Statistical Journal of the IAOS 36 (2020) 1129–1142
DOI 10.3233/SJI-200688
IOS Press

A statistical approach for assessing progress towards the SDG targets

Pietro Gennari[a,*] and Marcello D'Orazio[b]
[a]*Chief Statistician of the Food and Agriculture Organization (FAO) of the United Nations, Rome, Italy*
[b]*Office of the Chief Statistician, Food and Agriculture Organization (FAO) of the United Nations, and Italian National Institute of Statistics (Istat), Rome, Italy*

Abstract. The global SDG indicator framework establishes a set of measurement tools to assess country performances in a comparable way, and helps governments to identify appropriate policy interventions to achieve the SDG targets. Five years into the implementation of the 2030 Agenda, however, still different methods are being used by leading international organizations for assessing whether the SDG targets will be achieved or not. This may lead to different results, sometimes contradictory, generating confusion among users and policy-makers, who therefore cannot base their policy decisions on solid and coherent assessments. This article describes some of the solutions proposed by leading international organizations to address two distinct measurement objectives: (i) monitor the "current" status of achievement of a SDG target, i.e. the situation as pictured by the latest available data, and (ii) assess whether the SDG targets can be achieved by 2030. These distinct objectives are then translated in various methodological approaches, that often include also a way for identifying the targets when not explicitly set, and the procedure to obtain regional and global aggregates (as well as, aggregates by target and goal). This article provides a critical overview of the different approaches and proposes a unified coherent statistical approach for progress and status assessments, highlighting its advantages over the alternative approaches, and demonstrate its application to a specific FAO indicator. The article focuses mainly on the assessment of (i) and (ii), while is not intended to investigate the issues related the aggregation of results at target/goal level, a topic that is beyond the scope of this work.

Keywords: SDG indicators, status assessment, progress assessment, growth models

1. Introduction

The 2030 Agenda for Sustainable Development is the new, broader and more ambitious development framework that the Member States of the United Nations endorsed in September 2015 to replace the Millennium Development Goals. The new Agenda is, at the same time, universal and all encompassing, setting 17 Sustainable Development Goals (SDGs) and 169 targets that should set the path to economic, social and environmental prosperity for all the people of the entire planet up to year 2030. It is grounded on a set of universal principles, values and standards that are applicable in all countries of the world, rich and poor, in all contexts and circumstances and at all times.

The global SDG indicator framework, firstly endorsed by the UN Statistical Commission, and subsequently by the ECOSOC and finally by the UN General Assembly in July 2017, is the foundation of the 2030 Agenda's mutual accountability mechanism, establishing the agreed set of measurement tools to assess country performances in a comparable way, and helping governments to identify the most appropriate and timely policy interventions to actually achieve the SDG targets by the agreed deadline.

Five years into the implementation of the 2030 Agenda, is becoming more and more pressing the demand of governments, donors and international organizations to assess whether the established SDG targets

*Corresponding author: Pietro Gennari, Food and Agriculture Organization (FAO) of the United Nations, Viale delle Terme di Caracalla, Rome Italy. Tel.: +39 6 57053599; E-mail: pietro.gennari@fao.org.

will be achieved or not, at which level (global, regional or country), and on whether, beyond national averages, inequalities between different population groups and territorial areas within each country will be reduced and even eradicated by the end of the 2030.

To meet this demand, various methodological approaches have been developed by some leading regional/international agencies to convey their assessment of the situation, as pictured by the latest available SDG data, and what can be expected at the end of the reporting period. These different methodological approaches have often generated uncertainty and confusion among the users, as their results are often inconsistent and sometimes contradictory. This outcome is not only the result of adoption of different assessment approaches, but also due to the fact that sometimes the monitoring exercises are conducted using different sets of indicators, not always corresponding to those included in the global indicator framework. In this work we focus on the statistical aspects of the monitoring approaches. For this purpose, as a starting point, a distinction will be made between: 1) what ought to be measured; and 2) once this is decided, which measurement approach should be adopted. Regarding the first point, among the different solutions proposed by organizations like OECD, Eurostat, ESCAP, SDSN and UN, two distinct measurement objectives can be distinguished: (i) monitor the level of achievement, i.e. the situation as described by the latest available data, and (ii) assess whether the targets can be achieved by 2030. The (i) and (ii) measurement objectives are translated in various methodological approaches adopted to measure the latest and future status of the indicators, the way of identifying the targets when not explicitly set, and the geographic aggregation procedure (as well as the aggregation by target and goal). According to the OECD, the "heterogeneous nature of SDG targets means that setting desirable levels of achievement by 2030 on each indicator requires a variety of approaches" (see [1]), which creates the risk that different institutions may come up with different assessments of where the world stands in achieving the 2030 Agenda.

This paper provides a critical overview of the different approaches adopted by some leading international/regional agencies in preparing their progress reports and proposes a unified coherent approach for progress and status assessments that can serve as common basis for international organizations as well national authorities willing to conduct a comparable assessment of the different components of the 2030 Agenda. The proposed approach does not aim to be seen as the best possible method for each single indicator or country, rather to be considered a good compromise solution, easy to apply to the set of global SDG indicators given their main constraints and limitations. This has the objective of increasing the comparability and coherence of the monitoring exercises carried out at national or international levels. Section 2 will provide a brief overview of the key problems that monitoring the 2030 Agenda pose to the global statistical community and some proposals on how to address them. Section 3 will present different approaches for measuring the level of achievements of the SDG indicators, highlighting the pros and cons of the different solutions suggested, while the alternative methods used to assess the progress towards the various SDG targets will be presented in Section 4. Section 5 will describe the approach proposed in this article and its advantages versus the alternative solutions, while Section 6 will present the results obtained by applying the proposed methodology to one key SDG indicator under FAO custodianship. A few concluding remarks will summarize the main results of this research.

2. The Global SDG targets and the corresponding indicators

Despite the enlightening experience of the MDG era, which led the IAEG on MDG indicators to compile a list of recommendations aimed at ensuring that certain well-defined issues emerged in monitoring the MDG targets would not be replicated again in the new development framework [2], some of the most cumbersome problems are still present today when statisticians attempt to provide precise and unequivocal indications to policy-makers on the most appropriate policy interventions to be undertaken to achieve the SDGs.[1] Among the many critical issues related to the MDG process, one aspect which is in common with the SDG framework is that some of the targets are poorly specified or are too ambitious. Moreover one of the key weaknesses of

[1] The 2030 Agenda does not indicate the most appropriate policy interventions to achieve the SDG targets. It is worth noting that the vast majority of the targets included in the 2030 Agenda are development outcomes. The 2030 Agenda only includes some means of implementation (related to finance, trade, capacity building, science, technology and innovation) in almost every Sustainable Development Goal. For this reason, in order to achieve the 2030 Agenda, the global community should commit to implement the associated Addis Ababa Action Agenda (AAAA), which is a global framework for financing development.

the MDG monitoring framework was the fact that "discrepancies between national and international data, due to, among others, different methodologies, definitions, different choice of data sources, or time gap in release dates, created problems at the national level and tension in the international statistical community." Nowadays, the decision of the UN Statistical Commission to allow countries to develop "alternative" indicators to the official Global Indicator Framework is bound to lead to the same discrepancies between national and international data.

The fact that the lessons learned from the MDG process have not been applied in most cases to the 2030 Agenda, and the same problems are encountered once again in the SDG target setting, is due to the complete separation between the political and the statistical processes. The political process was centered on the so-called "Open Working Group" (OWG) which was mandated by the UN General Assembly with the task of proposing a set of Sustainable Development Goals and corresponding targets by September 2014. Subsequently, the Interagency and Expert Group on SDG indicators, under the oversight of the UN Statistical Commission, was mandated by the UN General Assembly to select the most condensed and, at the same time, relevant list of indicators for the targets already defined by the political process. This disconnect in the formulation of targets and indicators explains many of the problems mentioned above. "The target setting process was fully in the hands of policy-makers, diplomats and permanent representations, who usually have a limited understanding of whether a target is measurable, in principle, and whether relevant data are available for the target to be properly monitored, in practice." [3].

A first problem that clearly emerges from the analysis of the Global SDG monitoring framework is the mismatch in a number of cases between targets and related indicators, as the officially selected indicators measure only some components of the relevant SDG targets. The complexity of the formulation of many targets, which combine multiple and often complementary (if not discordant) policy objectives, has made it impossible for the global statistical community to select a limited number of indicators (max 1–2 per target) that could monitor in a comprehensive way the entire range of policy dimensions of each target. As a result, when existing indicators do not adequately address predefined targets, it becomes difficult for statisticians to assess whether these targets are going to be actually reached or not. A derived challenge stems from the difficulty in summarizing the information of different indicators monitoring one single target.

Another key problem of the global SDG monitoring framework is that only a minority of targets (about 30%) have clear numerical benchmarks, while the majority aim to achieve rather ambiguous qualitative trends. It is clear that without a numerical benchmark the possibilities to assess progress towards these targets face significant limitations.

Even when targets are quantifiable, an important distinction to be made is between relative and absolute benchmarks. Absolute numerical targets are often too ambitious to be reached by a large proportion of countries, as they do not take into account the initial conditions from which disadvantaged countries start their development trajectories. For example, based on the assessment of historical and current trends, the target of achieving zero hunger and zero poverty by 2030 cannot realistically be reached in many countries and globally. It is rather an aspirational target to orient public policies to improve the livelihood of the poor populations and to advocate for the provision of additional resources from development partners to alleviate the problem, rather than a feasible and tangible quantitative level of achievement to pursue. Also from a measurement perspective, values equal to zero are impossible to obtain just considering the presence of disparities, even minimal, within a country (or across countries, when considering regional or global aggregates). As a result, thresholds higher than zero need to be set under which the relevant countries can be considered to have achieved the target.

Another practical problem affecting the possibility of progress measurement, for an SDG monitoring framework developed only recently, is the availability and quality of baseline data (the baseline year coincides with 2015, as this is the year when the 2030 Agenda was adopted). These values are not available in many countries since many SDG indicators are new and their methodology has been developed and internationally endorsed only recently (certainly after 2015). In addition, normally the time lag at international level between data collection and data dissemination is normally 2–3 years for most SDG indicators and this, currently, limits considerably the possibility to have at least one data point after the baseline, which is the essential condition to calculate the rate of progress towards the target.

Finally, consolidating results at regional and global levels can assume completely different meanings and interpretation, according to the modalities of data aggregation. Global/regional targets can be considered achieved when the large majority of countries have

reached this objective or when the majority of the population living in these territories have achieved it. These are evidently two distinct results: in one case, countries are added one another independently from their size (in terms of land, GDP or population), i.e. without considering any weighting factor; in the other case, global results are obtained as a weighted average of the country results. As a result, when trends at global and regional level are obtained as weighted average, they tend to be dominated by a few countries with large population, GDP or land size (according to the weighting variable used).

3. Monitoring the implementation of the 2030 Agenda

Monitoring the implementation of the 2030 Agenda is a key cornerstone in the edifice of the SDG mutual accountability framework. Unfortunately, monitoring the results of the policy measures aimed at achieving the SDGs has been interpreted in various manners by different leading international organizations, generating confusion among users about what should be measured, and which methods should be applied. The confusion is also due to a terminology that sometimes is misleading; for instance, some reports focus on "measuring the distance to the target", which is usually considered as an assessment of the level of achievement of the targets as portrayed by the latest available data, but in practice it may also involve a nowcast of the current value of the SDG indicator, to compensate for the time lag with which SDG data are normally disseminated. This exercise would therefore also involve an assessment of progress over time (from the last available SDG value to the nowcast of the current one). Moreover, the monitoring exercises often consider only a subset of the whole set of official indicators, or replace some of them with proxies that do not fully correspond to the global indicator framework.

The Bertelsmann Foundation, in collaboration with the UN Sustainable Development Solutions Network (SDSN; see [4]), for example, compiles a SDG index and dashboard that summarizes countries' achievements and trends related to a specific set of indicators that do not correspond to global indicator framework. The SDG index measures the percentage of achievement of the 17 SDGs at country level and, in practice, it summarizes the situation described by the latest available data by offering a metric that is the opposite of a distance to the target (100 minus the score, is the distance to the target, in percentage points). In addition, the dashboard indicates whether, according to past trends of SDG indicators, a country is on track to achieve the various targets and goals by 2030. Similarly, the OECD in its latest report [1] focuses on "the distance that countries need to travel in order to meet the SDG targets by 2030". It is an assessment of the situation described by the latest available data compared to the target; however, under this umbrella OECD also performs an assessment of the trend over time. Eurostat [5] follows a different approach and analyzes solely the progress made towards the SDGs targets from the baseline in Europe (EU-28); this approach allows to assess whether the target will be reached or not by 2030 only when the target is quantitative, while for the remaining indicators, it evaluates just the speed of progress (for further detail see Section 3.2). Eurostat, however, considers a set of indicators that only partially correspond to those included in the global indicator framework. Finally, the UN, which is the ultimate custodian of the SDGs, agreed on the need of assessing both the level of achievements and the trend over time and since 2019 its Sustainable Development Goals Report comes along with a Progress Chart,[2] based on a limited number of indicators, that provides a summary picture of progress made at the global and regional levels towards selected targets of the 2030 Agenda.

The SDG reports prepared by leading international organizations usually include an assessment of the progresses towards the achievement of the SDG targets while only some of them provide also an evaluation of achievements associated to latest available data. The following sub-sections give a critical overview of the approaches suggested by some leading international agencies to assess both the level of achievement and the trend over time.

3.1. Measuring the level of achievement

The assessment of the level of achievement of an SDG indicator consists in measuring how close is its latest available value to the ideal value. This is straightforward when the ideal value of an indicator is explicitly set in the formulation of the relevant target. Measuring the distance from the target can be obtained in different ways, partially dependent also on the nature of the indicators themselves. To facilitate a comparison

[2] See https://unstats.un.org/sdgs/report/2019/progress-chart.pdf and https://unstats.un.org/sdgs/report/2020/progress-chart-2020.pdf.

across indicators and allow their aggregation at target or goal level, it is necessary to standardize the adopted measures.

OECD suggests using a traditional normalization transformation ("z-score", see [1]):

$$z_{it} = \frac{|x_{it} - x_i^*|}{s_{xt}}$$

where x_{it} is the value of the generic indicator of country i at time t; x_i^* is the target value for the indicator for country i (often the target set in the agenda is the same for all countries, i.e. $x_i^* = x^*$; see Section 3.1.1 for major details); finally, s_{xt} is the standard deviation of x calculated over all countries in year t. The larger is the variability of the indicator across countries, the shorter is its distance to the target, compared to the standard deviation. The z-score can also be viewed as the city-block distance (Manhattan distance; see [6]) to the target, standardized with the standard deviation. As opposed to other standardized distance measures, z_{it} does not vary between 0 (target reached) and 1; according to the OECD, this metric is seldom greater than 3 (see [1]). Following this logic, values of z_{it} greater than 2 would denote countries that are far from the target. The z-score, however, may not be the best approach for highly skewed distributions; in addition, the estimate of the standard deviation across countries (s_{xt}) can be affected by the presence of outliers. In this latter case, the impact of outliers can be attenuated by replacing s_{xt} by a robust estimator (see e.g. [7]).

The SDSN/Bertelsmann Stiftung report [4] adopts a different approach that considers the "worst" case as the reference:

$$x'_{it} = \begin{cases} \frac{x_{it} - x_t^{(w)}}{x_i^* - x_t^{(w)}}, & \text{when the desired direction is an increase over time} \\ \frac{x_t^{(w)} - x_{it}}{x_t^{(w)} - x_i^*}, & \text{when the desired direction is a decrease over time} \end{cases}$$

$x_t^{(w)}$ is the worst value observed among countries in year t. The x'_{it} is not a distance to the target since it takes value 0 in the worst case (maximum distance to the target) and 1 in the best-case scenario. Its complement to 1 ($d_{it}^{(\text{SDSN})} = 1 - x'_{it}$) can be viewed as a measure of distance and, since usually $x_t^{(w)}$ and x_i^* are the extremes of the distribution, it would correspond to the city-block distance standardized by the range of the observed values. Unfortunately, the range may be affected by the presence of outliers and consequently it would be preferable to remove them before the identification of the worst value, or set the worst value equal to a percentile close to the boundary (i.e. $P{-}2.5\%$ or $P{-}5\%$ for indicators whose target is in the right tail of the distribution).

The UNESCAP (see [8,9]) suggests the *baseline status index approach* to measure the progress made by each region or sub-region compared to the distance between its starting point and the target:

$$b_{gt} = \begin{cases} \frac{x_{gt} - x_{gt_0}}{|x_g^* - x_{gt_0}|}, & \text{when the desired direction is an increase over time} \\ \frac{x_{gt_0} - x_{gt}}{|x_g^* - x_{gt_0}|}, & \text{when the desired direction is a decrease over time} \end{cases}$$

x_{gt_0} is the value of the generic SDG indicator for region g in the year considered as "baseline" (t_0) (an alternative formula is suggested for parity indicators; see [9] for more details). This index measures the distance traveled by a region/sub-region from the baseline compared to the required distance to reach the target. A value of b_{gt} equal to 1 indicates that the target has been achieved; a value of 0 denotes no improvement from the baseline; negative values denote deterioration. Although b_{gt} has been designed to work with regional/sub-regional SDG data, it can also be applied at country level.

One problem of the indicator is that it cannot be calculated when the denominator is 0, i.e. when a region has already achieved the target in the baseline year (even though it may be distant from the target it in the latest available year in the time series, i.e. $x_g^* = x_{gt_0}$ but $x_{gt} \neq x_{gt_0}$). It is worth noting that, by assessing the situation with respect to the baseline, in practice the indicator measures the progress over time with respect to the baseline (assuming a linear growth over time), rather than measuring the level of achievement.

The UN in the 2020 Progress Chart (see footnote 2) measures the level of achievement of an SDG indicator in terms of a generic distance to the target which is categorized in five categories that go from "Very far from target" to "Target met or almost met". The technical note [10] associated to the Progress Chart does not clarify how the distance should be calculated nor how to decide which of the final five categories has to be picked up. The outcomes of the assessment (at global and regional level) of the level of achievement for the subset of the indicators included in the 2020 UN Progress Chart show that different criteria have been applied depending on the indicator and on the existence of an explicit numerical target.

3.1.1. Indicators without a numerical target

All the approaches proposed to assess the status of achievement described in the previous Section require an explicit numerical target (x_i^*). In some cases the

target set in the 2030 Agenda is expressed in absolute terms (a fixed value valid for all the countries/regions, i.e. $x_i^* = x^*$); in other cases the target is set in relative terms with respect to the baseline year (e.g. target 1.2 requests to reduce at least by half "the proportion of men, women and children of all ages living in poverty"). In the latter case, the target is specific to the initial condition of each country and changes from country to country. OECD in [1] noted that only 36% out of the 132 indicators selected for their report have a fixed target ($x_i^* = x^*$) explicitly set in the 2030 Agenda.

To assess the level of achievement in absence of an explicit target the various Agencies have suggested a series of procedures to determine "statistics targets" to be used in the assessment. OECD [1] suggests following a sequential set of procedures to set numerical values as benchmarks: 1) adopt quantified targets in international agreements or from subject-matter experts judgement; 2) if this is not possible, adopt the lowest value of the 10% best performing OECD countries (i.e. $\tilde{x}^* = P$ 90% or P 10% depending on the normative direction of the indicator). This procedure does not permit to identify a target for all the indicators, since some of them do not have a normative direction (about 17% of all indicators considered in [1]).

The SDSN/Bertelsmann Stiftung report [4] suggests a five-step decision tree do derive "statistics targets". The consensus reached by the scientific community is used for those indicators having a clear normative direction, i.e. the target corresponds to the best situation from the scientific viewpoint. In alternative, they calculate the average value of the top-five performing countries (or the three best performing OECD countries), i.e. the countries with the highest (lowest) SDG values when the normative direction corresponds to increase (decrease); this approach is also suggested in the case of several countries having already achieved the explicit target set in the Agenda. In this procedure, the outliers should be removed before calculating the averages or, in alternative, trimming can be applied, e.g. by removing 2.5% of the units in the left or right tail of the distribution, depending on the normative direction. The SDSN/Bertelsmann Stiftung report [4] considers 114 indicators all having a target set in the Agenda or determined according to the proposed strategy.

The top-performers approach is also adopted by UNESCAP [11] to set targets at regional level (about 70% or the 169 indicators included in the report). The procedure follows a two-stage approach: 1) for indicators with enough historical data (time series of at least 15 years), the regional targets are forecasted by applying the average rate of change estimated for the top performing countries; 2) for regions with relatively short time series, and many missing countries, the regional targets correspond to the average of the latest available values for the top five performing countries. Before identifying the top five performers, UNESCAP suggests removing potential outliers.

The UN Progress chart does not clarify how the assessment of a distance to the target is done in absence of a numerical target (see [10]). The practice adopted for some indicators indicates that ad hoc rules have been developed indicator by indicator. This, in practice, means that implicitly a "statistics" target has been set.

In general, the procedures that set the "statistical target" based on the top performing countries are rather questionable, given the structural diversity existing across countries. Some countries will never be able to reach a specific target, because their starting point is too distant and the specific country characteristics make this achievement an almost impossible event. For instance, let us consider the forest area as a proportion of total land area (SDG 15.1.1). In 2015 the estimated 90%-percentile worldwide corresponds to about 65% of forest coverage at country level. In case this proportion is set as the "statistics target", it is clear that countries in the arid regions will never be able to reach it. Even setting the "statistics targets" by region would not give satisfactory results, given the wide disparities present even at regional level, which makes the situation of some countries not comparable with others. In this context, the unique possible approach would be that of grouping countries in homogeneous clusters with respect to their characteristics relevant to the phenomena being studied and then use these groups to derive the corresponding targets based on the values observed on the best performers in each cluster. This clustering exercise can only be done when auxiliary data can be used for this purpose, but the chance of having a single grouping pattern that works for all type of indicators (economic, social and environmental) is rather low.

3.1.2. Aggregation issues

The methods proposed for measuring the distance to the SDG targets are mainly designed to assess the countries' situation as described by the latest available data. The only exception is the method suggested by UNESCAP that is uniquely tailored to assess the regions' distance to the SDG target. The UN progress Chart, performs the assessment only at regional and global level, even though the details of the method are not provided. The calculation of regional/global distance to the tar-

get is relatively simple for indicators having an explicit numerical target. For the OECD's z-score, the regional distance to the target is calculated as the weighted average of the distances of countries belonging to the region, where the weights correspond to the share of the country population over the total population of the region. The same approach (weighted average with population share) is considered in the SDSN/Bertelsmann Stiftung report [4].

The assessment of the regional status should obviously consider also the number of data gaps, i.e. the countries with missing values for the chosen indicator. The common practice to address this problem is to aggregate the valid data, without considering the missing cases, when their weight is greater that a fixed threshold (usually 50%, or more rarely 67%, of the total population of the whole region).

As already noted, using this weighting scheme, global and regional status tend to be dominated by a few countries with large population; the underlying assumption is that the target is achieved when the majority of the population living in the region has achieved it. When the assessment is made not with regard to the total population, but just in terms of number of countries, the aggregation is obtained by applying the simple mean (or, better, the median, which is more robust). However, both approaches (weighted or unweighted mean) suffer the usual problem that an average alone may lead to wrong conclusions in absence of a measure of variability or, more generally, a summary idea about the distribution of the distance to the target of countries belonging to a given region.

Usually the SDG monitoring reports also include aggregations at level of target or goal, in order to summarize the results of the various official SDG indicators selected to monitor the same target or the same goal. In this case, the preferred approach is to simply perform an average of the distance to the target for the various indicators' values. Weights are therefore not applied in order to avoid subjective choices on the importance of the various indicators under the same target/goal. Also in this case, the averaging may provide a misleading picture if it is not accompanied by a measure of variance of the different indicators under the same target/goal. In general, while aggregation at target level may work, since often the indicators under the same target are quite homogeneous, the aggregation at goal level seems less feasible due to the heterogeneity of the targets under the same goal.

Finally, averaging the outcomes of many indicators under the same target/goal does not consider that some of them may be highly correlated/associated and therefore provide redundant information which tends to attribute an inflated weight to certain indicators.

3.2. Measuring progress over time

The other major objective of the SDG progress reports is to assess the likelihood that a country or a region, at observed trends, will achieve the SDG targets by 2030 (the deadline established for most SDG indicators). This type of analyses can be carried out using different methodologies, from simple methods that ascertain the presence of the desired trend, to proper forecasting methods that predict the value of the indicator in the target year.

For an international organization whose aim is to prepare a report assessing the trends for the entire list of SDG indicators for all countries in the world, the main complexity is to define a simple enough methodology that can be applied seamlessly across all countries and indicators, producing internationally comparable results. This exercise is obviously simpler when the objective is the preparation of a national report where it is possible and preferable to adopt a methodology that considers the specific characteristics of the indicators available in the country. In practice, reports summarizing results related to all the countries in the world should adopt a compromise solution that takes into account all the limitations of the indicators monitoring the SDG targets. Estimating the parameters of a complex time series model, for example, requires in general a sufficient number of data points (greater than the number of parameters to estimate), which should be even larger in case of a high random variation of the indicator. Unfortunately, the type of SDGs data and the length of the time series often restricts the range of applicable methods for global monitoring reports. One important constraint in selecting the best methodology for global reports is the shortness of the time series, as most SDG indicators are relatively new. Even in the best scenario, their time series start in 2015 and the time lag from data collection to data dissemination is rather long (about 2 years): as a result, the number of data points in the series for most SDG indicators is very low (not more than 2 or 3). All these reasons have led the custodian Agencies[3] of the SDG indicators to adopt rather simple

[3]The UN Statistical Commission has assigned each official indicator included in the Global Indicator Framework to one (or more) custodian agency, which has the responsibility to develop the indicator methodology in consultation with countries; collect, harmonize and

methodological approaches, which could be suitable for all the possible situations encountered.

The SDSN/Bertelsmann Stiftung report [4] estimates the trends of SDG indicators by simply considering a *linear growth model* where the annual growth is estimated as:

$$\text{LAGR}_{Ai} = \frac{1}{t - t_0}(x_{it} - x_{it_0})$$

The observed linear trend is compared with the linear annual growth rate needed for achieving the target $\text{LR}_i = \text{LAGR}_{Ai}/\text{LAGR}_{Ri}$ where

$$\text{LAGR}_{Ri} = \frac{1}{T - t_0}(x_i^* - x_{it_0})$$

is the required linear growth to reach the target in the year T (namely 2030). The magnitude and the sign of LR_i guide the assessment: values equal or greater than 1, indicate that the target is expected to be reached by 2030 (country "on track"), if the observed trend is maintained over time. Conversely, negative values indicate that the indicator is moving in the opposite direction with respect to the required one (country "deteriorating"). Values close to 0 indicate stagnation, i.e. no growth over time compared to the baseline (according to [4] the situation is "stagnating" with $0 \leqslant \text{LR}_i < 0.5$); finally the progress is judged "moderately increasing" when $0.5 \leqslant \text{LR}_i < 1$.

The assumption of linear growth over time is rather simple, but it is often not adequate to describe the dynamics over time of complex phenomena. For instance, in case the SDG indicators are expressed as proportions (as in the majority of cases) it is difficult to maintain the linear trend assumption, because of the well-known problems of non-heteroscedasticity, non-Gaussian residuals, impossibility to guarantee that the predictions fit in the [0, 1] interval; and, finally, the presence of non-linear relationships close to the interval boundaries.

A better alternative would be to assume a geometric (or exponential) growth over time. In this case, the annual growth can be estimated applying the *compound annual growth rate* (CAGR) formula:

$$\text{CAGR}_{Ai} = \left(\frac{x_{it}}{x_{it_0}}\right)^{\frac{1}{t-t_0}} - 1$$

The required CAGR to achieve the target is:

$$\text{CAGR}_{Ri} = \left(\frac{x_i^*}{x_{it_0}}\right)^{\frac{1}{T-t_0}} - 1$$

The comparison between observed and required growth, $\text{CR}_i = \text{CAGR}_{Ai}/\text{CAGR}_{Ri}$ guides the assessment. Values of CR_i close to 1 indicate that the country i is "on track" to reach the target, while negative values, denote that the SDG indicator is moving away from the target. The approach based on estimation of CR_i is adopted by [5,10,12]. In practice, the geometric growth over time corresponds to assuming a linear growth for the log-transformed values of the indicator. To present a summary assessment, Eurostat (see [5]) suggests to categorize the values of CR_i in four classes: $\text{CR}_i < 0$ denotes a "movement away from the target"; $0 \leqslant \text{CR}_i < 0.6$ indicates an "insufficient progress towards the target"; $0.6 \leqslant \text{CR}_i < 0.95$ stands for "Moderate progress towards the target"; and finally $\text{CR}_i \geqslant 0.95$ indicates a "Significant progress towards the target". It is worth noting that Eurostat applies this approach to the EU (EU-28) as a whole and not to the single countries.

UNESCAP (see e.g. [8]) assumes also a geometric growth over time, but the compound growth rates are calculated by considering all the values in the time series (instead of just two data points, the initial and final value) by means of a weighted geometric mean. The weights decrease over time so to give more importance to the most recent values (for details see [8]). This approach corresponds to fitting a weighted linear regression model to the log-transformed values. The estimated annual growth rate is used to get an explicit prediction of the value of the indicator in the target year (2030) and to establish whether the region will reach the target.

The UN 2020 progress chart (see [10]) is based on the same approach suggested by Eurostat and, similarly, the values of CR_i are categorized in four classes but with different thresholds: in the UN progress chart $\text{CR}_i < -0.1$ indicates "Deterioration"; "Limited or no progress" is associated to $-0.1 \leqslant \text{CR}_i < 0.5$; "Fair progress but acceleration needed" is considered when $0.5 \leqslant \text{CR}_i < 0.95$; and, finally, $\text{CR}_i \geqslant 0.95$ indicates "Substantial progress/on track". The Progress Chart however covers only 36 SDG indicators and assesses the trend only at regional and global level; for some indicators the thresholds used for assessing the CR_i are not the standard ones.

The UNESCAP approach overcomes the main criticism addressed to the methodology that estimates the annual growth rate by just considering two values (the value of the baseline year and the latest value available) and not taking into account of what happens between

disseminate SDG data globally; provide technical support to support the capacity of countries to monitor the SDGs; and produce global monitoring reports.

them. In practice, the estimation of both LAGR_{Ai} and CAGR_{Ai} corresponds to drawing a linear regression line between two data points (the compound growth rate corresponds to drawing a line between the log-transformed values), and this approach makes sense only in the presence of relatively short time series. On the contrary, with relatively long time series, it would be preferable to estimate the slope of the regression line fitted across all the available data points (values vs. time in the linear case; and log-transformed values vs. time, in the case of geometric growth). In case of very long time series, it is advisable to adopt a weighting system as in the exponential smoothing approach, which gives progressively lower weighs, the more the observations are distant from the latest period.

Despite the fact that fitting a regression model allows using all information available, the estimated model may fit poorly and provide not fully reliable results in the presence of short time series and high random variation. Moreover, the presence of outliers, as well as the presence of missing values between the start and the end of the time series, may affect the estimation process. The presence of missing values within the reference period does not pose problems with the two-point estimate, which however maybe heavily affected by the presence of an outlier at the beginning or at the end of the series. The presence of outliers can be treated by adopting estimation techniques such as the M-estimations or other related methods (see e.g. [13]). In alternative, the slope of the regression line can also be estimated using the Sen's *slope* (see [14]), a robust nonparametric approach commonly associated to the Mann-Kendall test, used to detect the presence of a significant monotonic trend.

As fitting statistical methods to explicit forecast the value of the SDG indicators for the year 2030 is not a simple task, the practitioner before applying these techniques, should consider the nature of the SDG indicator, along with the length of time series and the variability of the data. For instance, when the SDG indicators are proportions, a case quite common as we have seen, then the forecasting methods should ensure that the predictions fall in the [0, 1] interval. Sometimes, an appropriate transformation can solve the problem; in the case of proportions, for example, the *logit* transformation is recommended ($y_{it} = \log(x_{it}/(1 - x_{it}))$). In alternative, more sophisticated models should be considered, like the weighted least squares models (WLS) or better, the beta regression.

All the above-mentioned problems, related to the choice of the underlying model, the need of data transformations and the potential impact of outliers, can be avoided by following the nonparametric approach suggested by the OECD (see [1]). In fact, the objective of the assessment is just aimed at checking for the presence of a monotonic trend over time of the indicator of interest and then to verify whether this trend follows the normative direction. Basically, the OECD suggests estimating the Spearman's rank correlation coefficient (i.e. the correlation coefficients between the ranks associated to the SDG values and the ranks of the time variable) and then test whether it is significantly different from 0. In addition, some criteria are established to assess whether the trend can be classified in one of the following categories: "progress towards the target", "movement away from the target", or "no significant trend is detected" (for details see [1, p. 137]). This approach is appealing because avoids making assumptions on the type of growth (linear or geometric), is completely nonparametric, is not influenced by the presence of outliers and, finally, it does not require transformations of the data (is unaffected by monotonic transformations of the data). It however requires a minimum number of observations in the time series to get a reliable estimate of the Spearman's correlation coefficient (the test requires a minimum of five values in the time series, but it would be better to have at least 10). Unfortunately, the suggested test suffers the presence of serial correlation in the data that may determine the rejection of the null hypothesis of absence of trend when a trend is actually present ("type I error"). The serial correlation of the values of the SDG indicator cannot be ignored, especially when the time series may include imputed values obtained through the carry forward method (replication of previously available value for a number of years). To compensate for this problem a few solutions have been suggested. A viable option could be the application of the Mann-Kendall's test, modified to account for serial correlation in the data, to reduce the risk of "false positives" (see e.g. [15]). It is worth noting that when the OECD method detects a monotonic trend that goes in the desired direction, it does not provide any information about the possibility that the observed progress will be sufficient to ensure that the target can be reached.

3.2.1. Assessing the trend of indicators without a numerical target

In order to compare the actual growth with the required growth to reach the target, a numerical target is needed. Similarly, all the approaches based on explicit forecasting, need to compare the 2030 forecast with the target. In absence of a numerical target,

the SDSN/Bertelsmann Stiftung report [4] and UN-ESCAP [8] use the same procedures shown in Section 3.1.1 to determine "statistic targets". Eurostat, on the contrary, in the absence of a quantified value for the target, suggests only to assess whether the estimated growth goes in the desired direction and how fast (see [5, pp. 26, 364]). The Eurostat's approach is also used in the UN 2020 progress Chart that introduces four criteria for judging the actual growth (see [10]). Finally, the OECD approach that carries out a test for the presence of a significant trend has the advantage of being applicable to all the indicators, no matter if they have a target or not.

3.2.2. Aggregation issues

The UNESCAP's approach for assessing progress towards the SDG target is designed for being applied at regional level, although it may work also at country level. For the other approaches presented in Section 3.2., the assessment of the trend at regional/global level poses the typical dilemma of forecasting: should the regional forecasts be obtained by aggregating countries' forecasts or should regional forecast be obtained directly by fitting the model to the time series of regional level data? The first option guarantees the consistency of forecasts at the various aggregations steps, but has the tendency to give poor aggregated forecasts due to the low signal-to-noise ratio, i.e. the fact that the underlying signal is masked by a high random variation (noise) in the single country time series data, usually due a high natural variation of the observed phenomenon. From this point of view, the need of a reliable assessment of the regional status favors an assessment of the trend by working directly with time series of the SDG data at regional level. This is the approach adopted in drafting the UN progress Chart (see [10]) that however does not extend the assessment at country level. Obviously, the assessment of the trend at regional level does not provide any information of the variety of the different trends shown by the countries belonging to the same region. For this reason, the OECD prefers to summarize the progress at regional/global level by estimating the frequency of the outcomes of the trend test, i.e. counting how many countries in the region are "moving away from the target" etc. (see e.g. [1, pp. 33–36]). Eurostat (see [5]), on the contrary, carries out its trend assessment only at EU-28 level and not for all the contributing countries.

The SDSN/Bertelsmann Stiftung report [4] adopts a weighted average to summarize at regional level the results of the trend assessment at country level (actual vs required growth) using country population as the weighting variable. In practice, however, it not clear if the weighted average is applied to LR_i or to the categorized values of LR_i.

Simple averages are used in [4] to aggregate of the outcomes of trend assessment at target or goal level, i.e. the overall contribution of the indicators to the target or the goal. However, the method lacks clarity on the modality of its implementation. Similarly, Eurostat applies a simple average of the trend assessed for the various indicators when the fraction of assessed indicators falling under a given goal is greater or equal to 75%. OECD [1], on the contrary, estimates the frequency distribution of the three outcomes of the test for all the countries and for all the indicators under a specific goal.

Adopting averages and similar methods, without an assessment of the heterogeneity of the targets and variability of the indicators may provide a biased assessment of the aggregates. In this regard, only the OECD approach provides, together with the aggregates, a summary representation of the different characteristics of the components making up the aggregates.

In general, the aggregation at goal level may produce rather questionable results, especially when the various contributing indicators are rather heterogeneous and measure quite different phenomena, although often these aggregations are mainly done for providing a general assessment tool for communication purposes, rather than to guide the decisions of the policy makers.

4. A proposal for assessing both the level of achievement and the progress over time

In our view, a comprehensive SDG progress report should include an assessment of both the level of achievement associated to the latest available data, as well as of the progress made to achieve the target. A country that, for example, is very close to the target, may not be in a favorable condition to achieve it if, in recent years, the trend shown is too slow or if, on the other hand, the indicators has moved in the opposite direction to the one required to reach the target. Assessing therefore the possibility to reach the target solely based on the latest level of achievement can provide a biased picture, even when, for instance, the distance is very small.

The methodological approaches used for assessing the latest level of achievement of the indicator, presented in the Section 3.1, seem all focused on calculating a sort of distance to the target. In this

sense, our suggestion is to follow the proposal of the SDSN/Bertelsmann Stiftung report [4], but working directly with a distance measure, i.e.:

$$d_{it} = 1 - x'_{it},$$

where the outliers, i.e. country values that are far from the mean (or median) of their distribution, are removed before estimating the range. In our view, the standardization may not be necessary for indicators expressed as proportions (whose theoretical range is equal to 1). When however the "observed" range of a proportion is much smaller than the theoretical one, the standardization of the indicator by the estimated range ($R = \max - \min$ of the available values), may provide better results.

This distance d_{it} can be calculated only for indicators with an explicit numerical target. In absence of it, in our view, using procedures to estimate "statistics targets" can be done only when it is possible to provide a valid and sound rationale for this decision. It is important to note, however, that setting targets is typically the role of policy-makers, not of statisticians, whose role is to identify the best metrics to monitor those targets and then analyze and interpret the data that they collect and publish. By setting "statistical targets", in absence of policy targets, the statisticians are overstepping the boundaries with the role of legislators. Statisticians therefore should avoid setting numerical target and avoid doing any assessment of the latest level of achievement when a policy target is absent.

A practical solution to measure the level of achievement in absence of numerical targets could consist in calculating summary statistical information on the worldwide distribution of country latest available values, like, for instance, the five-number summary (min, P25%, median, P75%, max) or, in alternative, the quintiles of the distribution (min, P20%, P40%, P60%, P80%, max).

From a country perspective, obviously, it is always possible to assess its distance to the target for a specific SDG indicator, even when the target is not explicitly defined, by setting up its own target according to its policy priorities and past progress over time.

The assessment of the distance to the target at region/global level should be preferably performed using simple averages or medians, avoiding the application of weighting schemes. Since the implementation of the SDGs is mainly a national responsibility, global/regional targets can be considered achieved when the large majority of countries have reached this objective. Considering the global/regional targets reached when the majority of the population living in these territories have achieved it, as implied by weighting with the country population, have a completely different meaning from the previous approach. It should be noted, in this respect, that weighted outcomes at global/regional levels tend to be dominated by a few countries with large population. The regional averages, whenever possible, should come along with an indication of the variability of the distribution of country values (e.g. five-number summary, etc.), to provide a summary picture of the differences between the countries belonging to the same region. The same considerations are also applicable when considering the entire world. Any assessment of the distance to the SDG targets should be based on a unweighted mean to avoid subjective choices that give more importance to a specific indicator, despite other important indicators monitor the same target.

Progress assessment analyses, in order to provide sufficiently robust results, should consider: (i) the type of SDG data (ratio, proportions, scores, etc.); (ii) the length of time series; and (iii) the amount of random variation in the time series. As mentioned before, even in the best scenario, the time series start in 2015 for most of the indicators and the number of data points currently available is not more than 2 or 3. In this context, we propose to postulate that the SDG indicators follow a geometric growth over time (hypothesis that applies to many phenomena) and then estimate the growth rate considering the first and the last available data point in the time series (CAGR_{Ai}). As showed, in the Section 3.2, the actual growth rate should be compared with the desired growth rate to reach the explicit target ($\text{CR}_i = \text{CAGR}_{Ai}/\text{CAGR}_{Ri}$). In absence of an explicit numerical target, following the suggestion by Eurostat, the assessment can be based solely on the judgment of the estimated CAGR_{Ai}. This approach works for SDG targets that can be considered as continuous variables (interval or ratio type variables), including proportions. On the contrary, for SDG targets expressed as scores (typically indicators that measure the "degree of implementation" of a certain policy, as for instance SDG indicator14.6.1), with a natural ordering among the categories (categorical ordinal variables), a more suitable approach for the trend assessment seems the one proposed by OECD: in particular, the OECD proposal consists in applying a trend test, based on the Spearman's rank correlation coefficient or, better, the Mann-Kendall test modified to account for serial correlation. Unfortunately, both tests require at least 5 observations in the time series. With fewer observations (from 2 to 4) the only way to conduct a trend assessment would

Table 1
Distance to the target for the PoU at regional and global level in the year 2017

Region	No. of Countries	Value 2017	Dist. to target	Equal to targ.	Type	Summary of distances			
						Mean	Median	P25%	P75%
Australia and New Zealand	2	0.025	0	100.0%	*Dist.*	0	0		
					Rel. dist.	0	0		
Central Asia and Southern Asia	12	0.145	0.120	8.3%	*Dist.*	0.0863	0.0635	0.0358	0.1205
					Rel. dist.	0.1815	0.1336	0.0752	0.2536
Eastern Asia and South-eastern Asia	17	0.087	0.062	23.5%	*Dist.*	0.0934	0.0680	0.0070	0.1090
					Rel. dist.	0.1964	0.1431	0.0147	0.2294
Latin America and Caribbean	28	0.065	0.040	10.7%	*Dist.*	0.0718	0.0545	0.0225	0.0768
					Rel. dist.	0.1510	0.1147	0.0473	0.1615
Northern America and Europe	40	0.025	0	82.5%	*Dist.*	0.0028	0	0	0
					Rel. dist.	0.0058	0	0	0
Sub-Saharan Africa	40	0.227	0.202	0.0%	*Dist.*	0.2035	0.1700	0.0860	0.2723
					Rel. dist.	0.4224	0.3577	0.1810	0.5729
Western Asia and Northern Africa	19	0.098	0.073	15.8%	*Dist.*	0.0655	0.0200	0.0060	0.0695
					Rel. dist.	0.1378	0.0421	0.0126	0.1463
World	*165*	*0.108*	*0.083*	*27.9%*	*Dist.*	0.0867	0.0430	0	0.1200
					Rel. dist.	0.1812	0.0905	0	0.2525

"Oceania, excl. Australia and New Zealand" is not reported because there are insufficient country data to get a regional estimate.

be to directly compare the initial and latest available scores, setting criteria to classify all the various possible combinations, so to reproduce a classification of 3–4 relevant outcomes ("moving away from the target", "stagnation", etc.) as proposed by most of the leading agencies.

The strategy for assessing progress over time can obviously be improved in the future when the time series will become longer and include more data points. In this case the CAGR_{Ai} can be assessed considering different baselines or replaced by an estimate of the slope of a linear regression model fitted to log-transformed values, or the Sen's slope (estimated from log-transformed values too).

The assessment of the progress over time at regional level should preferably be carried out using the regional time series of the SDG data. This assessment should be accompanied with a comparative measure of the differences between the trends of the countries belonging to the same region (i.e. five-number summary of the values of CR_i or CAGR_{Ai} or the estimated frequency distribution of their specific categorization).

5. An application to the SDG indicator 2.1.1

This section shows an application of the proposed approaches to real SDG data. The indicator chosen is one of the SDG indicators under FAO custodianship, the "Prevalence of Undernourishment" (PoU or SDG indicator 2.1.1). It is an estimate of the proportion of the population in a country whose habitual food consumption is insufficient to provide the dietary energy levels required to maintain a normal active and healthy life. The PoU is estimated for almost all FAO Members on an annual basis,[4] starting from the year 2000 up to year 2017.[5] The distribution of the habitual dietary energy intake at country level generally follows the log-normal distribution (sometimes skewed lognormal or skewed normal) and the estimation of its parameters[6] allows to derive the PoU as the probability that the habitual dietary energy intake is below the minimum dietary energy requirements of a representative average individual.

The numerical target for the PoU stems out directly from the formulation of the SDG target 2.1 ("By 2030, end hunger [. . .]") and is equal to 0. In practice, the threshold used for the analysis is set equal to 0.025, mainly because the uncertainty in the estimation process is such that an estimate of the PoU less than or equal to 0.025 cannot be said significantly different from 0. In addition, a target value equal to 0 would create problems when estimating the required growth under the assumption of geometric growth over time (CAGR_{Ri}).

[4] In 2017 the PoU is estimated for 165 countries. Note that the annual estimate for PoU at country level is in practice a three-years moving average centered on the reference year.

[5] http://www.fao.org/faostat/en/#data/FS, data accessed at the end of April 2020.

[6] The average dietary energy consumption, derived from the Food Balance Sheets; the coefficient of variation of the food consumption among the population, derived from the Household budget surveys: and the dietary energy requirements, derived by the characteristics of the population by sex and age. For more details see Annex 1B in [16].

Table 2
Summary results of trend of the PoU over time with different baseline years

Description		Baseline year	
		2015	2010
No. obs.		3	8
Values	Baseline year	0.106	0.118
	In 2018	0.108	0.108
	Target (2030)	0.025	0.025
Geometric Growth *(linear on log-transf. data)*	*Required growth*	*−0.0918*	*−0.0747*
	Actual growth 2 obs.	0.0093	−0.0126
	Actual growth ALL obs.		−0.0147
	Sen's slope		−0.0173

The Table 1 provides a summary assessment of the level of achievement of the target associated to the latest available data at country level (year 2017). The column "dist. to target", reports simply the difference between the regional estimate of the PoU and the target (0.025); the subsequent column shows the percent of countries in the region already reaching the target in 2017. Finally, the last four columns in the Table 1 provide summaries of the distances to the target for countries belonging to each region. The rows denoted as "dist." refer to summaries of the distance to the target (distance scaled by the theoretical range for proportions, i.e. 1) while "rel. dist." refers to the relative distances, where the distance is scaled with the observed range of PoU for the available 165 countries (range = 0.475, estimated excluding 1% of observations in the right tail of the distribution that are found to be outliers).

Table 1 shows that the Northern America and Europe region has a 0 distance to the target, but there are 17.5% countries in the region that have not reached the target, although they are very close to it. The Latin America and Caribbean regional estimate of PoU is not far from the target but only 10.7% of the countries in the region have already reached the target; however, 3/4 of them show a distance to the target below 0.08. In relative terms, this fraction of countries has a relative distance to the target below 0.17 (the maximum relative distance is equal to 1). Finally, at World level, the countries having already reached the target are about the 28% of the total and an additional 20% of countries show a relative distance to the target below 0.1 (the third quartile (P75%) of the relative distance corresponds to about 0.25).

Table 2 summarizes the results of the analysis of the trend over time for the PoU at global level ("World") with the approach suggested in the Section 4 (Fig. 1 shows the estimated PoU at global level from 2010 to 2017) whose objective is to understand whether the indicator target can be reached by 2030. By setting

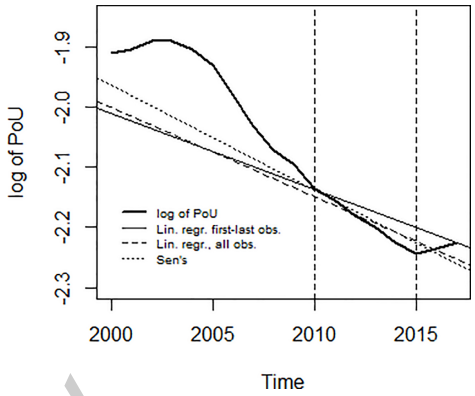

Fig. 1. Estimated PoU at World level (log-transformed values) with fitter regression lines when the baseline is the year 2010.

the baseline to year 2015, the year in which the 2030 Agenda was endorsed, the time series consist of just 3 values and this allows only to estimate the CAGR_A and compare it with the required growth to reach the target in 2030; the resulting ratio, $\text{CR} = -0.1$, indicates that globally the situation is moving away from target (direction opposite to the normative one). However, by checking the distribution of the CR ratio at country level, apart the 46 countries having already reached the target (28%), only 13 of them can be said "on-track" to reach the target in 2030 (8% with $\text{CR}_i \geqslant 0.95$); in addition, 25% of countries is moving in the desired direction, but at a pace too slow to reach the target by 2030 ($0.1 < \text{CR}_i < 0.95$); 16% of countries shows a stagnating situation ($-0.1 \leqslant \text{CR}_i \leqslant 0.1$) and, finally, the remaining 24% is moving away from the target ($\text{CR}_i < -0.1$).

In order to assess the results obtained with other possible estimation methods that work with longer time series, we have decided to move back the baseline to year 2010 (8 observations). Now the actual growth can be estimated by fitting a linear regression model to the log-transformed values or applying the Sen's method so to have a more robust estimate of the slope of this linear model.

The results of the estimation of the slope of the regression line, when the baseline is set back to 2010, are represented graphically also in Fig. 1, which also shows the whole time series of the log-transformed data of PoU at World level.

With a time series of 8 values, the estimated growth with the three different methods shows the same sign but results are slightly different. Generally speaking, in this case it is preferable to use all the observations in the time-series, instead of just the first and the last data point, unless the time series shows a very high random variability.

6. Conclusions

Monitoring the implementation of the SDG targets is a fundamental step for the implementation and achievement of the 2030 Agenda. As described in this article, different leading international organizations use different methodological approaches for this monitoring exercise. For this reason, these monitoring reports produce rather different results, sometimes even in contradicion one with another results, creating puzzlement and uncertainty among non-expert users and policymakers. An international agreement on what should be measured and which standard methodologies should be used for this purpose have not been reached yet. In addition, the approaches adopted to provide summary pictures at aggregated level (aggregation of geographical areas or per target/Goal) often have privileged the interpretability of the results, at the risk of providing a misleading picture of the situation, which is quite high when, for instance, we are aggregating the results in the presence of a high variability of the elementary indicators or when the aggregation is influenced by a couple of outliers.

In our view, a comprehensive SDG progress report aimed at monitoring the implementation of the 2030 Agenda should include an assessment of the level of achievement of the target, as well as of the progress made towards the target. Our proposal goes in the direction of using simple tools, easy to be implemented in different conditions, avoiding the adoption of complex statistical models with many parameters to be estimated. As stressed many times, a critical step is the aggregation of results at both geographical level and target/goal level. In our view, any geographical aggregation should avoid the adoption of weighting schemes because, as already stated in Section 2, the implementation of the SDGs is mainly a national responsibility and consequently global/regional targets can be considered achieved when the large majority of countries have reached this objective. More generally, the researchers preparing SDG reports should try to improve the quality of the assessments by accompanying the averages calculated at aggregated level with measures of the associated variability, while remaining easy to be interpreted by all the potential users of the report.

Acknowledgments

The authors wish to thank the anonymous reviewers for their valuable comments, which allowed to improve the clarity of this article.

References

[1] OECD. Measuring distance to the SDG targets 2019: An assessment of where OECD countries stand. Paris: OECD Publishing. 2019.
[2] UN. Lessons Learned from MDG monitoring from a statistical perspective. Report of the Task Team on Lessons Learned from MDG Monitoring of the IAEG-MDG. 2013; Available from: https://unstats.un.org/unsd/broaderprogress/pdf/Lesson%20Learned%20from%20MDG%20Monitoring_2013-03-22%20(IAEG).pdf.
[3] Gennari P, Navarro Kalamvrezos D. The challenge of measuring agricultural sustainability in all its dimensions. J Sustain Res. 2019; 1.
[4] Sachs J, Schmidt-Traub G, Kroll C, Lafortune G, Fuller G. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN). 2019.
[5] Eurostat. Sustainable development in the European Union, Monitoring report on progress towards the SDGs in an EU context (2019 edition). Uxembourg: Publications Office of the European Union. 2019.
[6] Craw S. Manhattan Distance. in: Encyclopedia of Machine Learning Boston: Springer. Sammut C, Webb GI, editors. 2011.
[7] Rousseeuw PJ, Croux C. Alternatives to the median absolute deviation. Journal of the American Statistical Association. 1993; 88(424): 1273-1283.
[8] Bidarbakhtnia A. A weighted extrapolation method for measuring the SDGs progress. ESCAP Statistics Division Working Paper Series, SD/WP/04/March. 2017.
[9] Bidarbakhtnia A. Measuring sustainable development goals (SDGs): An inclusive approach. Global Policy. 2020; 11(1): 56-67.
[10] UN. Sustainable Development Goals Progress Chart 2020 - Technical Note. New York, United nations. 2020; Available from: https://unstats.un.org/sdgs/report/2020/Progress_Chart_2020_Technical_note.pdf.
[11] UNESCAP. Asia, 2020 TPSPR. Bangkok: United Nations Publication. 2020.
[12] World Health Organization and the United Nations Children's Fund (UNICEF). Methodology for monitoring progress towards the global nutrition targets for 2025. Technical Report WHO-UNICEF Technical Expert Advisory Group on Nutrition Monitoring (TEAM). 2017.
[13] Huber P. Robust Statistics, 2nd Edition. Hoboken: John Wiley & Sons. 2009.
[14] Sen PK. Estimates of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association. 1968; 63(324): 1379-1389.
[15] Hamed KH, Rao AR. A modified Mann-Kendall trend test for autocorrelated data. Journal of Hydrology. 1998; 204(1): 182-196.
[16] FAO, IFAD, UNICEF, WFP, WHO. The State of Food Security and Nutrition in the World 2019. Safeguarding against economic slowdowns and downturns. 2019.