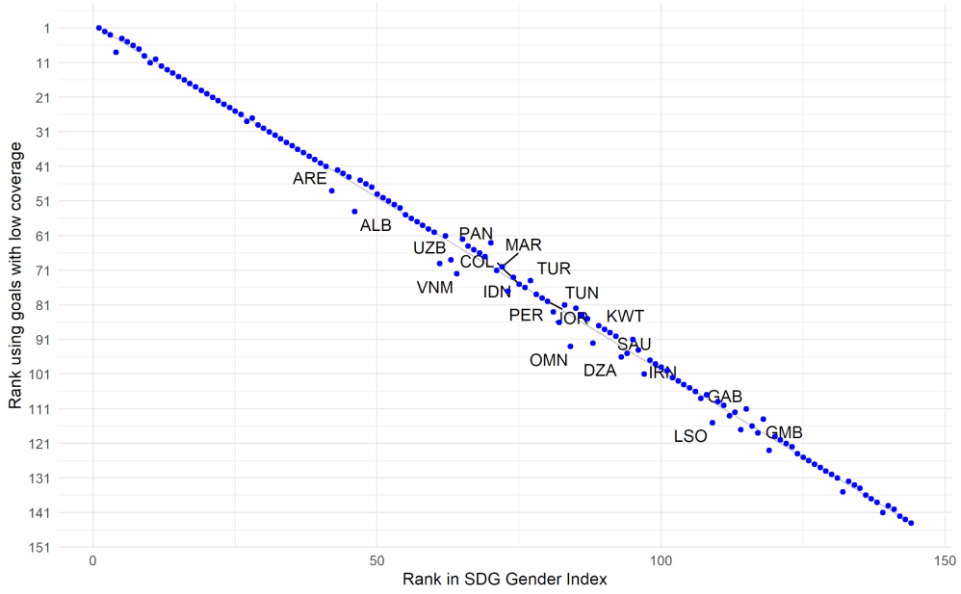
The results are depicted in **Figure 5**, with the x-axis reporting the country ranks after excluding the goals with insufficient data coverage and the y-axis indicating the ranks obtained after replacing missing indicators with the mean value of the remaining available indicators within the respective goal. In response to this change in the modelling assumptions, the SDG-GI ranks remain relatively stable. Excluding goals with insufficient coverage, according to the findings, has had a positive impact on a relevant number of countries. Indeed, countries dropping at least five positions (dot points below the diagonal) from their original SDG-GI ranking are somehow penalised by the inclusion of incomplete goals in the score.

Figure 4: Sensitivity Analysis: Ranks according to geometric and arithmetic means



Source: European Commission's Joint Research Centre, 2022.
Note: Labelled countries show a shift of at least five positions between the two aggregation formulas.

Figure 5: Sensitivity Analysis: Ranks according to exclusion and inclusion of goals with low coverage



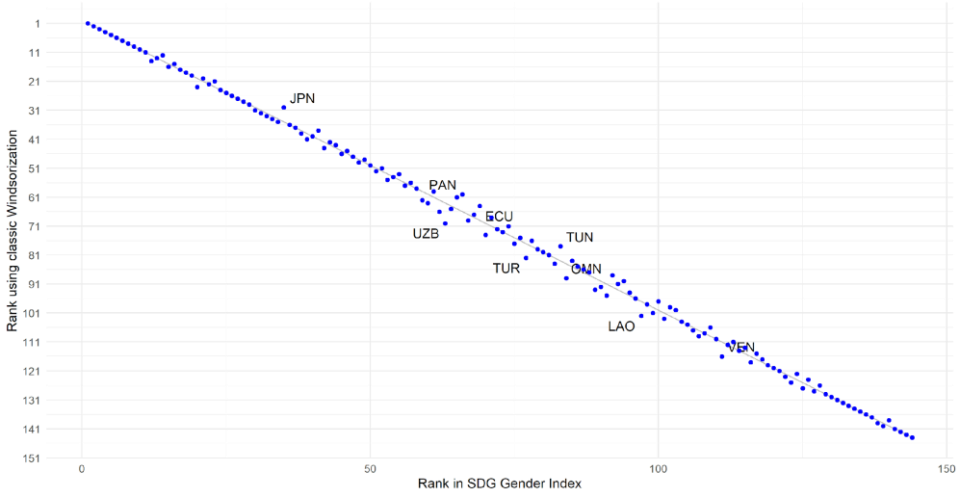
Source: European Commission's Joint Research Centre, 2022.
Note: Labelled countries show a shift of at least five positions.

In addition to the analysis included in the uncertainty, JRC-COIN investigated the role of normalisation and treatment of outliers in the computation of the SDG-GI: the indicators with Skewness and Kurtosis, respectively, above 2 and 3.5, were treated with the classic winsorisation approach, unless if the values needing trimming were 5 or more. Only one indicator (11.2) needed five observations to be treated. In this case, a logarithmic transformation has been used, and no single observation was trimmed. For what concerns normalisation, the decisions made by the developers were kept as they are. However, the MinMax normalisations have been based only on the raw values of the indicators for

countries included in the Index. It is possible to compare the original ranks with the ranks that would have been obtained with the alternative computation. This comparison makes it possible to further understand the ranks shown by the countries and identify the importance of the choices related to outliers and normalisation.

The results are depicted in **Figure 6**, with the x-axis reporting the countries' ranks in the SDG Gender Index and the y-axis indicating the ranks obtained using the alternative data treatment approach. In response to this change in the modelling assumptions, the SDG-GI ranks remain relatively stable. The methodology followed by the developers does not have a huge impact on the rankings. Among the nine countries with a rank shift of five or more positions, there is a balance between positive and negative shifts.

Figure 6: Sensitivity Analysis: Ranks according to data treatment approach



Source: European Commission's Joint Research Centre, 2022.
Note: Labelled countries show a shift of at least five positions.

Table 10: SDG-GI ranks and 90% intervals

ISO3	SDG-GI rank	Interval	ISO3	SDG-GI rank	Interval	ISO3	SDG-GI rank	Interval
DNK	1	[1-3]	CHL	49	[48-50]	LAO	97	[97-99]
SWE	2	[2-3]	ROU	50	[48-51]	GTM	98	[97-99]
NOR	3	[1-3]	MKD	51	[51-54]	HND	99	[98-101]
ISL	4	[4-5]	MDA	52	[51-54]	NAM	100	[98-103]
FIN	5	[4-5]	BLR	53	[48-53]	KHM	101	[100-103]
AUT	6	[6-6]	KAZ	54	[52-54]	BWA	102	[101-103]
NLD	7	[7-7]	TTO	55	[55-56]	GHA	103	[103-104]
CHE	8	[8-8]	ZAF	56	[55-60]	IRN	104	[98-104]
LUX	9	[9-9]	BIH	57	[56-61]	LBN	105	[105-107]
IRL	10	[10-11]	MNG	58	[57-64]	MMR	106	[105-107]
NZL	11	[10-11]	MYS	59	[55-60]	BGD	107	[105-108]
ESP	12	[12-15]	THA	60	[58-62]	RWA	108	[107-109]
BEL	13	[12-14]	PHL	61	[60-68]	LSO	109	[108-110]
AUS	14	[12-14]	CHN	62	[56-64]	KEN	110	[110-112]
EST	15	[15-17]	UZB	63	[60-66]	VEN	111	[108-112]
CAN	16	[15-17]	VNM	64	[62-67]	TZA	112	[111-113]
FRA	17	[17-21]	UKR	65	[62-67]	SEN	113	[112-115]
GBR	18	[17-20]	PAN	66	[64-71]	MOZ	114	[113-116]
SVN	19	[17-20]	MEX	67	[64-70]	GAB	115	[114-117]
SGP	20	[15-22]	KGZ	68	[66-70]	SWZ	116	[113-116]
DEU	21	[19-21]	ECU	69	[68-72]	BEN	117	[116-118]
CZE	22	[22-23]	QAT	70	**[59-70]**	GMB	118	[117-120]
PRT	23	[22-23]	RUS	71	[69-74]	CMR	119	[119-121]
LTU	24	[24-25]	MAR	72	[71-78]	ZWE	120	[117-120]
ISR	25	[24-25]	DOM	73	[72-77]	CIV	121	[121-123]
LVA	26	[26-27]	PRY	74	[72-76]	IRQ	122	[118-122]
CYP	27	[27-29]	COL	75	[75-83]	PAK	123	[122-124]
HRV	28	[28-29]	IDN	76	**[75-85]**	ZMB	124	[123-125]
MLT	29	[26-30]	TUR	77	[76-83]	ETH	125	[123-125]
SVK	30	[29-30]	BRA	78	[75-83]	BFA	126	[126-128]
URY	31	[31-33]	PER	79	**[78-89]**	TGO	127	[126-128]
ITA	32	[31-34]	JOR	80	**[78-88]**	AGO	128	[127-130]
HUN	33	[32-36]	JAM	81	[80-88]	UGA	129	[128-130]
POL	34	[33-37]	TJK	82	[79-85]	NGA	130	[128-132]
JPN	35	[31-36]	TUN	83	[81-89]	MWI	131	[130-133]
KOR	36	[36-39]	BLZ	84	**[73-88]**	LBR	132	[131-133]
GRC	37	[36-41]	OMN	85	**[68-87]**	MLI	133	[131-133]
USA	38	[32-41]	LKA	86	[82-89]	MDG	134	[133-134]
SRB	39	[37-42]	AZE	87	**[78-88]**	MRT	135	[135-136]
MNE	40	[38-42]	NIC	88	[86-92]	SLE	136	[136-139]
CRI	41	[40-43]	KWT	89	**[74-91]**	GIN	137	[135-137]
ARE	42	[35-42]	SLV	90	[90-93]	BDI	138	[137-139]
MUS	43	[41-43]	IND	91	[90-94]	COD	139	[139-141]
ARG	44	[44-46]	SAU	92	**[73-94]**	NER	140	[139-141]
ARM	45	[44-47]	DZA	93	[90-93]	SDN	141	[137-141]
ALB	46	[45-47]	NPL	94	[91-94]	YEM	142	[142-142]
BGR	47	[44-47]	BOL	95	[95-96]	AFG	143	[143-143]
GEO	48	[48-51]	EGY	96	[95-96]	TCD	144	[144-144]

Source: European Commission's Joint Research Centre, 2022.

6 Conclusions

The JRC statistical audit delves into the extensive work carried out by the developers of the SDG Gender Index with the aim of suggesting improvements in terms of data characteristics, structure and methods used. The analysis aims to ensure the transparency of the Index methodology and the reliability of the results.

The developers set thresholds for the inclusion of countries in the SDG Gender Index. Only countries with at least 13 goals out of 14 are included in the final composite indicator. Among these countries, only goals with at least 75% of coverage, are considered available. This implies that a country with two missing values in a goal with five indicators is treated as if it has no data in that goal (not even the available three). While this choice is completely legitimate and meaningful, JRC-COIN suggests considering an alternative approach to maximise the use of available data. That is, to maintain the developers' approach of not calculating the goal score when the aforementioned criteria are not fulfilled but at the same time use this subset of indicators - as a proxy of the goal – in the construction of the overall index.

Outliers are present in only two indicators after data treatment. However, due to the conceptual nature of these two indicators, it is fair to avoid any additional transformation. A possible alternative for the future would be the use of a logarithmic transformation. JRC-COIN suggests a detailed description of the methods used for data treatment and normalisation. Especially in the cases when the approach is tailored to a single indicator and does not adhere to a general rule.

According to the findings, the Index is statistically well balanced within its goals. The majority of the indicators have positive correlations with their corresponding goals, implying that they provide meaningful information on the variation of the scores. Some indicators have weaker correlations with the other indicators in their respective goals, resulting in lower correlations at the index level.

Our recommendation is to continue monitoring these specific indicators in the future editions, in order to check their behaviour and, if necessary, modify them. The general concepts' consistency is further confirmed by very good correlations between the goals and the overall Index. As a consequence, the inclusion of the goals intermediate step of aggregation proves to be undoubtedly positive.

The principal component analysis provides a clear understanding of the close relationship between the goals. Overall, given the variety of goals, the results demonstrate satisfying coherence, and the presence of an underlying single concept is clear, which is consistent with the correlation analysis results.

The results of the uncertainty analysis reveal that the SDG-GI is a robust summary measure for almost all countries. The simulated intervals are narrow enough for meaningful inferences to be drawn from the index for 94% of the units observed; there is a shift of less than 10 positions for about 94% of the countries included in the Index. Nevertheless, there are nine countries with 90% confidence interval widths of at least 10 positions, which is due to a lack of balance among their values on the fourteen goals. Significant variation of ranks associated to these countries is also observed in the sensitivity analysis.

Considering the specificity of the concept of sustainability and the related gender aspects, this audit confirms that the SDG Gender Index is reliable, and that the framework has a good statistical coherence. The audit also acknowledges the significant efforts by the developers’ team to obtain a balanced and transparent result.

References

European Commission, Joint Research Centre, Papadimitriou, E., Caperna, G., JRC statistical audit of the equal measures 2030 SDG gender index, Publications Office, 2019, https://data.europa.eu/doi/10.2760/09421

Groeneveld, R. A. and G. Meeden. 1984. 'Measuring Skewness and Kurtosis'. *The Statistician* 33: 391–399.

OECD/EC JRC (Organisation for Economic Co-operation and Development/European Commission, Joint Research Centre). 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide.* Paris: OECD.

Saisana, M., B. D'Hombres, and A. Saltelli. 2011. 'Rickety Numbers: Volatility of University Rankings and Policy Implications'. *Research Policy* 40: pp. 165–177.

Saisana, M., A. Saltelli, and S. Tarantola. 2005. 'Uncertainty and Sensitivity Analysis Techniques as Tools for the Analysis and Validation of Composite Indicators'. *Journal of the Royal Statistical Society* A 168 (2): pp307–323.

List of tables

Table 1: Conceptual framework of the SDG-GI 4
Table 2: Summary statistics of the indicators included in the SDG-GI 7
Table 3: Correlations between indicators in the same SDG 9
Table 4: Correlations between indicators and their aggregates 11
Table 5: Correlations between SDGs 12
Table 6: Correlations of SDGs with the index 13
Table 7: Eigenvalues and explained variance for the first ten principal components 14
Table 8: Distribution of rank differences between pillar and SDG Gender Index rankings 16
Table 9: Alternative assumptions considered in the robustness analysis 19
Table 10: SDG-GI ranks and 90% intervals 23

List of Figures

Figure 1: Factor map of the 14 goals and comparison with the overall SDG-GI 15
Figure 2: Average shifts in SDG Gender Index country rankings when one element is omitted at a time (goals and indicators) 17
Figure 3: Robustness analysis on ranks (SDG-GI rank vs median rank and 90% intervals) 20
Figure 4: Sensitivity Analysis: Ranks according to geometric and arithmetic means 21
Figure 5: Sensitivity Analysis: Ranks according to exclusion and inclusion of goals with low coverage 21
Figure 6: Sensitivity Analysis: Ranks according to data treatment approach 22

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications
You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-31016-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office of the European Union

doi:10.2760/993717

ISBN 978-92-76-49351-8