POLICY RESEARCH WORKING PAPER 8843

Tracking the Sustainable Development Goals

Emerging Measurement Challenges and Further Reflections

Hai-Anh H. Dang
Umar Serajuddin

Abstract

The Sustainable Development Goals recently adopted by the United Nations represent an important step to identify shared global goals for development over the next two decades. Yet, the stated goals are not as straightforward and easy to interpret as they appear on the surface. Review of the Sustainable Development Goals indicators suggests that some further refinements to their wordings and clarifications to their underlying objectives would be useful. This paper brings attention to potential pitfalls with interpretation, where different evaluation methods can lead to different conclusions about country performance. The review of the United Nations' Sustainable Development Goals database highlights the overwhelming challenge with missing data: data are available for just over 50 percent of all the indicators and for just 19 percent of what is needed for comprehensively tracking progress across countries and over time. The paper offers further reflections and proposes some simple but cost-effective solutions to these challenges.

Tracking the Sustainable Development Goals:
Emerging Measurement Challenges and Further Reflections

Hai-Anh H. Dang and Umar Serajuddin[*]

Key words: SDGs, monitoring, data challenges, survey data, international organization

JEL: F00, I3, O1

I. Introduction

We currently live in a globalizing world, where countries are becoming increasingly integrated beyond their physical border. For example, international trade as a share of the world's GDP increased from around a quarter in the 1960s to almost 40 percent in the 1980s, and has steadily climbed up to more than half since 2000 (World Bank, 2019). Environmental issues such as water and air pollution are not exclusively restricted to a country, but oftentimes affect its neighbors as well. The advance of technology has also significantly reduced the barriers posed in the past by physical distances, and rendered communications costs to a fraction of what they used to be.[1] It comes as no surprise that present-day challenges that are commonly faced by countries can be best solved by the joint efforts of the global community, rather than those of any single country.

The Sustainable Development Goals (SDGs) were adopted by the United Nations (UN) in 2015. They represent a major step by countries—poorer and richer alike—to identify shared global goals for development over the next 15 years. The SDGs build on the success of their predecessor, the Millennium Development Goals (MDGs), but the new goals are more ambitious and the indicators for measuring progress more comprehensive. The SDGs were also designed to be more realistic, with a view toward avoiding targets that are either implausible to measure or unfair for certain groups of countries (as pointed out with the MDGs).[2] Yet, the SDGs and their indicators may not be straightforward and easy to interpret as they appear on the surface. Many targets are complementary, but some may be contradictory. Complicating matters further, the data available are inadequate to measure progress.

---

[1] Multinational companies such as Facebook have discussed plans to provide free Internet connection to rural and hard-to-reach communities across the world (see, e.g., Vanian, (2016)).

[2] A major criticism of the MDGs is that these goals are unfavorable to low-income African countries (see, e.g., Clemens, Kenny, and Moss (2007), Easterly (2009), and Waage et al. (2010)).

We discuss in this paper various challenges related to the SDGs in terms of identification and measurement methods, and we place a special emphasis on the quantitative analysis angle.[3] We examine global data but focus on poorer countries. In particular, we revisit basic questions that are often assumed as settled, but that on more careful thought may actually lead to different answers. For example, is the current list of SDG indicators appropriate for the stated development goals? Is there any overlap or potential conflict among these goals? Could progress on some indicators be interpreted in one way only, or more than one way? If the latter, how best to interpret progress? Should we evaluate a country's performance as far as we can go back with the available data, or only within a recent time period? Since it is well-known that data are unavailable for a number of indicators, could this missing data challenge affect how we evaluate a country's performance? Also given this missing data issue (and other potential interpretation issues), what viable alternatives are there to monitoring progress on all 232 indicators: could we instead track a core group of indicators or an overall index? While we do not claim to raise all these questions for the first time (see our discussion on the relevant literature in later sections), we aim to shed new light and offer more reflections on them.

We combine our own data analysis with insights gleaned from related studies to enrich our discussion. We analyze data from various data sources, including the World Bank World Development Indicators (WDI) database, the UN SDG database, as well as the nascent (but growing) literature on the SDGs. Building on the analysis and taking into account the political settings of the SDGs, we then offer further reflections on future data collection efforts for the SDGs and other related issues. In short, we aim to expose deeper issues hidden behind the various

---

[3] For recent discussions on other substantial aspects of the SDGs, see, for example, Joshi et al. (2015) on governance, Waage et al. (2015) on health indicators, and Klopp and Petretta (2017) on urban development. See also Fukuda-Parr (2019) for a recent discussion on the politics of selecting the inequality measure for the SDGs.

indicators that—if superficially treated with casual thoughts—can interfere with monitoring progress toward the SDGs. We do not claim that we could offer answers to the all challenges that we identify, but we strive to make a contribution to refining the SDG indicators and advance their monitoring. We add to the nascent, but growing literature on measurement issues with the SDGs, by discussing jointly the various identification, interpretation, and data challenges with the SDGs.[4] In addition, we also offer a technical review of common approaches to measuring country progress, most of which have been applied to the MDGs.

Our findings suggest that several SDG indicators could be refined in terms of their wording, and their underlying objectives could be further clarified. We also point out potential pitfalls in interpreting progress on the SDGs, since different evaluation methods can lead to different conclusions. Furthermore, our review of the UN's SDG database also illuminates the overwhelming challenge with missing data. Even when we consider the relatively more comprehensive data from the most recent five years, there are data only for just over half of all the SDG indicators, and just 19 percent of the data needed to thoroughly track the progress across countries and over time. We subsequently propose some simple but cost-effective solutions to these challenges.

This paper consists of five sections. We provide a brief introduction to the SDGs in the next section before discussing measurement challenges in Section III, which range from identification of appropriate indicators to interpretation of their progress, and various related data challenges. We subsequently offer in Section IV further reflections, as well as some cost-effective solutions; here we also briefly discuss how international organizations can help address these challenges. We

---

[4] Earlier studies that discuss measurement issues include, for example, MacFeely (2018), Fukuda-Parr and McNeill (2019), and Ordaz (2019).

finally conclude in Section V. To avoid distracting from the measurement and policy issues, we leave all the technical details to Appendix 1.

II. Overview of the SDGs

In September 2015, at the UN, the Heads of State and Government adopted the 2030 Agenda for Sustainable Development to set all the countries in the world on a common path towards sustainable development. This agenda consists of 17 Sustainable Development Goals (SDGs) and 169 targets aimed at quantifying shared global development for the next 15 years in its various aspects, covering social, economic, and environmental dimensions. In 2017, the UN General Assembly anchored these goals and targets in 244 indicators, which have been subsequently revised to the current list of 232 indicators (United Nations, 2018a).

The SDG indicators are classified into three tiers, depending on a combination of their underlying methodology and data availability: those with an established methodology and good data coverage (Tier 1), those with an established methodology but lack good data coverage (Tier 2), and those that lack both (Tier 3). Notably, the tier classification can change as new methodologies are developed and data become more available. As of April 2019, there are 101 Tier 1 indicators, 91 Tier 2 indicators, 34 Tier 3 indicators, and six indicators with multiple tiers (United Nations, 2019). In addition, there are six indicators that belong to multiple tiers as different components of these indicators are classified under different tiers. This tier classification also implies that we only have some data coverage for at most around 80 percent of all the indicators (which are in Tier 1 and Tier 2 categories).

The SDGs are unique in that they represent a first endeavor of its type by world leaders to combine such diverse national and global development policies in a collaborative partnership.

While the SDGs can be regarded as a successor to the Millennium Development Goals (MDGs)—which were adopted in 2000 for the target year of 2015—the former are far broader in scope and deeper in reach. Indeed, the MDGs consist of eight goals and 60 indicators, less than half the SDGs and a quarter of its indicators; the number of the targets of the former is 18, one-ninth that of the latter. The SDGs also introduced for the first time such goals as reduced inequalities and sustainable urbanization. We return to further discussion on the political settings of the SDGs in Section IV.

III. Emerging Measurement Challenges

While the richer nature of the SDGs represents progress, it also gives rise to certain measurement challenges. We discuss in this section challenges related to identification, interpretation of progress, and data availability.

III.1. Identification of Goals and Indicators

Overlapping Indicators

The 232 indicators of the SDGs address various aspects of societal development. Since the SDGs have far more indicators than the MDGs, they might be expected to produce a more comprehensive picture of the multi-dimensional and complex development process. Yet, concerns have been raised over whether there should be fewer indicators to better focus on priorities, or whether they should at least be refined for better management (Easterly, 2015; Economist, 2015). Indeed, given the magnitude of this global undertaking, efforts should be made to minimize inaccuracies with the goal-setting process at the starting point. Any overlap among the goals can lead to confusion in communications, and even worse, duplication and inefficiencies with monitoring efforts down the road. The following examples illustrate the problem.

We provide in Table 1 several (non-exhaustive) examples of SDG indicators that overlap. For example, indicators 1.5.2 and 11.5.2 both aim to measure the direct disaster economic loss in relation to the global gross domestic product (GDP). Similarly, indicators 11.7.2 and 16.1.3 both track the proportion of persons who are victims of physical or sexual harassment. But standing out even more, the wording in indicator 4.7.1 is identical to that of indicator 12.8.1. The objectives of both these indicators are the same: measuring the extent that global citizen education and education for sustainable development are mainstreamed in the national education policies, curricula, teacher education, and student assessment.

These examples support the notion popular among many development practitioners that the list of SDG indicators can benefit from further refinement. In fact, as adopted by the UN in 2017, the list originally consisted of 244 indicators, which had as many as nine groups of identical indicators. A year later, the UN refined this list to remove duplications, which results in the current list of 232 indicators.[5] Thus our findings suggest that the current list can be refined further, including whittling out overlapping entries.

Potentially Conflicting Goals

The solution to the issue of overlapping indicators discussed above may simply be to comb through the list to reorganize or combine these indicators. In many cases, perhaps we can just use more appropriate wordings. But a more substantive issue is that some goals may conflict with one another.[6] In other words, this situation may require a far more substantial re-thinking of the underlying goal itself. For example, how should we evaluate a country's performance if this

---

[5] These duplicate indicators are shown on the UN's official website https://unstats.un.org/sdgs/indicators/indicators-list/.

[6] Kanbur, Patel, and Stiglitz (2018) point out a related challenge with SDG number 17 that it may be interpreted as a potential catch-all goal consisting of disparate components that do not form a coherent whole.

country manages to reduce poverty (Goal 1) but also increases inequality (Goal 10) at the same time? Or what is the trade-off between strong economic growth (Goal 8) and deteriorating environmental quality (Goal 11)? The next two examples illustrate such situations.

Table 2 presents recent estimates on the dynamic changes with poverty and vulnerability for 21 African countries, drawing with some modifications on Dang and Dabalen (2019), who analyzed household consumption surveys mostly implemented in the late 2000s. These countries together account for two-thirds of the population in Sub-Saharan Africa. We rank the countries in a decreasing order of reduction in the headcount poverty rate (column 3), and show the changes to the population shares of the vulnerable group and the middle class (columns 4 and 5). We also show the growth in the mean consumption levels for the bottom 40 percent of the consumption distribution (column 6) versus the growth for the whole distribution (column 7), which represents the World Bank's definition of shared prosperity (see, e.g., Basu, 2013; Jolliffe et al., 2015). This measure of shared prosperity is the first SDG indicator for tracking inequality reduction (Indicator 10.1.1). However, since concerns have been raised that this measure tracks only the (anonymous) growth of the bottom 40 percent, we also show for comparison another measure of shared prosperity that tracks dynamic changes of the population shares of the poor and the vulnerable groups. This latter measure is based on a simple typology of pro-poor growth scenarios recently proposed by Dang and Lanjouw (2016).[7]

---

[7] In particular, we consider three welfare categories that correspond respectively to the poor, the vulnerable, and the middle class groups. There are in total six possible growth scenarios depending on whether (the population share for) each of the three categories is expanding or shrinking. The most positive pro-poor growth scenario is one where both the poor and vulnerable categories decrease while the middle class expands (denoted by three stars in Table 2). The opposite happens with the worst pro-poor growth scenario (denoted by three minuses in Table 2) where both the poor and vulnerable categories expand while the middle class shrinks. The first three scenarios relate to the reduction of the poor group and thus indicate positive pro-poor growth, and the remaining scenarios suggest negative pro-poor growth. The growth of the vulnerable category helps further determine the rate of pro-poor growth, for example, whether pro-poor growth is more positive or simply positive. See Table 2.1 in Appendix 2 for further details.

Several remarks for Table 2 are in order. First, poverty reduction does not necessarily go hand in hand with inequality reduction. Indeed, among 14 of the countries (from Chad to Malawi) that could reduce poverty, inequality increases in one-half (i.e., Chad, Ghana, Mozambique, Ethiopia, Togo, Eswatini, and Malawi had less growth for the bottom 40 percent than the whole population). Several countries (i.e., Ethiopia, Togo, and Malawi) even witness their bottom 40 percent of the population having negative growth, while the whole population has positive growth on average.

Second, interestingly, the opposite situation also happened: increased poverty does not necessarily accompany increased inequality. For the remaining seven countries that suffered from worsened poverty, the majority of them (i.e., Burkina Faso, Zambia, Madagascar, Côte d'Ivoire, and Cameroon) actually managed to lessen inequality. Third, even within the group of countries that saw both poverty and inequality decreasing (or increasing), more progress on one front does not necessarily imply more progress on the other. For example, Sierra Leone ranked as number 10 in poverty reduction, but managed to achieve an impressive growth rate of 14.9 for its bottom 40 percent, which is 14 percentage points more than its overall growth rate (i.e., subtract column 7 from column 6). This places Sierra Leone as number 2, only after Botswana, in terms of keeping inequality down.

Why does this inconsistency between poverty and inequality exist? One answer is that, much depends on the relative position of a country's poverty line and its 40th percentile on the income distribution. As an example, if the poverty line is sufficiently below this 40th percentile cut-off point, there can be an increase in poverty together with growth in the average income of the bottom 40 percent—while the poorest are seeing a decline in incomes, those that are richer but still below the 40th percentile threshold could see their income rising significantly. In these circumstances, it

is unclear whether and how the concept of shared prosperity defined in this way would resonate within a country.[8]

The shared prosperity measure offered by Dang and Lanjouw (2016) provides a complementary angle by explicitly focusing on the poorer income groups. In particular, the 14 countries with lower poverty rate are considered as having positive pro-poor growth (i.e., marked with stars, column 8); similarly, the remaining countries with more poverty are regarded as not having pro-poor growth (i.e., marked with minuses, column 8). Furthermore, countries with less vulnerability but in the same group are judged as having better performance. For example, among the 14 countries with positive pro-poor growth, five with less vulnerability (including Mauritania, Ethiopia, Togo, Eswatini, and Malawi) are ranked as having even more positive growth. Of the seven countries with more poverty, five (including Burkina Faso, Zambia, Madagascar, Côte d'Ivoire, and Cameroon) managed to reduce vulnerability. If we combine this second measure with the first measure of shared prosperity, certain countries come out as consistently performing very well, like Botswana or Mauritania, or near the bottom, like Senegal and Nigeria.

We turn next to discussing the trade-off between economic growth and the quality of the environment. We plot in Figure 1 the global GDP per capita level against air quality, as measured by the number of micrograms per cubic meter (μg/m3) PM 2.5 using the WDI database for 1990 to 2016. Figure 1 indicates that all countries in the world as a whole almost tripled their GDP per capita during this period. While this good news should clearly be celebrated, it did not come

---

[8] Other issues exist as well. For example, why should the income threshold be set at the 40th percentile, rather than say, other percentiles such as the 30th percentile or 50th percentile? Since different countries can use different national poverty lines with their concomitant poverty reduction and social protection policies, why should a blanket bottom 40 percent apply globally? Furthermore, one might also ask why we do not consider growth of the income distribution as a whole instead (see also Jolliffe et al. (2015, chapter 5) for further discussion of this last point). As an alternative, we provide in Table 2.2 (Appendix 2) estimates using the Gini coefficients, which also show inconsistency between poverty and inequality. For example, of the same 14 countries that could reduce poverty, in almost half (i.e., Chad, Ghana, Uganda, Ethiopia, Togo, and Malawi) inequality as measured by the Gini coefficient worsened.

without a cost. The whole world also saw its air quality declining rapidly. Our results concur with those in a recent study, which suggests that by 2013 about 87 percent of the world's population were living in areas with polluted air that exceeded the World Health Organization's annual safety threshold of 10 micrograms per cubic meter PM2.5 (World Bank, 2016). Furthermore, the same study also estimates that exposure to ambient PM2.5 drove up premature mortality by 30 percent, from 2.2 million deaths to 2.9 million deaths per year between 1990 and 2013; moreover, the corresponding figure for global welfare losses rose 63 percent over the same period.[9]

While certain countries clearly demonstrate the trade-off between their economic growth and their environment quality, a few others stand out as a bright model that could manage to accomplish economic growth and clean up their environment at the same time. For a concrete example, China's spectacular economic growth has dramatically reduced poverty in the past decades, and has moved it up to the list of upper-middle income countries (World Bank, 2019). Yet, the country has witnessed its environment quality worsening considerably during this process. Figure 2 graphs the trends over the past 25 years of GDP against the PM 2.5 pollutant for both China and, for comparison, Norway. While China's GDP has solidly climbed up, its environmental quality has also steadily gone down; in fact, its air pollution appeared to be worsening at a much faster pace than its speed of economic growth. This stands in sharp contrast to the opposite pattern of improving both its economy and environment achieved by Norway. Even more worrisome is the fact that, China already reached an average PM2.5 air pollution of 56 micrograms per cubic

---

[9] A country's polluted air can cause (economic) damages to its neighboring countries as well. Zheng and Kahn (2017) observe that Hong Kong SAR, China, has paid manufacturers in its neighbor Guangdong Province, China, about $150 million every year to install pollution-reducing equipment to help prevent polluted air from coming into its territory. Another study by Jia and Ku (forthcoming) find more pollution in China to increase mortality from respiratory and cardiovascular diseases in neighboring districts in the Republic of Korea, with the most vulnerable being the elderly and children under five.

meter roughly at a GDP per capita level of $US7000 in 2016. This amount of air pollution is two-fifths higher than the global average for that level of income (Figure 1).

The solutions to these conflicting goals may not be straightforward, and may even place national priorities against global ones. In fact, we are not the first to make this point. For example, Bourguignon et al. (2010) already raised a similar question on whether a similar trade-off between the MDGs should be determined by countries or the international community. But perhaps the first step to solving any challenge starts with obtaining high-quality data that allow for comparison over time both within a country and across countries. Such data are needed to facilitate a well-informed decision-making process.

III.2. Interpretation of Progress
Which Metrics to Use for Evaluating Progress?

After identifying a final list of consistent indicators that do not overlap—which can lead to confusion in communications and inefficient monitoring efforts as discussed earlier—the next questions to ask naturally concern the best ways to interpret their changes. For example, should we just weight all indicators equally, such that we count up the number of positive changes to compare with the number of negative changes? Put differently, applying the same equal weight to all indicators is equivalent to using no weight at all. Or should we give more weight to certain indicators when aggregating them? This is a critical decision to make, since applying weights can lead to a different conclusion from not doing so.[10] If we use weights, how should these weights be

---

[10] Let us briefly illustrate this by comparing poverty estimates for India and the South Asia region. While poverty can be estimated for each state in India, for the all-India national poverty rate, we would usually want to report a population-weighted for the whole country. Yet, for South Asia, we may report two types of poverty numbers depending on the objectives. One type is the unweighted average of the poverty rates for all South Asian countries that gives equal weight to (poverty reduction progress for) each country. The other type is the population-weighted average for the region, which gives more weight to countries with a larger population size.

defined (e.g., arithmetically or geometrically)? Should we look at the changes in absolute numbers or the changes in relative numbers over time, or both (i.e., the absolute number of the poor in the population or the percentage)?

The importance of employing appropriate metrics to monitor progress cannot be overemphasized.[11] We turn to discussing some of these issues with concrete data. Table 3 provides the latest poverty numbers using the international poverty line of $1.9/day in 2011 PPP from the WDI database for four countries: Brazil, China, Ethiopia, and India (i.e., 2015 for the first three countries and 2011 for India). Judging by the headcount poverty rate (Table 3, Panel A, Row 1)—or the percentage of the population who are poor—China at 1 percent and Brazil at 3 percent have far lower poverty rate than the 21 percent for India, and 27 percent for Ethiopia. Brazil and China are thus strong performers in terms of having a smaller percentage of their population that are poor. Yet, these four countries vary widely in population size, which leads to the interesting result that 1 percent of the population in India (12 million) is even larger than 10 percent of the population in Ethiopia (10 million) (Panel A, Row 5). Consequently, although Ethiopia's poverty rate is one-fourth higher than India's, its (absolute) number of poor is just around one-tenth that of the latter (Panel A, Row 2). Similarly, Ethiopia has more than 20 times the poverty rate of China, but has only three times as many poor.

As such, while driving down the poverty rate by one additional percentage point may represent similar progress on this metric for each country, when translated into the number of people living

---

[11] Indeed, Easterly (2009) famously points out that, depending on whether we target a relative change or an absolute change in poverty rates (or primary school enrolment), Africa's performance with the MDGs can be viewed as a success or a failure. He also highlights the importance of distinguishing between change targets versus level targets, as well as positive indicators versus negative indicators. Fukuda-Parr, Greenstein, and Stewart (2013) similarly argue that the criterion of success with the MDGs should focus on the pace of progress rather than on achieving the targets. Applying an alternative measurement method based on this criterion, they find that African countries outperformed global averages in their progress toward achieving the MDG targets. We offer more discussion on the technical details in Appendix 1 (Part A).

in poverty that same percentage point can offer a different perspective. The choice of measure would clearly lead to different interpretations with the pace of poverty reduction in a (regional or global) comparative setting. It can thus be useful to report statistics on the number of poor people as an additional indicator for SDG number 1 on poverty reduction. For example, the absolute number of the people living in poverty—at least globally or regionally—can be reported alongside the headcount poverty rate.

Time Dimension

When measuring poverty reduction, another issue that merits serious consideration is the time dimension. Indeed, what is a reasonable interval length to study the trends: is it a five-year or a ten-year period?[12] Which is more important, the trend or the current level of achievement?[13]

The numbers discussed above with Table 3 are static and provide only one snapshot of poverty at a single point in time. If we consider historical trends, more interesting results emerge. We show the average poverty reduction rate for each country in two periods, the past decade (Table 3, Panel A, Row 3) and the past two decades (Panel A, Row 4). Given the available data, the past decade is defined as the period 2005-15 for Brazil and China, 2004-15 for Ethiopia, and 2004-11 for India; and the past two decades is defined as the period 1995-2015 for Brazil and Ethiopia, 1996-2015 for China, and 1993-2011 for India. For the past decade, China is the best performer, having reduced poverty by an average of 28 percent per year, followed in order by Brazil, India, and

---

[12] An indirect, but related issue is what is the benchmark year from which we start evaluating a country's progress? Clearly, selecting a different start year can result in different conclusions with progress over time. We would like to thank an anonymous reviewer for suggesting this point.

[13] A related but deeper question is that, how do we compare the performance of a (very) poor country with a solid record of decreasing poverty in the past decade versus another not-so-poor country with a slow reduction, or even an uptick, in poverty? If more weight is placed on the poverty trend, the former would be regarded as the better performer, but if more weight is placed on the current poverty level, we would have the opposite result.

Ethiopia. But this order changes when we consider a larger time horizon over the past two decades: China is still the best performer, but now the order following is Ethiopia, Brazil, and India.[14] Thus, using a different study period can result in a different ranking.

Table 3, Panel B summarizes the results discussed above by showing the different rankings for a country, depending on the specific measure of poverty reduction. Perhaps what is most interesting is that, no measure yields the same ranking. Countries can see their position dramatically switching from the top performer to the next-to-worst performer (Brazil) or from the second-best performer to the worst performer (Ethiopia) just according to the measure of poverty reduction considered.

Combining Different Metrics

Taking into consideration both the time dimension and absolute and relative numbers, we next briefly sketch progress with poverty reduction in the past four decades for all the six regions across the world. Figure 3 graphs trends in poverty reduction, with Panel A plotting the poverty rate and the Panel B the number of poor people. To obtain the maximum number of observations, we plot the mean poverty rate for each region over the three decades starting from the 1980s—1980-89, 1990-99, 2000-09—and the past five years where data are available, 2010-2016. Still, no data are available in the 1980s for two regions, Europe and Central Asia and Sub-Saharan Africa.

It is useful to note some observations with this figure. First, in the most recent period, 2010-16, the order of countries by poverty rate mostly, although not perfectly, coincides with that by the number of poor. For the poverty rate, Sub-Saharan Africa (SSF) is the poorest region, followed by

---

[14] In addition, estimates using data dating back to the earliest years—the early 1980s for some countries—that we have available data show another different set of results (not shown). This further supports our discussion.

South Asia (SAS), East Asia and Pacific (EAS), Latin America (LCN), Europe and Central Asia (ECS), and Middle East and North Africa (MEA). For the number of poor, the order is similar except that the last two regions, ECS and MEA, now switch places. Second, all the regions have steadily reduced the poverty rate over time (Figure 3, Panel A), although the speed of reduction is strongest for East Asia and Pacific and South Asia. This result is qualitatively similar when we consider the decrease of the number of poor people over time (Figure 3, Panel B), except that Sub-Saharan Africa has in fact seen an increase in the number of poor people. Furthermore, South Asia has reduced the number of poor more slowly than it has reduced the poverty rate.

We end this subsection on a cautious note that using a different evaluation method can result in a different conclusion for a country's progress on the SDGs; and analysis using time series data, or varying lengths of these time series, may show a quite different result from a static snapshot approach.

III.3. Data Challenges

Our illustrative examples have so far assumed that the data we analyze are complete, comparable, and of good quality. Put differently, underlying our analysis are two implicit assumptions that the data are consistent and are not missing for each and every year. As it happens, these crucial assumption plays a very important role in helping us correctly interpret the data and their trends. These assumptions cannot be taken for granted. We discuss in the next example the interpretation issues that come from data inconsistencies and unavailability, before we review the current status of the UN SDG database.

Different Data Sources

Notably, the data used to evaluate performance with the SDGs can come from different sources. This can pose a cause for concern, since it is common knowledge that even for the same variable, different sources can provide different numbers. As an example, trusted data providers, such as the United Nations, World Bank, and IMF, may publish different country estimates of macroeconomic aggregates such as GDP. As these statistics are founded in the same conceptual framework for all UN member states,[15] why does this happen? There are two reasons. First, many countries with lower statistical capacity publish GDP estimates with a time lag. As international organizations use GDP estimates for operational purposes (e.g., means testing the World Bank's member countries for IDA eligibility, or informing development policy discussions with member state policy makers), they employ different approaches to fill the data gap between the latest published number and the present. Second, international organizations also forecast economic time series, and the schedules for revising (recalibrating) model estimates can vary among the different organizations.

Similarly, two household surveys implemented by a national statistical office to collect data on the same employment characteristics may not produce the same statistics (see, e.g., Dang, Lanjouw, and Serajuddin (2017) for a discussion for Jordan). Furthermore, this issue is not only pertinent to poorer countries but is frequently observed for richer countries as well.[16]

We discuss next a specific example of how analyzing data coming from different sources can complicate the conclusions. In particular, shared prosperity as measured by growth in the income

---

[15] This framework is available at https://unstats.un.org/unsd/nationalaccount/docs/sna2008.pdf. See also the IMF (2018) for an interesting discussion on some potential differences between their macroeconomic databases.
[16] A salient case is the U.S., where the inconsistency between different surveys is well documented in the literature. For example, Abraham *et al.* (2013) examine the differences between employment data between the Current Population Surveys (CPS) and employer-reported administrative data. Bavier (2014) finds spending and income poverty in the Consumer Expenditure Survey (CES) to be an outlier compared with those in other surveys including the Panel Study of Income Dynamics (PSID).

per capita of the bottom 40 percent of the population (Indicator 10.1.1) will most likely—if not always—be measured using data from a household consumption or income survey. However, the GDP growth rate (Indicator 8.1.1) is measured using the national account. Deaton (2005) observes that the estimated growth rate of consumption based on the national account tends to be larger than that based on the household survey, both across countries and over time for major countries.[17] We plot in Figure 4 the growth rates of mean consumption levels from the household survey and GDP per capita over past decades for two major middle-income countries, India and China. Indeed, our estimates using more recent data are consistent with Deaton's findings: the growth rate of the survey mean (blue bars) is almost always less than the growth rate of GDP per capita (red bars). For both countries, over the whole study period, the average growth rate of the survey mean (dashed blue line) is also less than that of GDP per capita (dotted red line).[18] This result also holds on average for more than 150 countries for 1981-2014 in our sample that combines the PovCalNet database and the WDI database.

It can be argued that Deaton's finding, supplemented by our more recent estimates, applies to the whole distribution of consumption, rather than just a chunk of it, such as the bottom 40 percent. Still, this may be a cause for concern if the discussed inconsistencies extend to all parts of the consumption distribution. Consequently, at least for the purpose of comparison with estimates on shared prosperity (Indicator 10.1.1), estimates of the growth rate of consumption using the

---

[17] Deaton (2005) argues that when rich households are less likely to cooperate with the survey than poor people, survey-based estimates of consumption will understate mean consumption. On the other hand, various differences with national accounts estimates can overstate the rate of growth of average consumption, both over time in poor countries, and in comparisons between poor and rich countries at a moment in time. This finding also concurs with the results from an earlier study by Ravallion (2003).

[18] We provide a similar figure for some other countries, including Bangladesh, Indonesia, Nigeria, and Turkey (Appendix 2, Figure 2.1). All these countries, except for Indonesia, saw their GDP per capita grow faster than the survey mean in the period under consideration. Moreover, for Nigeria, the growth rate of the survey mean is even negative while that of GDP per capita is positive.

household surveys should perhaps be provided as a robustness check for estimates on the GDP growth rate using the national account (Indicator 8.1.1).

Missing Data

We have discussed in the previous section the important role of time trends with interpreting poverty reduction results, which similarly holds for other development outcomes. Building on this result, we show in Figure 5 the different patterns of GDP per capita growth over a five-year period during 2011-15 for all countries in the WDI database. We divide growth patterns during this 5-year interval into two groups: decreasing (Panel A) or increasing (Panel B). For most readers, the notion of increasing GDP growth may be continuous growth over all five years (Figure 5, Panel B, group 1); in fact, the real growth patterns are much more complicated. There may, for instance, a decrease in GDP per capita in the first year, and a continuous increase for the remaining years (Panel B, group 2). Or growth may be continuous for the first two or three years, then slip back in the remaining years (Panel B, group 4 and group 5). Or growth may be continuous for the first three years, decrease in the fourth year, and then increase again in the fifth year (Panel B, group 3).

These results help highlight the impacts that missing data can have on interpreting the SDG progress. More precisely speaking, missing data can result in an incorrect interpretation of the actual trends. For example, if the data for country groups 4 and 5 are available not for the whole five years but only for the last two, the growth pattern for these two years would be decreasing despite the general increase over the 5-year period. The pattern can be even more complex if not all countries have data in the same years, and some groups have data only sporadically—which is

true for many SDG indicators. In such cases, missing data can lead to severe misinterpretations of progress over time.

We turn next to examining the data set perhaps most relevant for monitoring progress on the SDGs, the UN's official SDG database.

Overview of the UN SDG Database

The United Nations (2018b) curates a rich database that tracks the SDG indicators for all countries with data coming from various sources and dating back to as early as 1983. Our preliminary assessment indicates that this database consists of data coming from around 200 data sources, where each source is defined as one that contributes 40 or more country-year-indicator observations. Clearly, identifying these different sources and harmonizing any potential consistencies would be an important task that demands time and resources, but that can help improve the quality of this database.

Such a task is beyond the scope of this paper. Thus for now, we assume that all the data are comparable, and we focus on the easier task of identifying missing data. For this purpose, to obtain an overview of this database, we consider only the most recent 5-year period, 2012-16, where the data are most complete.[19] We simply check on the data by summarizing, for each SDG in Table 4, the number of countries (Table 4, column 1) and the number of indicators (Table 4, column 1) covered in the database. We also show the total number of country-indicator-year data points available (column 3), and the percentage of non-missing data points (column 4) for each SDG. For comparison, we show the full numbers of countries, indicators, and all country-indicator-year data

---

[19] Data are not incomplete for 2017 and 2018 in this database, we return to more discussion in the next example.

points in Appendix 2, Table 2.3, assuming that there were no missing data.[20] The latter statistics is simply the ratio of the figures in Table 4, column 3 and the total numbers of country-indicator-year data points that may be available (Appendix 2, Table 2.3, column 3).

Table 4 points to the data challenge with tracking the SDGs, which varies widely by each goal. First, we never have full data for all the 249 countries formally recognized by the UN, with the number of countries with missing data ranging from four (SDG 15, Life on land) to as many as 96 (SDG 13, Climate action). Furthermore, for two countries during the period considered, we do not even have any single data point.[21] Second, we also never have full data for all the indicators in each goal. Again, the number of indicators varies widely from goal to goal, and ranges from one (SDG 4, Quality education or SDG 9, Industry, Innovation, and Infrastructure) to 17 (SDG 17, Partnership for global development). In fact, the last row of Table 4, column 2 indicates that the total number of indicators that we have some data accounts for just over half of all the SDG indicators (i.e., 134 out of 244 indicators). While this empirical finding is consistent with our earlier discussion in Section II (i.e., we can only have some data coverage for at most 80 percent of all the indicators in theory), it also points to a seemingly worse picture on data availability in practice.

But note that the goal-by-goal "tests" for missing data offered by columns 1 and 2 in Table 4 may not uncover the full severity of the missing data challenge, since the availability of just one country data point for one indicator satisfies these tests. What is most relevant to analysis is the absolute number of country-indicator-year data points available (column 3), and its relative

---

[20] Since the UN's database does not remove the duplicate indicators, we also keep all the 244 indicators with duplicates in Appendix 2, Table 2.4 for better comparison.

[21] This may perhaps partly explained by the fact that the two countries with completely missing data are the island nations Bouvet Island and Sark, which have a small population. Indeed, Bouvet Island is considered to be an uninhabited island (CIA, 2018), while a recent estimate puts the population of Sark at only 600 people (BBC, 2012).

number equivalence to the percentage of non-missing data (column 4). Column 4 indicates that the percentage of non-missing data is less than 10 percent for six goals, 10-30 percent for seven goals, and 30-40 percent for the remaining four goals. On average, the percentage of non-missing data is just 19 percent (Table 4, column 4, last row).

Do these results change if we consider all the data in the 20-year period, from 2000 to 2018? A longer period may increase the number of countries and indicators covered, which are easier tests as discussed earlier, but may decrease the percentage of non-missing data if we have fewer data points in earlier years. Table 2.4 in Appendix 2 provides supportive evidence for these hypotheses: over the longer period, the number of countries and indicators are both larger, but the percentage of missing data is lower at 16 percent.

IV. Further Reflections and Tentative Suggestions

Before further reflecting in this section on some aspects of the issues and challenges discussed earlier, it can be useful to briefly review the political background against which the SDGs came into being. This can offer us more insights about changes that can potentially be made to them. We then discuss the areas where data and measurement challenges are most severe and may be addressed with some relatively cost-effective solutions. Where relevant, we also attempt to offer suggestions to address technical issues based on the latest estimation methods in the literature. But we keep the technical discussion to the minimum and offer more specific details in Appendix 1. To facilitate comparison with the previous section, we use headings similar to those in the previous sections.

Political Settings for the SDGs

22

Different from the MDGs which were considered to be a donor-led agenda, the SDGs were the outcome of a far more participatory process where lower-income countries (e.g., Brazil, Colombia) and smaller Western European countries could take an active role. Furthermore, the constituency in this process includes not only the development community, but also the environment community, which traditionally have different political alignments from that of the former. As such, the SDGs offer more diverse ideas on development outcomes, as well as a significantly richer agenda than that of the MDGs (Dodds, Donoghue, and Roesch, 2017; Fukuda-Parr and McNeill, 2019).

Yet, there are concerns about a mismatch between the (goals and) targets and the indicators with the SDGs. On one hand, MacFeely (2018) raises the concern that although the SDGs were agreed to by all UN states, it is the global statistical community that effectively select the indicators, thus ultimately determine whether the targets will be a success or failure. On the other hand, technical issues (such as data availability) aside, there were concerns that the selection of an indicator can be highly political.[22] A recent special issue of the journal *Global Policy* offers an interesting discussion of the dynamics behind the politics of indicator selection—particularly the process of moving from goals to targets to indicators—for various outcomes including education, environment, justice, and sustainable agriculture. For example, Fukuda-Parr (2019) provides a detailed account of the struggle to include inequality as a stand-alone goal, and observes that while various stakeholders, ranging from academics, civil society groups, UN agencies, to governments in the South, support this goal, most of the governments of the North consider it to be redundant. The end outcome is that inequality was primarily treated under the SDG framework from the

---

[22] Notably, the drafting of the goals and targets themselves was considered as part of the political negotiations as well (UNSC, 2015).

poverty and exclusion angle, rather than from one that highlights potential issues with concentration of wealth and income in elite population groups.

This helps illustrate that selecting the final targets is a complex and political process. Refining the indicators may be similarly complicated. As a result, our recommendations below should be best interpreted after taking into account this background.

Identification

Our earlier discussion suggests that 232 indicators appear onerous, with undesirable consequence of overlap and likely conflicting goals. In fact, achieving some goals requires setting different priorities, which may conflict with each other. For instance, countries—rich or poor—oftentimes have to make the difficult decision of how much resources to invest in newer technology that is environment-friendly but also more expensive. This decision-making process is further complicated by the fact that technology is fast advancing, and climate changes concerns may affect certain countries (e.g., island nations or those with a long coastline) more than others. This again poses challenges to both data collection and interpretation.

Furthermore, not every country may be able to achieve excellent performance on *all* these indicators in practice. Countries at different levels of development clearly have different priorities, and probably adopt different strategies to accomplish their goals. For example, for a poor country, the more pressing concern is to reduce absolute poverty and ensure economic growth. But for a (much) richer country, the priorities might shift toward other issues, such as job creation or

reducing inequality. Countries may also differ in their statistical capacity, which is indispensable for monitoring the SDGs well.[23]

Seen in this light, we may want to focus on the ultimate goal of development, rather than certain indicators. For this purpose, we offer in Table 5 a simple grouping of the SDG indicators by thematic areas (the full list of indicators is provided in Appendix 2, Table 2.5) We propose four areas: Economics, Health and Human Development, Governance, and Environment. These are related to the UN's five themes of People, Planet, Prosperity, Peace, and Partnership, with our Governance theme incorporating the UN's themes of Partnership and Peace. These areas offer a general way of summarizing the 17 goals (and 232 indicators), which can be particularly useful when data are unavailable to plot indicator trends over time. Progress on the SDGs can then be measured as a two-step process, with one step setting out general trends in the four areas, and the next much more granularities on the trends for each of the 232 indicators.

In fact, our proposed grouping for better identification is consistent with other ongoing efforts to find better ways to interpret progress on the SDGs. For example, Kanbur et al. (2018) suggest that if an African country were forced to prioritize to five indicators only, they would recommend the following: 1) per capita income, 2) income inequality and poverty, 3) employment, 4) a multidimensional deprivation index (based on access to basic public services), and 5) long-term environmental degradation. For another example, Sach et al. (2018) recently propose an SDG index that tracks a country's overall progress, based on the principle of distance to the frontier. This index standardizes a country's progress from the minimal score (i.e., the numerator in

---

[23] This challenge is also relevant to international organizations. For example, of the 232 indicators, the World Bank is directly responsible for monitoring the progress of 20 indicators, and involved in the production of an extended group of 23 additional indicators.

Equation (1) below) by the range between the minimal and maximal scores (i.e., the denominator in Equation (1) below)

$$R_i^s = \frac{c_i - c_i^{min}}{c_i^{max} - c_i^{min}} \qquad (1)$$

where c_i represent the current achievement on an indicator I, for $i = 1, \ldots, N$. c_i^{max} and c_i^{min} respectively denote whether the achievement is at the maximal level or the minimal level.

Given this new tool to measure overall progress on the SDGs, it appears not unreasonable to consider a three-step process to measure progress on the SDGs, with subsequent steps offering more granularities than earlier steps. That is, Step 1 employs a relevant index such as that proposed by Sachs et al. (2018) to plot out the general trend in development for countries. Step 2 disaggregate this overall trend into trends in the four thematic areas, or a set of core indicators. Step 3 tracks the trends for each of the 17 goals, and finally all the 232 indicators where data are available.[24]

Interpretation

As discussed earlier, different methods for evaluating performance on the SDGs can produce different results. In particular, analysis using time series may show a quite different result from snapshots of the data. A common analytical framework would thus be a prerequisite for obtaining comparable evaluation results.

For further illustration on the importance of evaluation methods, we summarize in Table 6 methods that have been recently proposed. We also briefly review two related indexes, the World

---

[24] Again, we also assume that all the indicators (and goals) should be made consistent and non-overlapping for their grouping to be meaningful. See, for example, Barbier and Burgess (2018) for a recent analysis that quantifies the trade-offs (and complementarities) between the SDGs. We return to discuss other relevant indexes in the next section.

Bank's human capital index (HCI) and country statistical capacity index (SPI). (We provide more technical details on these methods in Appendix 1, Part A.) There are, unsurprisingly, pros and cons with each method. The simplest is the dashboard method employed by the United Nations (2018c), which analyzes the SDG indicators as is; that is, it tracks progress on each and every indicator. This method offers straightforward interpretation, but its disadvantage is that, with so many indicators, it is difficult to evaluate progress for them as a whole. For example, it is hard to compare a country with good poverty reduction but weak environment protection with another which has the opposite performance.

Sachs et al. (2017) combine the dashboard approach with an index that is an average (i.e., the arithmetic mean) of all the standardized indexes obtained from Equation (1). While this procedure is somewhat more complicated than the dashboard approach and requires making certain assumptions, it offers a way to evaluate overall progress for all the indicators. But note that this index is driven by data availability.[25] In fact, Sachs et al. (2017) also offer a dashboard-based approach to keep track of progress on each of the 17 goals.

An alternative to Sachs et al.'s (2018) method of constructing the SDG index by the arithmetic mean is to construct it using either the geometric mean or the product of all components. (The former is the latter raised to the power of $1/N$, where N is the number of components in the product.) Put differently, we can multiply them together in a similar way to the World Bank's recent HCI (Kraay, 2018). Alternatively, after obtaining this product, we can then take the geometric mean, as does the UNDP's human development index (HDI) (UNDP, 2010). The advantage of the latter is the emphasis on the lower values, that is, countries with weaker performance on one indicator

---

[25] We offer further discussion in Appendix 1, Part A.

will have a weaker overall index than if the arithmetic mean is used. However, the disadvantage of the geometric mean is that, if any indicator is 0, the overall index will be 0 by construction. Another disadvantage is that, a product of all 232 indicators is perhaps far more unwieldy and harder to interpret than a product of just three indicators.

Finally, a recent approach proposed by Cameron et al. (2019) in the context of measuring a country's statistical capacity can be relevant. This method builds on the widely used counting approach by Atkinson (2003).[26] It also uses the arithmetic mean to aggregate indicators, but in a more complex way. The procedures to construct the statistical capacity index (SPI) consists of the following three steps. First, different levels of dimension should be identified. For example, we can choose two levels of dimension: the first can naturally be the 17 goals (or alternatively, the four thematic areas proposed earlier in Table 5), and the second is the indicators themselves. Second, the goals (or thematic areas) on the first level will have the same weight that is equal to the inverse of the number of goals; for example, the weight is 1/17 if we choose the 17 goals. Similarly, the indicators within the same goal (or thematic areas) on the second level will have the same weight that equals the inverse of the number of indicators for this goal. Finally, the overall index will be constructed as a weighted average of all indicators from the two levels.

Compared to the other methods, the SPI has some flexibility with its functional form (i.e., more indicators can be added, but they do not change the weight of the main dimension), and it allows for decomposition by subgroups, such as geographical regions. On the other hand, this approach requires that the main dimensions be clearly defined, with a good justification for which indicators going into which dimension. This latter concern perhaps poses no challenge, since we already have

---

[26] A well-known example is Alkire and Foster's (2011) multi-dimensional poverty index.

the 17 goals as a natural grouping for all the indicators. Alternatively, we can use the results from our Table 5 for another grouping option. (See Appendix 1, Part A for technical details.)

But one substantive issue with this counting approach is that, we give equal weight to all the goals—as well as all the indicators within each goal—when aggregating them. This implies that, progress on one indicator is linearly related with that on any others, so that a country can keep their overall performance the same by trading the progress on one indicator with that on another. Another substantive interpretation issue is that, the same amount of progress at different levels of development may have quite different meanings. In particular, an accepted hypothesis is that once a country has reached its technological production frontiers, its economic growth will slow down, unlike a country that has still to fully develop its potential (see, e.g., Cowen (2011) and Gordon (2016)). For example, economic growth can be harder to achieve for richer countries that are already operating at full capacity than for an emerging economy. As another example, it may be quite hard, if not downright impossible, to reduce infant mortality rate to 0 percent. As such, a country that has already achieved a very low infant mortality rate would find it much harder to further reduce it by the same amount as another where infant mortality is still high.

Should we take into account this difference in evaluating country progress? Or do we need to apply different standards for measurement, depending on a country's level of development? Answers to these questions are not straightforward, and call for more thoughts.[27]

---

[27] Mathematically speaking, however, it can be rather straightforward to operationalize the idea of giving different weights to performance according to level of development. For example, one way to achieve this with the Sachs et al. (2018) index is to simply raise it to the power of α as follows ($\frac{c_i - c_i^{min}}{c_i^{max} - c_i^{min}})^{\alpha}$, where α assumes different values depending on the country's level of development. For a recent alternative approach that applies a multidimensional synthesis of indicators, see Casini et al. (2019).

Data Challenges

The last, but far from least, thorny issue with monitoring the SDGs is missing data. Missing data can simply indicate that data are not collected for some indicators. For example, the infant mortality rate is so low for some Nordic countries that these countries do not typically collect data on it. But oftentimes missing data points to either low performance, particularly for low-income countries, or inadequate statistical capacity.

There are two ways to address this data challenge: collect better and more frequent data, or employ recent developed statistical methods that can help impute the missing data. The first approach is also the more popular long-term data collection approach, and it should be implemented under the ideal circumstances. Yet, in practice, collecting data on all the 232 indicators requires both coordination of different government agencies and careful budget planning. This is not to mention the technical capacity required to ensure that the collected data have good quality. These practical issues have resulted in many indicators that are either missing or are available at infrequent intervals.

As an example, estimating the poverty rate—which is the first indicator under the first goal (Indicator 1.1.1)—may not be that simple to achieve, particularly among poorer countries. A recent study by Serajuddin et al. (2015) find that, over the period 2002-11, more than one-third (57) of the 155 countries for which the World Bank monitors poverty data have only one poverty data point or no data at all in the WDI database. Even where countries collect data on poverty, these may not be comparable over time due to bad quality. Indeed, Beegle et al. (2016) point out that just over half (27) of the 48 countries in Sub-Saharan Africa had two or more comparable household surveys for the period 1990-2012. These examples further illustrate the missing data challenge we discussed earlier with Table 4.

The second approach of employing statistical modelling techniques to impute missing data can offer a promising alternative when data are scarce. Within the imputation approach, there are also two directions: one is imputation at the micro level, using household (or individual) data from household surveys, and the other is imputation at the global level using country data. We refer interested readers to Dang, Jolliffe, and Carletto (2019) for a more detailed review of the literature on micro imputation methods. As for macro imputation methods, Bonjour et al. (2013) applied mixed modeling techniques to impute estimates for solid fuel use for household cooking for 155 countries over the period 1980-2010 and obtained encouraging results. Their approach has also been employed to impute estimates for missing electrification rates in a recent joint global study by the UN Statistics Division, the World Health Organization, and the World Bank (International Energy Agency et al., 2018).

For a brief illustration, we adopt a modified version of this statistical model with some further refinements, and provide estimates of the electrification rate for several African countries in Figure 7.[28] The selected countries represent different levels of electrification rate and include Uganda (low electrification), Nigeria (medium electrification), and Tunisia (high electrification). While the imputed rates (solid line) do not perfectly coincide with the actual data points, they closely track the latter in all three cases.[29] Estimates for all countries in the African region as a whole are also encouragingly close to those based on the actual data.

Estimations results using both micro and macro imputation methods thus appear promising and could be applied to provide estimates for other indicators where data are unavailable. There

---

[28] We offer more discussion on the technical details in Appendix 1.

[29] The actual data points themselves are the best available estimates of electrification that come from different data sources including household surveys and population censuses, so some inconsistency may be expected; see International Energy Agency et al. (2018) for further details on the data.

are advantages and disadvantages with both data collection approaches. High-quality surveys are clearly the long-term solution to the data missing challenge, but they are costly, in terms of both finances and time. Meanwhile, imputation-based estimates can fill in data gaps when actual survey data are not available (at least in the short term or where there is a need to obtain estimates going back in time), but they require certain levels of technical capacity.

Role of International Organizations

Finally, we offer some further thoughts about the role of international organizations in tracking the SDGs. International organizations are the key actors that are actively engaged in all aspects of the process, from data collection, coordination, and standardization to analysis. Indeed, international organizations are assigned as the custodian agencies for reporting on the SDG targets in the areas of their specialty. In this capacity, they work with national statistical systems to develop methodologies for indicators to measure progress on the SDGs. The agencies also work with countries to compile data for SDG indicators, which they submit to the UN SDG database (which is reviewed in the previous section).

For example, FAO is responsible for indicators related to food and agriculture, and UNICEF for those regarding child welfare. Figure 6 illustrates the complex and multi-step process for tracking indicators related to SDG number 6 on water, based on a UN website dedicated to that goal. First, the custodian agency requests data from a country, or retrieves such data from public official sources. Second, the country sends the custodian agency the requested data. Third, the custodian agency validates the data in consultation with the country. Fourth, the country signs off

on the validation. Fifth, the custodian agency sends the validated data to the UN Statistics Department, which finally published the data. This process is fairly standard.[30]

Given the magnitude of their role, international organizations not only curate and provide quality assurance for SDG-related data, but they are also uniquely positioned to make a significant contribution to improving their quality, as well as generating new research on the SDGs. As an example, the international financial market has become more developed and is now accessible to most, if not all, poorer countries. This may result in a more diminished role for the lending operations of international financial organizations like the World Bank. As such, there have been stronger calls for the World Bank to move more vigorously to becoming a knowledge bank that generates new data and trend-setting research (see, e.g., Clemens and Kremer (2016) and Ravallion (2016)). This would, in fact, build on its recognized strengths in data and analytics.[31]

In this regard, international organizations may be expected to offer their vision for the development landscape in the decades to come. They may also be expected to spearhead new data initiatives to meet evolving global data needs beyond the SDGs, such as statistics on displaced population groups like refugees, who are usually not captured well in traditional surveys and censuses. Another example would be the increasingly common use of subjective well-being data, such as life satisfaction, to supplement the traditional money-metric data used to measure welfare

---

[30] For an example, the World Bank often estimates a country's poverty rate based on a joint consultation process between Bank staff and government officials (often from the national statistical office) after intensive analysis of household surveys.

[31] Indeed, Birdsall (2015) proposes that the World Bank can invest more resources in supporting researchers, particularly in poorer countries. Clemens and Kremer (2016) even suggest that the World Bank has had more influence on policies in poorer countries through its policy advocacy than its lending portfolios; see also Gavin and Rodrik (1995) for a similar viewpoint. Ravallion (2016) further argues that in its activities related to data and research, the World Bank has not fully reached its potential as a "knowledge bank". It is not clear to what extent that similar arguments can apply to other international organizations like the UNDP that are also well known for their technical assistance.

outcomes.[32] At the same time, data provided by countries remain an integral part of monitoring the SDGs, thus further statistical and analytical capacity activities, particularly for low-income countries, should perhaps receive more attention.

V. Conclusion

We offer in this paper a review of various challenges regarding identification and measurement methods related to the SDGs. We place an emphasis on the data angle, and we focus on poorer countries. Our findings point to the need to further refine the SDG indicators in terms of their wordings—while we acknowledge that it can be a difficult process to make (even) minor changes to indicators—as well as to clarify their underlying objectives. We also bring attention to potential pitfalls with interpretation of progress on the SDGs, where different evaluation methods can lead to different conclusions. One particularly demanding challenge is the severe shortage of data for tracking progress across countries and over time.

We also propose relatively simple solutions to identify and interpret progress. We propose a three-step process to measure progress on the SDGs, with each subsequent step offering more granularity than the previous one. In particular, this process can well consist of tracking an overall index, some major groups, and then all the SDG indicators. We also consider imputation-based statistical methods to be cost-effective alternatives to addressing the missing data challenge. Furthermore, we view international organizations as playing a most relevant role in producing and curating data to track progress on the SDGs, which should be implemented in close collaboration

---

[32] For example, the OECD produces an annual life index that aims to go beyond GDP figures (OECD, 2017). See also the recent annual world happiness report by Helliwell et al. (2018).

with countries. International organizations may, and should, also take new data initiatives as global

data needs evolve beyond the SDGs.

References

Abraham, K. G., Haltiwanger, J., Sandusky, K., and Spletzer, J. (2013). "Exploring Differences in Employment between Household and Establishment Data". *Journal of Labor Economics*, 31, S129-S172.

Alkire, Sabina, and James Foster. (2011). "Counting and multidimensional poverty measurement." *Journal of Public Economics,* 95(7): 476-487.

Atkinson, Anthony B. (2003). "Multidimensional deprivation: contrasting social welfare and counting approaches." *Journal of Economic Inequality,* 1(1): 51-65.

Barbier, Edward B. and Joanne C. Burgess. (2018). "Sustainable Development Goal Indicators: Analyzing Trade-offs and Complementarities". *Paper presented at the Sustainable Development Conference*. University of Michigan, Ann Arbor.

Basu, K. (2013). Shared prosperity and the mitigation of poverty: In practice and in precept. *Policy Research Working Paper* # 6700. Washington, DC: The World Bank.

Bavier, R. (2014). "Recent Trends in U.S. Income and Expenditure Poverty". *Journal of Policy Analysis and Management*, 33, 700–718.

Beegle, Kathleen, Luc Christiaensen, Andrew Dabalen, and Isis Gaddis. (2016). *Poverty in a Rising Africa*. Washington, DC: The World Bank.

Birdsall, Nancy. (2015). "A New Mission for the World Bank". Accessed on October 24, 2018 at https://www.project-syndicate.org/commentary/world-bank-global-public-goods-cgiar-by-nancy-birdsall-2015-06?barrier=accesspaylog/

Bonjour, Sophie, Heather Adair-Rohani, Jennyfer Wolf, Nigel G. Bruce, Sumi Mehta, Annette Prüss-Ustün, Maureen Lahiff, Eva A. Rehfuess, Vinod Mishra, and Kirk R. Smith. (2013). "Solid fuel use for household cooking: country and regional estimates for 1980–2010." *Environmental Health Perspectives,* 121(7): 784-790.

Bourguignon, François, Agnès Bénassy-Quéré, Stefan Dercon, Antonio Estache, Jan Willem Gunning, Ravi Kanbur, Stephan Klasen, Simon Maxwell, Jean-Philippe Platteau, and Amedeo Spadaro. (2010), "The Millennium Development Goals: An Assessment." In Kanbur, R. and A.M. Spence (eds). *Equity in a Globalizing World*. World Bank for the Commission on Growth and Development, pp. 17-39.

British Broadcasting Corporation. (BBC). (2012). "Sark Election 2012: Two conseillers lose seats". Accessed on October 24, 2018 at https://www.bbc.com/news/world-europe-guernsey-20697765

Cameron, Grant, Hai-Anh Dang, Mustafa Dinc, James Foster, and Michael Lokshin. (2019). "Measuring the Statistical Capacity of Nations". *World Bank Policy Research Paper 8693*. World Bank: Washington, DC.

Casini, Margherita, Simone Bastianoni, Francesca Gagliardi, Massimo Gigliotti, Angelo Riccaboni, and Gianni Betti. (2019). "Sustainable Development Goals indicators: A methodological proposal for a Multidimensional Fuzzy Index in the Mediterranean area." *Sustainability*, 11(4): 1198.

Central Intelligence Agency. (CIA). (2018). *The World Factbook.* Accessed on October 24, 2018 at https://www.cia.gov/library/publications/the-world-factbook/geos/print_bv.html

Clemens, Michael A. and Michael Kremer. (2016). "The New Role for the World Bank." *Journal of Economic Perspectives*, 30(1): 53-76.

Clemens, Michael A., Charles J. Kenny, and Todd J. Moss. (2007). "The trouble with the MDGs: confronting expectations of aid and development success." *World Development,* 35(5): 735-751.

Cowen, Tyler. (2011). *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better.* New York: Dutton.

Dang, Hai-Anh and Andrew Dabalen. (2019). "Is Poverty in Africa Mostly Chronic or Transient? Evidence from Synthetic Panel Data." *Journal of Development Studies*, 55(7): 1527-1547.

Dang, Hai-Anh and Peter Lanjouw. (2016). "Toward a new definition of shared prosperity: A dynamic perspective from three countries". In K. Basu & J. Stiglitz. (Eds.), *Inequality and Growth: Patterns and Policy* (pp. 151-171). New York: Palgrave MacMillan Press.

Dang, Hai-Anh, Peter Lanjouw, Umar Serajuddin. (2017). "Updating Poverty Estimates at Frequent Intervals in the Absence of Consumption Data: Methods and Illustration with Reference to a Middle-Income Country." *Oxford Economic Papers*, 69(4): 939-962.

Dang, Hai-Anh, Dean Jolliffe, and Calogero Carletto. (2019). "Data Gaps, Data Incomparability, and Data Imputation: A Review of Poverty Measurement Methods for Data-Scarce Environments". *Journal of Economic Surveys*, https://doi.org/10.1111/joes.12307.

Deaton, Angus. (2005). "Measuring poverty in a growing world (or measuring growth in a poor world)." *Review of Economics and Statistics,* 87(1): 1-19.

Dodds, Felix, Ambassador David Donoghue, and Jimena Leiva Roesch. (2017). *Negotiating the sustainable development goals: a transformational agenda for an insecure world*. New York: Rouledge.

Easterly, William. (2009). "How the millennium development goals are unfair to Africa." *World Development* 37(1): 26-35.

---. (2015). "The trouble with the sustainable development goals." *Current History,* 114(775): 322.

Economist. (2015). "*The 169 Commandments".* Accessed on September 30, 2018 at https://www.economist.com/leaders/2015/03/26/the-169-commandments.

Fukuda-Parr, Sakiko. (2019). "Keeping out Extreme Inequality out of the Agenda: SDGs and the Politics of Measurement Tools". *Global Policy,* 10: 61-69.

Fukuda-Parr, Sakiko, and Desmond McNeill. (2019). "Knowledge and Politics in Setting and Measuring the SDGs: Introduction to Special Issue." *Global Policy,* 10: 5-15.

Fukuda-Parr, Sakiko, Joshua Greenstein, and David Stewart. (2013). "How should MDG success and failure be judged: Faster progress or achieving the targets?" *World Development*, 41: 19-30.

Gavin, Michael and Dani Rodrik. (1995). "The World Bank in Historical Perspective." *American Economic Review,* 85(2): 329–34.

Gordon, Robert J. (2016). *The Rise and Fall of American Growth: The US Standard of Living since the Civil War*. New Jersey: Princeton University Press.

Helliwell, J., Layard, R., & Sachs, J. (2018). *World Happiness Report 2018*. New York: Sustainable Development Solutions Network.

International Energy Agency, International Renewable Energy Agency, United Nations, World Bank Group, and World Health Organization. (2018). Tracking SDG7: The Energy Progress Report 2018. World Bank, Washington, DC.

International Monetary Fund (IMF). (2018). *What explains differences between WEO and IFS data and/or the latest data available from the source?* Accessed on October 15, 2018 at https://www.imf.org/external/pubs/ft/weo/faq.htm#q1f

Jia, Ruixue, and Hyejin Ku. (forthcoming). "Is China's Pollution the Culprit for the Choking of South Korea? Evidence from the Asian Dust." *Economic Journal*.

Jolliffe, D., Lanjouw, P., Chen, S., Kraay, A., Meyer, C., Negre, M., Prydz, E., Vakis, R. & Wethli, K. (2015). *A measured approach to ending poverty and boosting shared prosperity: concepts, data, and the twin goals*. Policy Research Report. Washington, DC: World Bank.

Joshi, Devin K., Barry B. Hughes, and Timothy D. Sisk. (2015). "Improving governance for the post-2015 sustainable development goals: scenario forecasting the next 50 years." *World Development*, 70: 286-302.

Kanbur, Ravi, Ebrahim Patel, and Joseph Stiglitz. (2018). "Sustainable Development Goals and Measurement of Economic and Social Progress". In Stiglitz, J., J. Fitoussi and M. Durand. (eds.) *For Good Measure: Advancing Research on Well-being Metrics Beyond GDP*, OECD Publishing, Paris,

Klopp, Jacqueline M., and Danielle L. Petretta. (2017). "The urban sustainable development goal: Indicators, complexity and the politics of measuring cities." *Cities*, 63: 92-97.

Kraay, Aart. (2018). "Methodology for a World Bank Human Capital Index". *Policy Research Working Paper No. 8593*. World Bank, Washington, DC.

MacFeely, Steve. (2018). The 2030 Agenda: An Unprecedented Statistical Challenge. *International Policy Analysis*. Friedrich Ebert Stiftung.

OECD. (2017). *How's Life? 2017: Measuring Well-being*. OECD Publishing.

Ordaz, Enrique. (2019). "The SDGs Indicators: A Challenging Task for the International Statistical Community." *Global Policy*, 10: 141-143.

Ravallion, Martin. (2003). "Measuring aggregate welfare in developing countries: How well do national accounts and surveys agree?" *Review of Economics and Statistics*, 85(3), 645-652.

---. (2016). "The World Bank: Why it is still needed and why it still disappoints." *Journal of Economic Perspectives*, 30(1): 77-94.

Sachs, J., Schmidt-Traub, G., Kroll, C., Lafortune, G., Fuller, G. (2017). *SDG Index and Dashboards Report 2017*. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN).

---. (2018). *SDG Index and Dashboards Report 2018*. New York: Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN).

Serajuddin, Umar, Hiroki Uematsu, Christina Wieser, Nobuo Yoshida, and Andrew Dabalen. (2015). "Data deprivation: another deprivation to end." *World Bank Policy Research Paper no. 7252*, World Bank, Washington, DC.

United Nations. (2018a). *Global Indicator Framework for the Sustainable Development Goals and Targets of the 2030 Agenda for Sustainable Development*. New York: United Nations.

---. (2018b). *SDG Indicators Database*. Accessed on October 15, 2018 at https://unstats.un.org/sdgs/indicators/database/

---. (2018c). *The Sustainable Development Goals Report 2018.* New York: United Nations.

---. (2019). *Tier Classification for Global SDG Indicators.* Accessed on April 15, 2019 at https://unstats.un.org/sdgs/files/Tier%20Classification%20of%20SDG%20Indicators_4%20April%202019_web.pdf

UNDP. (2010). *Human Development Report 2010: The Real Wealth of Nations*. New York: Palgrave Macmillan for the UNDP.

United Nations, Department of Economic and Social Affairs, Population Division. (UNDESA). (2014). *World Urbanization Prospects: The 2014 Revision.* New York: United Nations.

United Nations Statistical Commission. (UNSC). (2015). *Technical report by the Bureau of the United Nations Statistical Commission on the process of the development of an indicator framework for the goals and targets of the post-2015 development agenda.* Accessed on April 16, 2019 at https://sustainabledevelopment.un.org/content/documents/6754Technical%20report%20of%20the%20UNSC%20Bureau%20%28final%29.pdf

Vandemoortele, Jan. (2009). "The MDG conundrum: meeting the targets without missing the point." *Development Policy Review,* 27(4): 355-371.

Waage, Jeff, Christopher Yap, Sarah Bell, Caren Levy, Georgina Mace, Tom Pegram, Elaine Unterhalter, Niheer Dasandi, David Hudson, Richard Kock, Susannah Mayhew, Colin Marx, and Nigel Poole. (2015). "Governing the UN Sustainable Development Goals: interactions, infrastructures, and institutions." *Lancet Global Health*, 3(5): e251-e252.

Waage, Jeff, Rukmini Banerji, Oona Campbell, Ephraim Chirwa, Guy Collender, Veerle Dieltiens, Andrew Dorward, Peter Godfrey-Faussett, Piya Hanvoravongchai, Geeta Kingdon, Angela Little, Anne Mills, Kim Mulholland, Alwyn Mwinga, Amy North, Walaiporn Patcharanarumol, Colin Poulton, Viroj Tangcharoensathien, Elaine Unterhalter. (2010). "The Millennium Development Goals: a cross-sectoral analysis and principles for goal setting after 2015." *Lancet*, 376(9745): 991-1023.

World Bank. (2016). *The cost of air pollution: strengthening the economic case for action*. Washington, D.C.: World Bank Group.

---. (2019). *World Development Indicators Online*.

Vanian, Jonathan. (2016). "Facebook Unveils Plans to Bring Internet to Both Cities and Rural Areas". *Fortune*. http://fortune.com/2016/04/13/facebook-terragraph-project-aries-internet/

Zheng, Siqi and Matthew E. Kahn. (2017). "A new era of pollution progress in urban China?" *Journal of Economic Perspectives,* 31(1): 71-92.

Table 1: Indicators that Overlap

No	Goal	Indicator	Example
1	Goal 1. End poverty in all its forms everywhere (Target 1.5)	1.5.2	Direct disaster economic loss in relation to global gross domestic product (GDP)
	Goal 11. Make cities and human settlements inclusive, safe, resilient and sustainable (Target 11.5)	11.5.2	Direct economic loss in relation to global GDP, damage to critical infrastructure and number of disruptions to basic services, attributed to disasters
2	Goal 4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all (Target 4.7)	4.7.1	Extent to which (i) global citizenship education and (ii) education for sustainable development, including gender equality and human rights, are mainstreamed at all levels in (a) national education policies; (b) curricula; (c) teacher education; and (d) student assessment
	Goal 12. Ensure sustainable consumption and production patterns (Target 12.8)	12.8.1	Extent to which (i) global citizenship education and (ii) education for sustainable development (including climate change education) are mainstreamed in (a) national education policies; (b) curricula; (c) teacher education; and (d) student assessment
3	Goal 11. Make cities and human settlements inclusive, safe, resilient and sustainable (Target 11.7)	11.7.2	Proportion of persons victim of physical or sexual harassment, by sex, age, disability status and place of occurrence, in the previous 12 months
	Goal 16. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels (Target 16.1)	16.1.3	Proportion of population subjected to (a) physical violence, (b) psychological violence and (c) sexual violence in the previous 12 months

Source: Global indicator framework adopted by the General Assembly (A/RES/71/313) and annual refinements contained in E/CN.3/2018/2. The full list of indicators is available at https://unstats.un.org/sdgs/indicators/indicators-list/.

Table 2: Change in Shared Prosperity for Sub-Saharan African Countries (percentage)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Growth in the population share of each welfare category			Growth in mean consumption for bottom 40%	Growth in overall mean consumption	Pro-poor growth scenario
No	Country	Poor	Vulnerable	Middle class			
1	Chad	-36.1	42.6	192.1	35.3	53.5	**
2	Botswana	-28.7	6.6	12.5	28.9	-0.5	**
3	Mauritania	-27.7	-18.8	34.7	13.2	12.4	***
4	Ghana	-20.7	4.4	21.7	14.9	20.3	**
5	Uganda	-19.1	26.3	24.9	21.3	20.7	**
6	Congo, Dem. Rep.	-13.5	149.7	249.4	75.0	69.7	**
7	Mozambique	-12.6	70.0	30.9	20.9	21.3	**
8	Rwanda	-8.7	19.3	24.4	27.0	20.5	**
9	Tanzania	-6.7	10.5	3.6	14.6	5.6	**
10	Sierra Leone	-6.5	16.6	-6.7	14.9	0.5	*
11	Ethiopia	-3.9	-1.2	26.7	-4.0	2.5	***
12	Togo	-1.4	-4.5	14.9	-7.8	3.8	***
13	Eswatini	-1.3	-1.3	4.0	-7.4	-3.8	***
14	Malawi	-1.0	-3.8	28.4	-8.1	5.9	***
15	Senegal	0.9	3.3	-7.9	-3.1	-2.5	---
16	Nigeria	5.5	1.4	-6.4	-1.7	-0.5	---
17	Burkina Faso	6.3	-4.4	-18.0	7.6	-5.6	--
18	Zambia	7.8	-12.0	-12.1	3.7	-4.4	--
19	Madagascar	9.5	-32.5	-23.6	-5.6	-16.3	--
20	Côte d'Ivoire	15.1	-5.3	-5.6	-3.4	-6.9	--
21	Cameroon	34.5	-12.3	-8.9	-5.7	-10.5	--
	Regional average	-5.2	12.1	27.6	11.0	8.8	**

Note: Authors' calculation based on household survey data. Household heads' age is between 25 and 55 in the first survey round and adjusted accordingly for the second survey round. The poverty line and vulnerability line are respectively set at $1.9/day and $4.3/day in 2011 PPP dollars for both periods. Pro-poor growth scenarios are based on the classification provided in Appendix 2, Table 2.1. Countries are ranked in a decreasing order of reduction in headcount poverty (column 3). The regional average is a simple average (unweighted). Most household surveys were implemented in the late 2000s. Adopted with modifications from Table 5 in Dang and Dabalen (2019).

Table 3: Poverty Reduction as Measured by Different Methods

Panel A: Estimation Results

No	Outcomes	Country			
		Brazil	China	Ethiopia	India
1	Headcount poverty rate ($1.90/ day, percent)	3.4	0.7	27.3	21.2
2	Number of poor people (million)	7.0	9.6	27.3	264
3	Average reduction rate within the past decade (percent)	-8.9	-28.0	-2.8	-8.1
4	Average reduction rate within the past two decades (percent)	-3.9	-15.1	-4.5	-2.2
5	Population (million)	206.0	1371.2	99.9	1247.2

Panel B: Rankings

No	Outcomes	Country			
		Brazil	China	Ethiopia	India
1	Headcount poverty rate	2	1	4	3
2	Number of poor people	1	2	3	4
3	Average reduction rate within the past decade	2	1	4	3
4	Average reduction rate within the past two decades	3	1	2	4
	Average ranking	2.0	1.3	3.3	3.5

Note: All estimates are based on the WDI database. The headcount poverty rates, the number of poor people, and the population figures are in 2015 for Brazil, China, and Ethiopia, and in 2011 for India. Given available data, the past decade is defined as the period 2005-15 for Brazil and China, 2004-15 for Ethiopia, and 2004-11 for India. The past two decades is defined as the period 1995-2015 for Brazil and Ethiopia, 1996-2015 for China, and 1993-2011 for India. The rankings in Panel B are based on the estimation results in Panel A.

Table 4: Overview of United Nations' SDG Database, 2012-2016

No	Sustainable Development Goal	(1) Countries	(2) Indicators	(3) Data Points	(4) Coverage (%)
1	No poverty	198	7	1,951	11.2
2	Zero hunger	206	9	3,279	20.3
3	Good health and well-being	220	25	12,138	36.1
4	Quality education	210	10	2,901	21.2
5	Gender equality	196	8	1,650	9.5
6	Clean water and sanitation	235	6	3,431	25.1
7	Affordable and clean energy	227	4	2,912	39.0
8	Decent work and economic growth	217	12	6,164	29.1
9	Industry, Innovation, and Infrastructure	222	11	5,186	34.7
10	Reduced inequalities	203	5	1,810	13.2
11	Sustainable cities and communities	204	6	1,066	5.7
12	Responsible consumption and production	196	2	1,121	6.9
13	Climate action	153	2	436	4.4
14	Life below water	184	2	946	7.6
15	Life on land	245	8	6,434	36.9
16	Peace, justice, and strong institutions	215	9	2,009	7.0
17	Partnership for global development	236	8	5,219	16.8
	Overall	**247**	**134**	**58,653**	**19.3**

Note: The SDG database was downloaded from the UN's database on October 20 2018. The number of countries in the world is 249 and the total number of SDG indicators (with duplicates) is 244, as shown in more details in Appendix 2, Table 2.3. The coverage (column 4) is the percentage of the available data points in the UN's SDG database for each goal, which is calculated as the ratio of column 3 in Table 4 and column 3 in Table 2.3.

Table 5: A Grouping of SDG Goals and Indicators by Theme

| No	Theme	UN Theme	Number of Indicators	Topics	Example
1	Economic	Prosperity	89	Poverty; Hunger; GDP; Employment; Industrialization & Innovation	1.1.1. Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)
2	Health & Human Development	People	80	Health; Education; Gender Equality; Human Settlement; Technology	4.2.1. Proportion of children under 5 years of age who are developmentally on track in health, learning and psychosocial well-being, by sex
3	Governance	Partnership and Peace	14	Laws; Global Governance; Justice	16.3.1. Proportion of victims of violence in the previous 12 months who reported their victimization to competent authorities or other officially recognized conflict resolution mechanisms
4	Environment	Planet	61	Water; Energy; Sustainable Development; Climate Change	6.1.1. Proportion of population using safely managed drinking water services
Total			**244**		

Note: The full list of indicators is provided in Table 2.5 in Appendix 2.

Table 6: Comparing Different Aggregation Methods to Interpret the SDGs and Some Other Indexes

No.	Method	Definition	Example	Advantages	Disadvantages
1	Dashboard	Use the 232 indicators as is	SDG Indicators (United Nations, 2018)	i) (Almost) No theory, and no need for aggregation function ii) Leave each indicator in its raw form, and thus offers straightforward interpretation for each indicator iii) May offer easier empirical analysis	Difficult to interpret the performance of all indicators as a whole; performance of one indicator may be misinterpreted as that of all indicators
2	Empirical arithmetic mean	Take average value of all indicators	SDG Index (Sachs et al., 2017)	i) Simple theory ii) Average values offer straightforward interpretation iii) Allow for decomposition	Either equal weights or unequal weights need to be justified well
3	Geometric mean (or their product)	Multiply all N components together and raise to the power of $1/N$	Human Development Index (UNDP, 2010); Human Capital Index (Kraay, 2018)	i) Multiplicative form emphasizes lower values	If any indicator equals 0, the index will be 0.
4	Counting approach	Calculates the sum of the values of the achievements as a share of the maximum total value that could be achieved.	Country Statistical Capacity Index (Cameron et al., 2018)	i) General counting approach has been widely used ii) Flexible functional form (i.e., more indicators can be added, but they do not change the weight of the main dimension) iii) Allow for decomposition	More complex theory Must clearly define the main dimensions and sub-dimension indicators.

Note: More technical details on the aggregation methods are provided in Appendix 1, Part A.

Figure 1: Global Trends of GDP and PM 2.5 Matter, 1990-2016
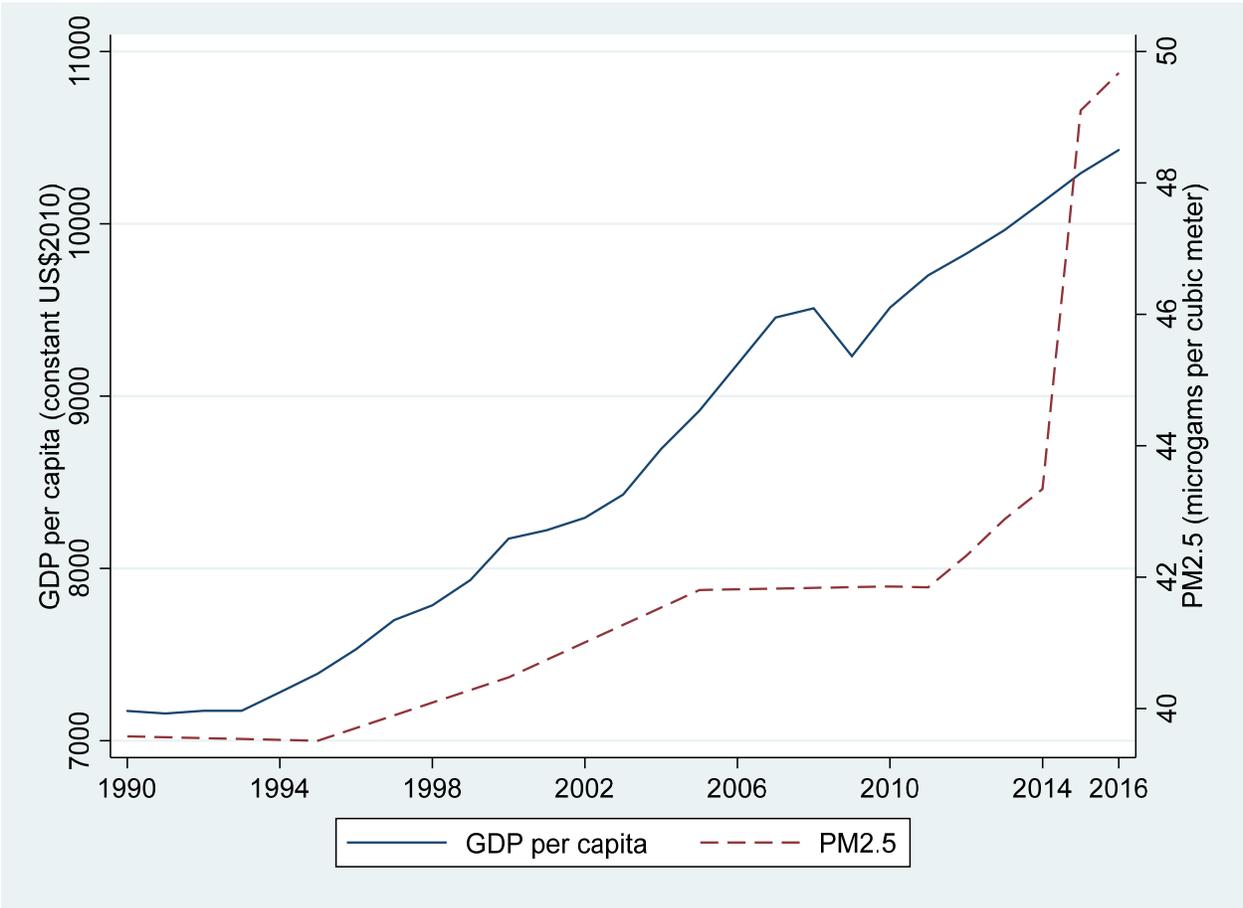
Source: Authors' calculation from World Bank's World Development Indicators Database.

Figure 2: Trends of GDP and PM 2.5 Matter for Two Countries, 1990-2016
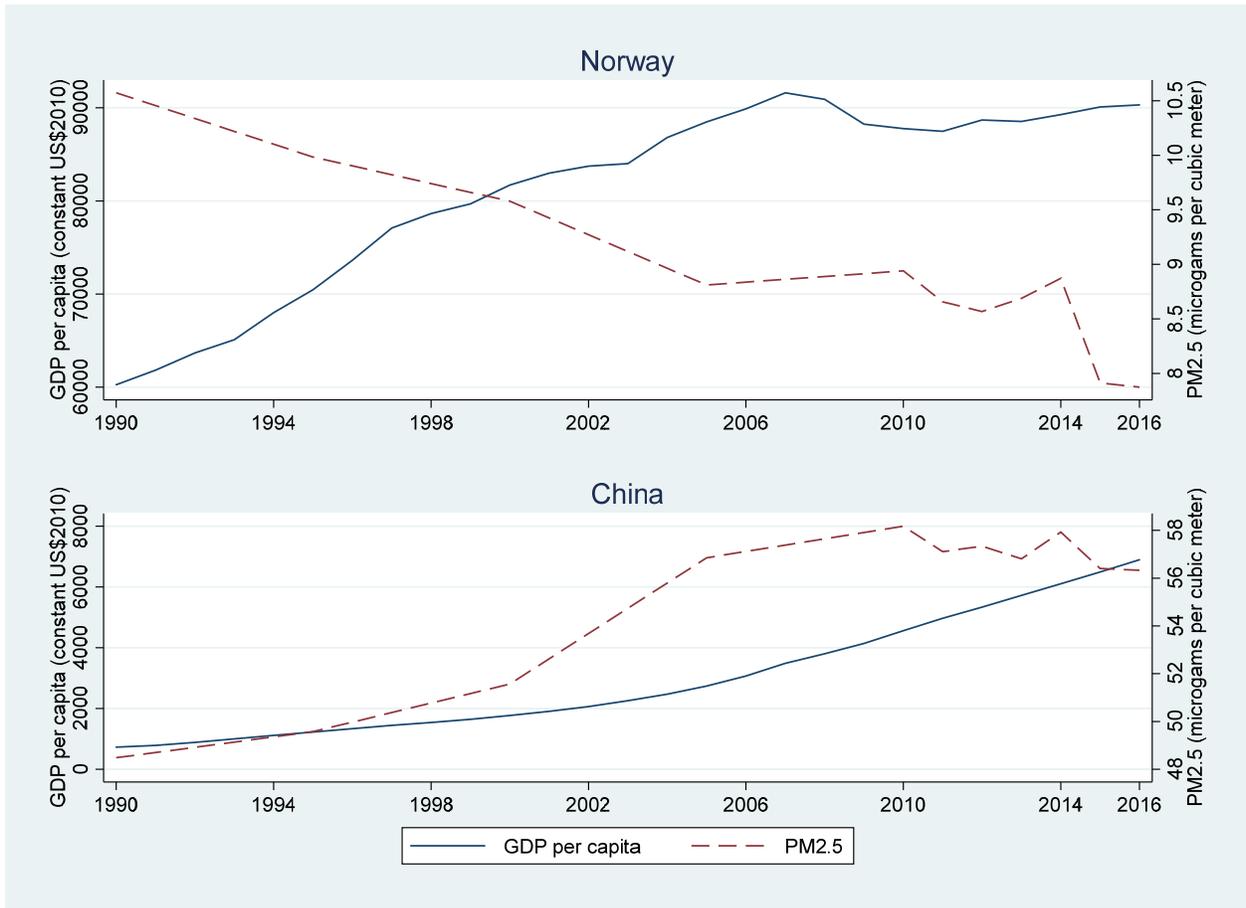
Figure 2: Trends of GDP and PM 2.5 Matter for Two Countries, 1990-2016

Source: Authors' calculation from World Bank's World Development Indicators Database.

Figure 3: Trends in Poverty Rate and Number of Poor People by Region, 1980-2016
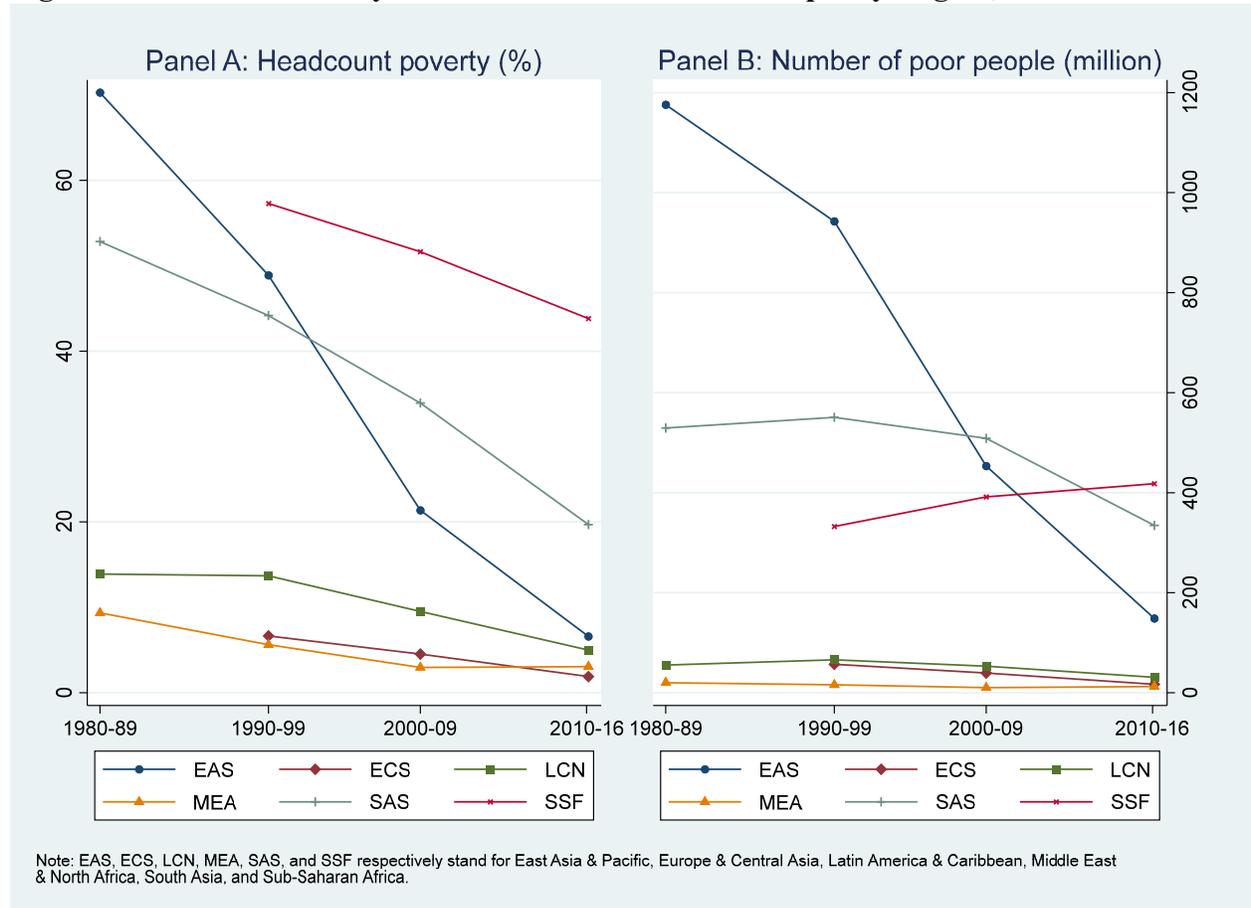
Note: EAS, ECS, LCN, MEA, SAS, and SSF respectively stand for East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, South Asia, and Sub-Saharan Africa.

Source: Authors' calculation from World Bank's World Development Indicators Database.

Figure 4: Different Levels in Consumption Growth from Household Surveys and National Accounts, China and India
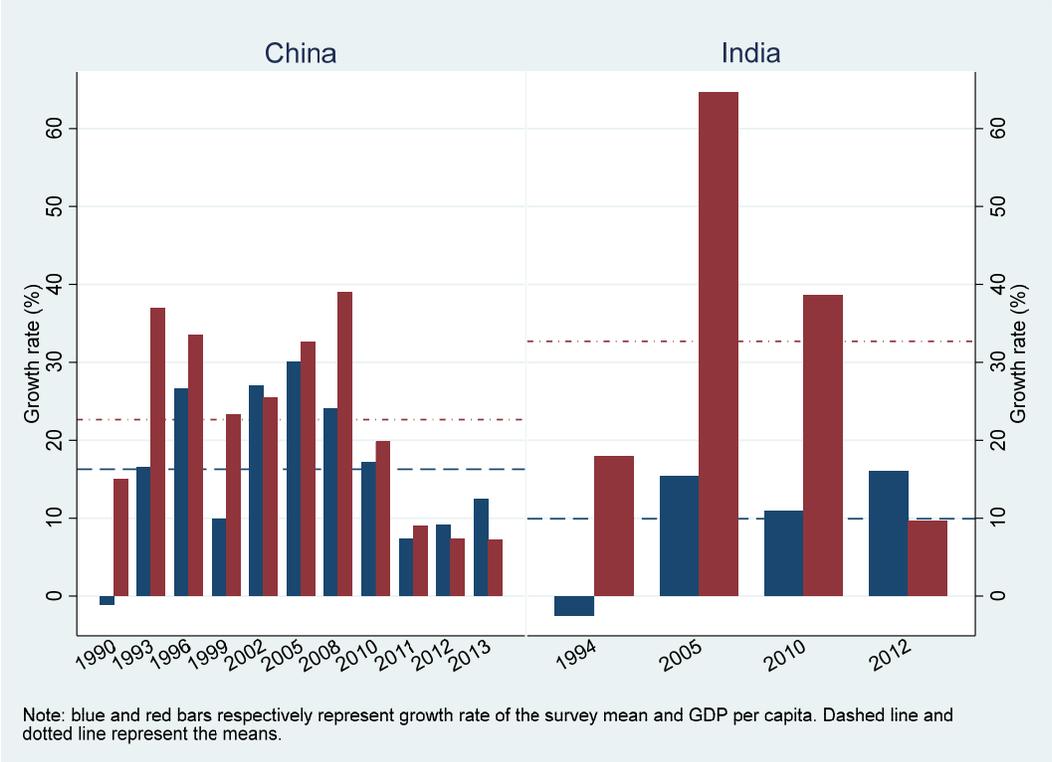
Note: blue and red bars respectively represent growth rate of the survey mean and GDP per capita. Dashed line and dotted line represent the means.

**Source**: Authors' calculation from World Bank's PovCalNet and World Development Indicators Databases.

Figure 5: Different Patterns of GDP per capita Growth over Time, 2011-2015
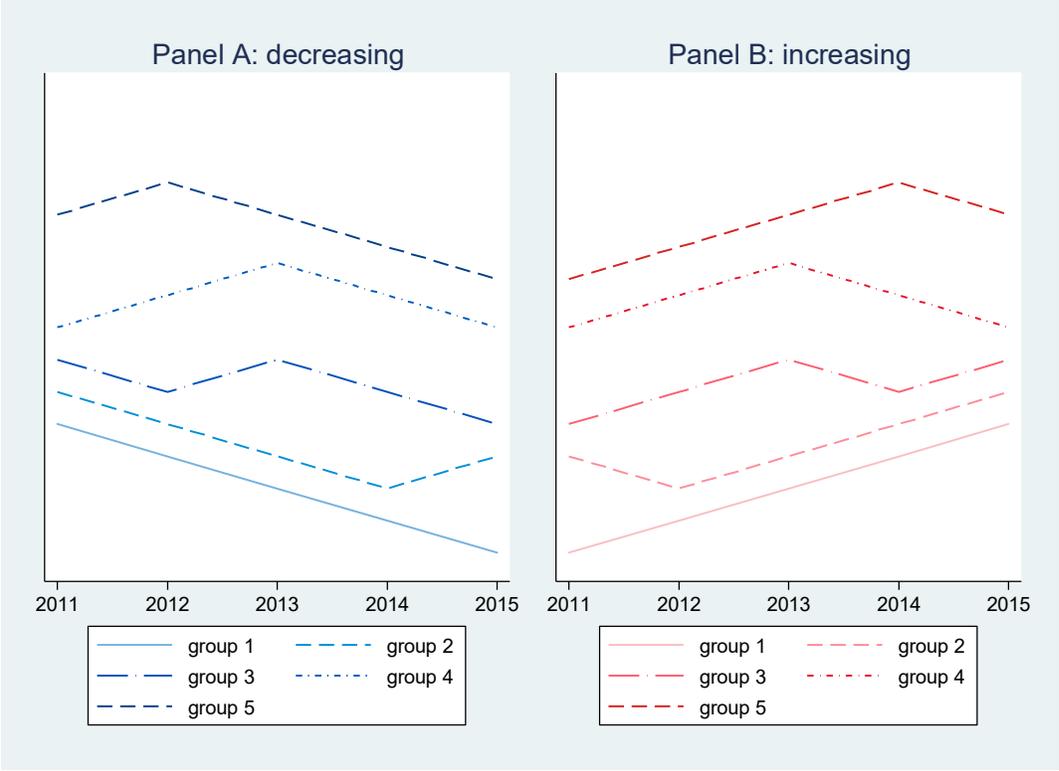
Figure 6: Data Flow for SDG Number 6 Related to Water
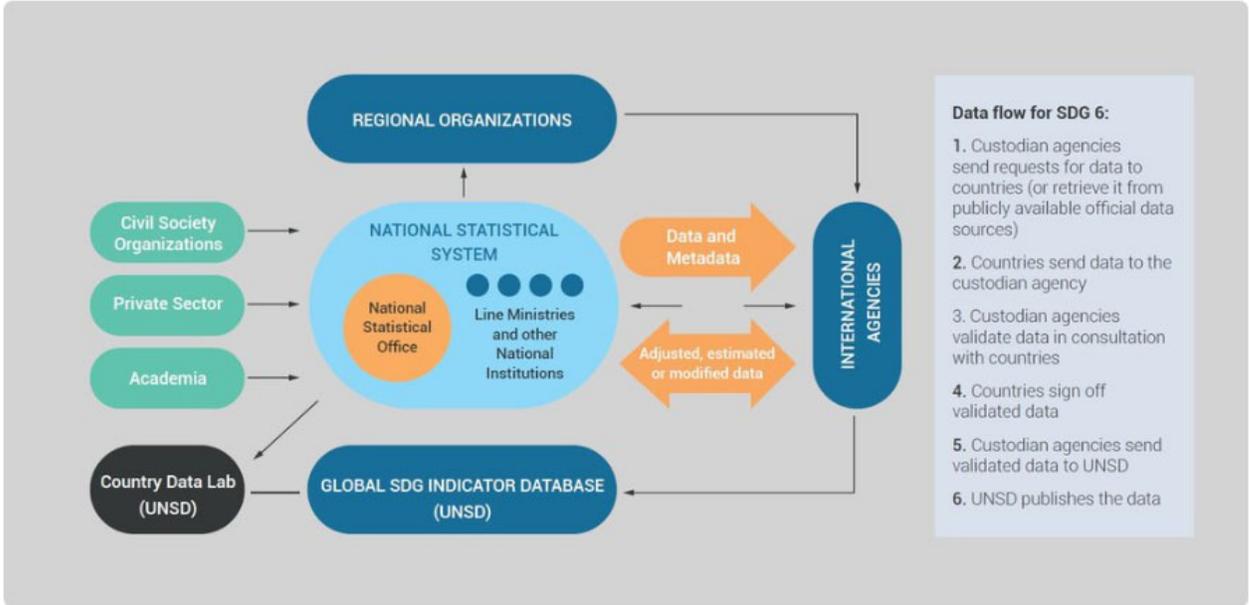
Figure 7: Imputation-Based Estimates of Electrification Rates for African Countries, 1990-2016
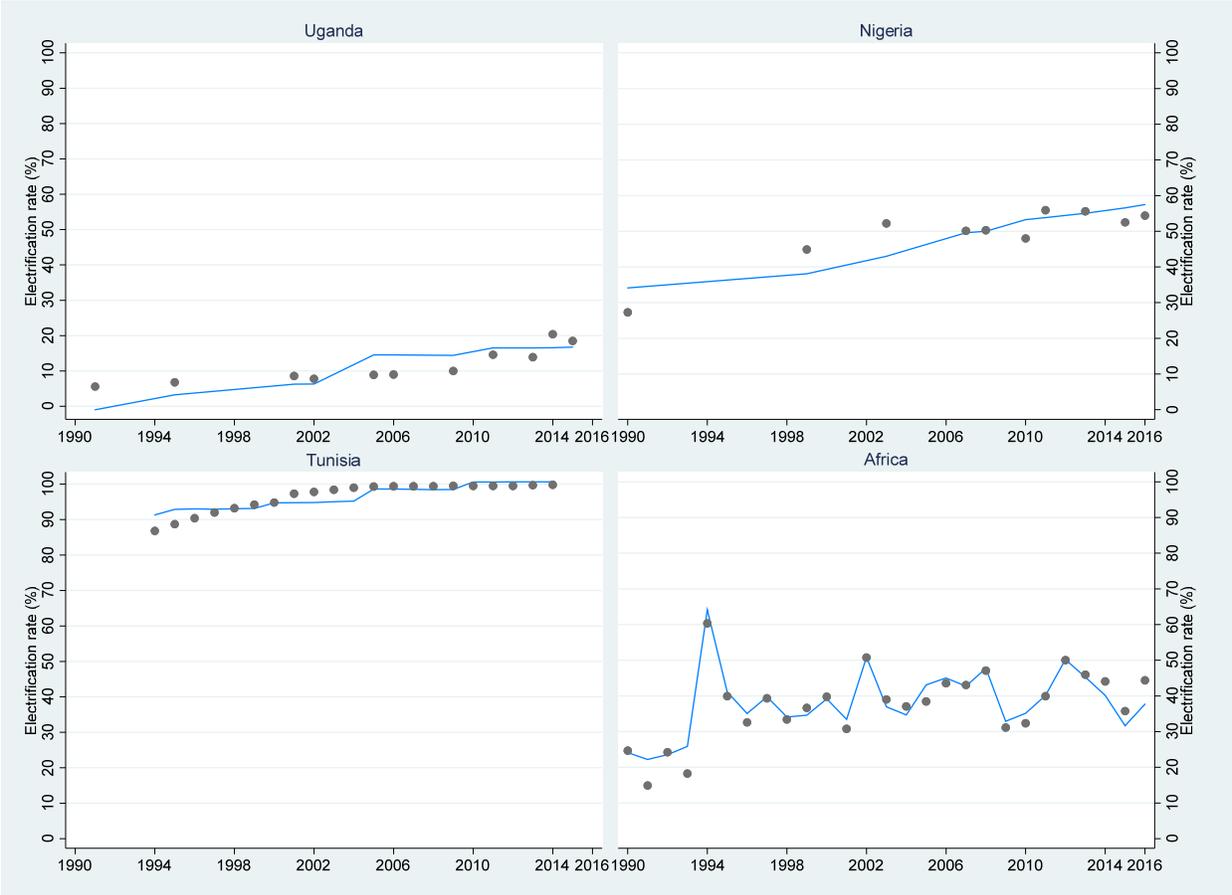
Appendix 1: Technical Appendix

Part A. Techniques to Track (MDG and SDG) Progress

We provide in Part A of this Appendix a brief review of the main techniques in selected works that have been employed to track progress on the MDGs and the SDGs. While these techniques appear to be not very complex, they have been observed to be highly prone to misinterpretation. We also add further details to expand on certain points where it is useful to do so. Our objectives are twofold: i) first, clearly lay out the technical details to make them more accessible, ii) and second, highlight the different conclusions that may be reached when different techniques are employed.

Let C_{it} represent the current achievement on an indicator *I*, for *i*= 1,…, 232 at year *t, t= 1,…T.* A superscript *h* can be added to C_i to denote whether the achievement is at the maximal level (C_i^{max}) or the minimal level (C_i^{min}). But to make notation less cluttered, we also leave out the subscript *t* when is it not necessary to discuss the change over time.

Dashboard Approach

To measure the current achievement on an indicator against a desirable benchmark, two common ways are used. One is to look at the absolute difference (A_i) between the two

$$A_i = C_i - C_i^{max} \tag{1.1}$$

the other is to look at the relative difference (R_i)

$$R_i = \frac{C_i}{C_i^{max}} \tag{1.2}$$

The different implications of using either the absolute difference or the relative difference to measure progress is well illustrated by the following hypothetical example from Easterly (2009). Suppose Latin America could halve the poverty rate from 10% to 5%, and Africa could reduce its poverty rate by around one-third from 50% to 35%. Certainly, by this relative difference metric, Latin America does better than Africa. However, by the absolute difference metric, we have the opposite result where Africa's reduction is 15%, three times the corresponding figure of 5% for Latin America. Yet, as discussed earlier (Section II.2), if we further assume that the population in Latin America were more than three times larger than that in Africa, and use a different metric of the number of people being lifted out of poverty, then Latin America is the winner.

We can then expand Equations (1.1) and (1.2) to measure progress between year *1* and *2* as

$$\Delta D_i = \frac{D_{i2} - D_{i1}}{D_{i1}} \tag{1.3}$$

where Δ denotes the change (or growth rate), and D_i equals either A_i or R_i . One interesting fact that has been observed about ΔD_i is that, this quantity depends to a large extent on the value of its denominator A_{i1}. This observation can be subsequently translated into progress (on the MDGs) as indicators with a low initial value would likely have larger growth rate, and vice versa (see, e.g., Easterly (2009) and Vandemoortele (2009)).

To avoid this issue, Fukuda-Parr et al. (2013) propose that we should compare the average rate of change in two periods, one before the commitment year and one after.[33] As such, we should consider countries that have a better growth rate as having better performance, rather than insist on countries having to reach the set targets. In particular, their formula for each period p, for $p= 1, 2$, is

$$\Delta D_{ip} = \frac{D_{it_2} - D_{it_1}}{t_2 - t_1} \tag{1.4}$$

An alternative to using Equation (1.2) to define the relative difference is to standardize both the numerator and the denominator in this equation by their respective differences from the minimal value. This is also known as a distance-to-the-frontiers approach, and is employed by Sachs et al. (2018).[34]

$$R_i^s = \frac{C_i - C_i^{min}}{C_i^{max} - C_i^{min}} \tag{1.5}$$

Index Approach

Let j denote the group of the indicators, for $j= 1,\dots, J$ and k denote the country, for $k= 1,\dots, K$. Cameron et al. (2018) propose a weighted arithmetic mean method in the context of measuring countries' statistical capacity, which is motivated by the counting approach of Atkinson (2003). This method defines weights by the level, where all components at the same level (or in the same group) would be assigned an equal weight.

We can use either the 17 SDGs or the proposed four categories discussed earlier (in Table 5) for the grouping. By this method, these groups would have an equal weight of $\frac{1}{J}$ in their contributions to the total scores (i.e., $J= 17$ or 4 depending on the specific grouping that is used). The indicators within each goal also have the same weight, which is defined as the inverse of the number of indicators in this goal. More specifically, the overall index can be calculated as follows

$$I^c = \sum_{j=1}^{J} \frac{1}{J} \sum_{i=1}^{N_{jk}} \frac{1}{N_{jk}} R_{ijk}^s \tag{1.6}$$

where N_{jk} is the number of indicators for the jth group. In other words, this index follows a nested structure where each goal is assigned an equal weight of $\frac{1}{J}$, and each indicator within goal j is assigned an equal weight of $\frac{1}{N_{jk}}$ (i.e., the inverse of the number of indicators under this goal).

Yet, in practice, as we discussed in Section II.3, data can be missing for indicators (and goals) for countries. Let us use the tilde sign (~) to denote data availability. Sachs et al. (2017) employ a modified version of Equation (1.6)

---

[33] Note that Fukuda-Parr et al.'s (2013) commitment year was 2000 in their analysis for the MDGs, and can be 2015 if applied to the SDGs.
[34] See Sachs et al. (2018) for a detailed discussion on the criteria to define C_i^{max}.

$$I^S = \sum_{j=1}^{\widetilde{N}_j} \frac{1}{\widetilde{N}_j} \sum_{i=1}^{\widetilde{N}_{jk}} \frac{1}{\widetilde{N}_{jk}} R_{ijk}^S \tag{1.7}$$

where $\widetilde{N}_{jk}$ is the number of indicators for the jth SDG that country k has data for (rather than the number of indicators for the jth SDG as formally stated for all countries). Similarly, $\widetilde{N}_j$ is now the number of SDGs for which country i has data for (rather than just the 17 goals). In other words, Equation (1.7) offers an empirical arithmetic mean calculation for Equation (1.6).

Clearly, compared to Equation (1.6), Equation (1.7) is driven by the data that are available for both i) the indicators ($\widetilde{N}_{jk}$) and ii) the goals ($\widetilde{N}_j$). As such, the weights placed on the indicators ($\frac{1}{\widetilde{N}_{jk}}$) and the goals ($\frac{1}{\widetilde{N}_j}$) can vary from country to country, depending on data availability for each country. If the missing data issue affects all countries randomly, then these variations in weights should not affect the country scores. But if it is not the case, this can lead to biased results.[35]

Kraay's (2018) human capital index offers an alternative method, whereby instead of aggregating using the arithmetic mean as in Equation (1.6), we can just multiply all indicators (or goals) together

$$I^K = \prod_{j=1}^{J} \prod_{i=1}^{N_{jk}} R_{ijk}^S \tag{1.8}$$

An earlier variant of Equation (1.8) is UNDP's human development index, which would provide the geometric mean of the product obtained from Equation (1.8) (UNDP, 2010). But as discussed earlier (Section III), there are two disadvantages with these measures. One is that, if any indicator is 0, the overall index will be 0 by construction, and the other is that, a product of all 232 indicators is much more unwieldy and harder to interpret than a product of just three indicators (components).

Part B. Imputation Models

We present in Part B of this Appendix the statistical models that can be employed to provide imputation-based estimates for missing electrification rates. We can apply a two-level nonparametric model without covariates in the spirit of Bonjour et al.'s (2013) for this purpose, which is defined as follows

$$Y_{tk} = \beta_0 + \boldsymbol{\beta_1'} \boldsymbol{X}_{tk} + \mu_k + \varepsilon_{tk} \tag{1.9}$$

where Y_{tk} is the electrification rate for country k in year t. The vector of variables $\boldsymbol{X}_{tk}$ includes regional dummy variables, and the time splines that are generated in the way described by Bonjour et al.'s (2013). That is, the time variable is centered at the median date of the database, and then transformed into a natural cubic spline with four knots. The covariance model is chosen to be unstructured. μ_k is the country random effects.

---

[35] In an extremely hypothetical scenario, country X that is determined to manipulate their scores can just focus on improving one indicator in one goal only, since by Equation (1.8), this country is not penalized for missing data on all the remaining 16 goals and other indicators. Perhaps this is one of the reasons that, in their most recent report, Sachs et al. (2018) now switch to using a fixed equal weight for every SDG instead of the empirical weight $\frac{1}{\widetilde{N}_j}$ employed in their earlier reports.

To improve the model estimation, we adjust Equation (1.9) by extending it to a three-level model, where the new third level is the region, and we also add some covariates. The model is defined as follows

$$Y_{tkh} = \beta_0 + \boldsymbol{\beta'_1}\boldsymbol{Z}_{tkh} + \mu_{kh} + \eta_h + \varepsilon_{tkh} \qquad (1.10)$$

where Y_{tkh} is the electrification rate for country k in region h in year t. The vector of variables $\boldsymbol{Z}_{tkh}$ now includes log GDP per capita, the share of urban population, and dummy variables for 5-year periods, and the interaction terms between the first two variables and the period dummy variables. The covariance model is chosen to be unstructured. μ_{kh} is the country random effects (that is nested within region h), and η_h is the region random effects.

We construct a data set that consists of the IEA et al.'s (2018) data on actual electrification rate for the period 1990-2016, the WDI's GDP data, and UNDESA's (2014) population data. We apply Equation (1.9) and Equation (1.10) to this data set and find Equation (1.10) to better perform Equation (1.9) on several test statistics, including the AIC, the square root of mean square error (MSE), and the mean absolute error (MAE). For example, the MSE and MAE based on Equation (1.9) are respectively 20.6 and 2.8; both these numbers are larger than the corresponding figures of 15.1 and 2.2 obtained from Equation (1.10). We subsequently produce Figure 7 using Equation (1.10).

Appendix 2: Additional Tables and Figures

Table 2.1. Typology of Welfare Transition Dynamics over Two Periods

Scenario	Pro-poor Growth	Welfare Category			Notes
		1st group	2nd group	3rd group	
		Poor	Vulnerable	Middle Class	
1	Strongest/ Most positive	-	-	+	first and second group reduce, and third group expands
2	More positive	-	+	+	first group reduces, and second and third group expands
3	Positive	-	+	-	first and third group reduce, and second group expands
4	Negative	+	-	+	first and third group expand, and second group reduces
5	More negative	+	-	-	first group expands, and second and third group reduce
6	Weakest/ Most negative	+	+	-	first and second group expand, and third group reduces

Note: The signs (-) and (+) respectively stand for decrease and increase. Pro-poor growth is defined as the dynamics that are most beneficial to the different categories in this order: Lowest Income, Middle Income, and Top Income. This typology is modified based on Dang and Lanjouw (2016).

Table 2.2: Change in Shared Prosperity and Gini Coefficients for Sub-Saharan African Countries (percentage)

(1)	(2)	(3)	(4)	(5)	(6)	(7)
No	Country	Change in poverty	Gini coefficient		Change in Gini coefficient	Pro-poor growth scenario
			1st period	2nd period		
1	Chad	-36.1	0.40	0.42	0.02	**
2	Botswana	-28.7	0.64	0.61	-0.03	**
3	Mauritania	-27.7	0.41	0.38	-0.03	***
4	Ghana	-20.7	0.40	0.43	0.02	**
5	Uganda	-19.1	0.44	0.44	0.01	**
6	Congo, Dem. Rep.	-13.5	0.44	0.42	-0.02	**
7	Mozambique	-12.6	0.48	0.46	-0.02	**
8	Rwanda	-8.7	0.53	0.53	0.00	**
9	Tanzania	-6.7	0.41	0.38	-0.02	**
10	Sierra Leone	-6.5	0.39	0.34	-0.05	*
11	Ethiopia	-3.9	0.29	0.32	0.02	***
12	Togo	-1.4	0.42	0.46	0.03	***
13	Eswatini	-1.3	0.53	0.52	-0.01	***
14	Malawi	-1.0	0.41	0.48	0.07	***
15	Senegal	0.9	0.41	0.41	0.00	---
16	Nigeria	5.5	0.35	0.36	0.01	---
17	Burkina Faso	6.3	0.43	0.40	-0.04	--
18	Zambia	7.8	0.55	0.57	0.01	--
19	Madagascar	9.5	0.39	0.41	0.02	--
20	Côte d'Ivoire	15.1	0.42	0.43	0.00	--
21	Cameroon	34.5	0.42	0.43	0.00	--
	Regional average	-5.2	0.44	0.44	0.00	**

Note: Authors' calculation based on household survey data. Household heads' age is between 25 and 55 in the first survey round and adjusted accordingly for the second survey round. The poverty line and vulnerability line are respectively set at $1.9/day and $4.3/day in 2011 PPP dollars for both periods. Pro-poor growth scenarios are based on the classification provided in Appendix 2, Table 2.1. Countries are ranked in a decreasing order of reduction in headcount poverty (column 3). The regional average is a simple average (unweighted). Most household surveys were implemented in the late 2000s. Adopted with modifications from Table 5 in Dang and Dabalen (2019).

Table 2.3: Official Numbers of Countries and Indicators, 2012-2016

No	Sustainable Development Goal	(1) Countries	(2) Indicators	(3) Data Points
1	No poverty	249	14	17,430
2	Zero hunger	249	13	16,185
3	Good health and well-being	249	27	33,615
4	Quality education	249	11	13,695
5	Gender equality	249	14	17,430
6	Clean water and sanitation	249	11	13,695
7	Affordable and clean energy	249	6	7,470
8	Decent work and economic growth	249	17	21,165
9	Industry, Innovation, and Infrastructure	249	12	14,940
10	Reduced inequalities	249	11	13,695
11	Sustainable cities and communities	249	15	18,675
12	Responsible consumption and production	249	13	16,185
13	Climate action	249	8	9,960
14	Life below water	249	10	12,450
15	Life on land	249	14	17,430
16	Peace, justice, and strong institutions	249	23	28,635
17	Partnership for global development	249	25	31,125
	Overall	**249**	**244**	**303,780**

Note: The number of countries in the world is 249 (United Nations, 2018b) and the total number of SDG indicators (with duplicates) is 244 (United Nations, 2018a). The number of the data points in column 3 is calculated as the product of the number of countries (column 1), the number of indicators (column 2), and 5 years for the period 2012-2016.

**Table 2.4: Overview of United Nations' SDG Database, 2000-2018**

No	Sustainable Development Goal	(1) Countries	(2) Indicators	(3) Data Points	(4) Coverage (%)
1	No poverty	211	8	6,138	9.3
2	Zero hunger	216	9	11,273	18.3
3	Good health and well-being	224	25	37,678	29.5
4	Quality education	212	10	7,887	15.2
5	Gender equality	197	8	5,381	8.1
6	Clean water and sanitation	235	9	10,790	20.7
7	Affordable and clean energy	227	4	10,707	37.7
8	Decent work and economic growth	217	12	18,948	23.6
9	Industry, Innovation, and Infrastructure	223	11	17,994	31.7
10	Reduced inequalities	205	5	5,559	10.7
11	Sustainable cities and communities	209	7	2,497	3.5
12	Responsible consumption and production	196	2	3,490	5.7
13	Climate action	167	3	1,054	2.8
14	Life below water	203	2	3,583	7.6
15	Life on land	245	9	22,065	33.3
16	Peace, justice, and strong institutions	215	10	4,306	4.0
17	Partnership for global development	237	12	16,419	13.9
	Overall	**248**	**146**	**185,769**	**16.1**

Note: The SDG database was downloaded from the UN's database on October 20 2018 for the period 2000-2018. The number of countries in the world is 249 and the total number of SDG indicators (with duplicates) is 244, as shown in more details in Appendix 2, Table 2.3. The coverage (column 4) is the percentage of the available data points in the UN's SDG database for each goal, which is calculated as the ratio of column 3 in Table 4 and column 3 in Table 2.3.

Table 2.5: A Grouping of SDG Goals and Indicators by Theme, with Full List of Indicators

No	Theme	Number of Indicators	Topics	Indicators
1	Economic	89	Poverty; Hunger; GDP; Employment; Industrialization & Innovation	1.1.1; 1.2.1; 1.2.2; 1.3.1; 1.4.1; 1.4.2; 1.5.1; 1.5.2; 1.5.3; 1.a.1; 1.a.2; 1.b.1
				2.1.1; 2.1.2; 2.2.1; 2.2.2; 2.3.1; 2.3.2; 2.4.1; 2.5.1; 2.5.2; 2.a.1; 2.a.2; 2.b.1; 2.b.2; 2.c.1
				8.1.1; 8.2.1; 8.3.1; 8.4.1; 8.4.2; 8.5.1; 8.5.2; 8.6.1; 8.7.1; 8.8.1; 8.8.2; 8.9.1; 8.9.2; 8.10.1; 8.10.2; 8.a.1; 8.b.1
				9.1.1; 9.1.2; 9.2.1; 9.2.2; 9.3.1; 9.3.2; 9.4.1; 9.5.1; 9.5.2; 9.a.1; 9.b.1; 9.c.1
				10.1.1; 10.2.1; 10.3.1; 10.4.1; 10.5.1; 10.6.1; 10.7.1; 10.7.2; 10.a.1; 10.b.1; 10.c.1
				16.6.1; 16.6.2
				17.18.2; 17.1.1; 17.1.2; 17.3.1; 17.3.2; 17.4.1; 17.5.1; 17.9.1; 17.10.1; 17.11.1; 17.12.1; 17.13.1; 17.14.1; 17.15.1; 17.16.1; 17.17.1; 17.18.1; 17.18.3; 17.19.1; 17.19.2
2	Health & Human Development	80	Health; Education; Gender Equality; Human Settlement; Technology	3.1.1; 3.1.2; 3.2.1; 3.2.2; 3.3.1; 3.3.2; 3.3.3; 3.3.4; 3.3.5; 3.4.1; 3.4.2; 3.5.1; 3.5.2; 3.6.1; 3.7.1; 3.7.2; 3.8.1; 3.8.2; 3.9.1; 3.9.2; 3.9.3; 3.a.1; 3.b.1; 3.b.2; 3.c.1; 3.d.1
				4.1.1; 4.2.1; 4.2.2; 4.3.1; 4.4.1; 4.5.1; 4.6.1; 4.7.1; 4.a.1; 4.b.1; 4.c.1
				5.1.1; 5.2.1; 5.2.2; 5.3.1; 5.3.2; 5.4.1; 5.5.1; 5.5.2; 5.6.1; 5.6.2; 5.a.1; 5.a.2; 5.b.1; 5.c.1
				11.1.1; 11.2.1; 11.3.1; 11.3.2; 11.4.1; 11.5.1; 11.5.2; 11.6.1; 11.6.2; 11.7.1; 11.7.2; 11.a.1; 11.b.1; 11.b.2; 11.c.1
				16.1.1; 16.1.2; 16.1.3; 16.1.4; 16.2.1; 16.2.2; 16.2.3
				17.6.1; 17.6.2; 17.7.1; 17.8.1
3	Governance	14	Laws; Global Governance; Justice	16.3.1; 16.3.2; 16.4.1; 16.4.2; 16.5.1; 16.5.2; 16.7.1; 16.7.2; 16.8.1; 16.9.1; 16.10.1; 16.10.2; 16.a.1; 16.b.1
4	Environment	61	Water; Energy; Sustainable Development; Climate Change	6.1.1; 6.2.1; 6.3.1; 6.3.2; 6.4.1; 6.4.2; 6.5.1; 6.5.2; 6.6.1; 6.a.1; 6.b.1
				7.1.1; 7.1.2; 7.2.1; 7.3.1; 7.a.1; 7.b.1;
				12.1.1; 12.2.1; 12.2.2; 12.3.1; 12.4.1; 12.4.2; 12.5.1; 12.6.1; 12.7.1; 12.8.1; 12.a.1; 12.b.1; 12.c.1
				13.1.1; 13.1.2; 13.2.1; 13.3.1; 13.3.2; 13.a.1; 13.b.1
				14.1.1; 14.2.1; 14.3.1; 14.4.1; 14.5.1; 14.6.1; 14.7.1; 14.a.1; 14.b.1; 14.c.1
				15.1.1; 15.1.2; 15.2.1; 15.3.1; 15.4.1; 15.4.2; 15.5.1; 15.6.1; 15.7.1; 15.8.1; 15.9.1; 15.a.1; 15.b.1; 15.c.1
Total		**244**		

Figure 2.1. Different Levels in Consumption Growth from Household Surveys and National Accounts, Other Countries

Note: blue and red bars respectively represent growth rate of the survey mean and GDP per capita. Dashed line and dotted line represent the means.

Source: Authors' calculation from World Bank's PovCalNet and World Development Indicators Databases.