2019/06

Social Development Working Papers

LEAVING NO ONE BEHIND: A METHODOLOGY TO IDENTIFY THOSE FURTHEST BEHIND IN ACCESSING OPPORTUNITIES IN ASIA AND THE PACIFIC

Predrag Savic
Yichun Wang

UNITED NATIONS
ESCAP
Economic and Social Commission for Asia and the Pacific

Leaving no one behind: A methodology to identify those furthest behind in accessing opportunities in Asia and the Pacific

Predrag Savic and Yichun Wang

Abstract

With the adoption of the 2030 Agenda for Sustainable Development, the United Nations Member States pledged to reduce inequality in all its forms through Sustainable Development Goal 10, while ensuring that "no one will be left behind." This paper proposes a methodology that governments and stakeholders can use in their countries to measure inequality of opportunity, using data from nationally representative household surveys. The Dissimilarity Index (D-index) allows a comparison of inequality of opportunity levels among countries. Its decomposition enables the researchers to determine which circumstances contribute mostly to the observed inequality. The paper also proposes an innovative methodology, the classification tree analysis, to identify households and individuals furthest behind in access to basic opportunities in the Asia-Pacific region. The classification trees offer a practical way of operationalizing the pledge to leave no one behind (LNOB), accelerating national progress to achieve the Sustainable Development Goals (SDGs).

Key words: Inequality of opportunity, Asia and the Pacific, Leaving No One Behind

***

Authors

Predrag Savić
Associate Social Affairs Officer
Social Development Division
United Nations Economic and Social Commission for Asia and the Pacific
Office: (+66) 2 288 1516 | predrag.savic@un.org

Yichun Wang
Statistician
Carrer Melissa 45, 7-14
Paterna, 46980, Spain
yichun.wang@gmail.com

Acknowledgements

The research of this paper was conducted at the Sustainable Social Transformation Section (SETS) of the Social Development Division of the United Nations Economic and Social Affairs Commission for Asia and the Pacific (ESCAP). It was developed under the overall supervision of Patrik Andersson, Chief, Sustainable Socioeconomic Transformation Section, and with inputs from Ermina Sokou, Social Affairs Officer.

CONTENTS

1. Introduction

Over the past decades, the Asia-Pacific region has made considerable strides in social development, driven by economic growth, which generated new jobs, increased incomes and improved access to basic services and other opportunities. In several countries, more than 80 per cent of the extreme poor have been lifted out of poverty.

Despite this sustained economic development and substantial reductions in poverty, progress has disproportionately benefited the richest members of society, increasing inequalities between the rich and poor in many parts of Asia and the Pacific. 400 million people are estimated to still live below the extreme poverty line of $1.90 a day and another 800 million people are trapped on incomes above $1.90, but below the moderate poverty level of $3.20 a day. An even larger number of people are deprived of basic services and opportunities. Less than 4 out of 10 people in the region have access to health care, and out-of-pocket expenditures for health care are among the highest in the world.[1] One and a half billion people in the region lack access to improved sanitation. The largest numbers of poor and deprived populations live in rural areas.

The region's income inequality, measured by the Gini coefficient, has increased by over 5 percentage points in the last 20 years. This increase goes contrary to almost all other regions. Steep increases in the incomes of the richest have often coincided with an increased concentration of wealth.[2] High inequality not only stifles economic progress, but also negatively affects feelings of trust and social cohesion.[3] It thus poses a formidable barrier to sustainable development. These rising levels of inequality within countries have sparked public concern and academic interest and contributed to a stand-alone goal on inequality in the 2030 Agenda for Sustainable Development. Sustainable Development Goal (SDG) 10 to 'reduce inequalities within and among countries' is thus a core policy priority to ensure a sustainable and prosperous future for all.

Large and often increasing inequalities also exist in access to opportunities, such as in educational attainment, ownership of a bank account and access to clean fuels. For example, despite high enrolments, in one third of the countries in the Asia-Pacific region, attendance rates in secondary education for the poorest quintile remain below 30 per cent, with four countries at 10 per cent or below. Meanwhile, attendance rates for children from the upper quintile in these countries can be as high as 80 per cent. This inequality of opportunity has sparked interest among policymakers and researchers, particularly as it is found to also perpetuate income inequality.

2. What is Inequality of Opportunity?

Income inequality is a long-standing concern across the developed and developing world, with academics, policymakers and civil society debating on its causes and consequences, as well as the best course of action for remedying the ills associated with high income inequality. The picture is muddled. Over the past twenty years, higher inequality has been shown to be linked to both

[1] Social Outlook for Asia and the Pacific 2018: Poorly Protected https://www.unescap.org/publications/social-outlook-asia-and-pacific-2018

[2] Inequality in Asia And The Pacific in The Era Of The 2030 Agenda For Sustainable Development, available at: https://www.unescap.org/sites/default/files/publications/ThemeStudyOnInequality.pdf

[3] United Nations, Economic and Social Commission for Asia and the Pacific (ESCAP) (2017). Sustainable Social Development in Asia and the Pacific: Towards a People-Centred Transformation. Sales No. E.17.II.F.15

higher growth and to lower growth.[4] For every country, especially in a region as diverse as Asia and the Pacific, a different story can be told.

Opposing worldviews, however, converge around the importance of ensuring equality in access to opportunities, or basic access to services to all. Beyond the debate of an ideal level of income or wealth inequality, what is more important is that people are given equal chances to improve their socioeconomic outcomes.

The inequality literature distinguishes between inequality of outcome and inequality of opportunity. While the former depicts the consequences of unequally distributed income and wealth, inequality of opportunity is concerned with access to key dimensions necessary for decent quality of life. The philosophical foundations of this approach lie in the work of John Rawls[5] and Amartya Sen[6]. Rawls was among the first modern political philosophers who articulated the importance of balancing personal liberties with distributive justice and fair options for all, arguing that public policy choices should focus on raising the welfare of the poorest people.[7] Rawls argued that a set of 'primary goods' should be made available for everyone, so that she or he would be able to realize their 'life plan'. Sen, later, argued that inequality could be re-examined from the perspective of human capability, looking at the 'means' rather than the 'ends' of development. Without equal opportunity, equitable outcomes cannot be secured.

A more complete theoretical framework for understanding and analysing inequality of opportunity has since then been developed. John Roemer refers to inequality of opportunity as the portion of inequality of outcome that can be credited to differences in 'individual circumstances.'[8]

Inequality of opportunity also combines concerns of equity and efficiency. The equity motivation is to create a level playing field. The efficiency motivation stipulates that we still do not want to equalize at the bottom and bring everyone down to the welfare level of the poorest person.[9] Instead, the goal is to ensure that everyone has access to basic services, as also envisioned in the 2030 Agenda for Sustainable Development.

Inequality of opportunity is reflected in the Agenda's vision of a "just, equitable, tolerant, open and socially inclusive world in which the needs of the most vulnerable are met." Target 10.3 of SDG 10 calls for ensuring equal opportunity and reducing inequalities of outcome, including by eliminating discriminatory laws, policies and practices and promoting appropriate legislation, policies and action. Target 10.3, however, is not the only one in the 2030 Agenda that relates to inequality of opportunity. The indicators of opportunities presented in this paper refer to specific SDGs that call for universal access to key services, such as education, health care, clean water and sanitation, among others.

---

[4] A brief but comprehensive literature review can be found in the IMF working paper: *Sharing the Growth Dividend: Analysis of Inequality in Asia*

[5] Rawls, John. 1971. A Theory of Justice. Cambridge, MA: Harvard University Press

[6] Sen, Amartya, and Geoffrey Hawthorne. 1985. The Standard of Living (Tanner Lectures in Human Values). Cambridge: Cambridge University Press.

[7] Ravallion, Martin, 2016. The Economics of Poverty. Oxford University Press.

[8] Roemer, John E. 1998. Equality of Opportunity. Cambridge, MA: Harvard University Press.

[9] Ravallion, Martin, 2016. The Economics of Poverty. Oxford University Press.

Focusing on inequality of opportunity also serves as a reminder that inequality is not a static phenomenon. Inequality of outcome among parents is also transmitted to their children, creating inequality of opportunity for the next generation. An adult's outcomes in terms of education and health will likely become the children's predetermined circumstances, creating more income or wealth inequality.[10]

3. LNOB in the context of the 2030 Agenda

To eliminate inequality of opportunity, the 2030 Agenda takes an ambitious yet pragmatic approach, stressing that no one should be left behind in any of its Goals - and that the furthest behind should become the focus of policymaking. It states:

"*As we embark on this great collective journey, we pledge that no one will be left behind. Recognizing that the dignity of the human person is fundamental, we wish to see the Goals and targets met for all nations and peoples and for all segments of society. And we will endeavour to reach the furthest behind first.*" (Paragraph 4)

Member States have also explicitly called on the United Nations and its agencies, funds and programmes to implement the Leave No One Behind (LNOB) pledge. The United Nations system has responded promptly, bringing the LNOB pledge at the core of its programming. The United Nations System Shared Framework for Action calls for "greater data disaggregation across a wider range of grounds for all SDG indicators; systematic analysis of available (disaggregated) data on marginalized groups; new tools for analysing horizontal and vertical inequalities, as well as discrimination, stigma, exclusion, and equity issues; identification of subjects of multiple and intersecting forms of discrimination; joined-up analysis of the drivers, root causes and underlying determinants of inequalities and discrimination."[11]

Further, the UN Sustainable Development Cooperation Framework (previously the UN Development Assistance Framework) places the pledge to *leave no one behind* at the core of its four principles for unifying programming and advocacy, requiring all UN entities "to prioritize its programmatic interventions to address the situation of those most marginalized, discriminated against and excluded, and to empower them as active agents of development" [12] The methodology presented here is of thus of direct use for generating discussions on this topic, corresponding to Steps 1, 4 and 5 of the five-step methodology developed by the UNSDG Operational Guide for UN Country Teams, assisting Member States in operationalizing the pledge to Leave No One Behind (LNOB) and reach the furthest behind first.[13]

The goal of this working paper is to propose new methodological tools that will help the UN

[10] The World Bank, Berlin Workshop Series, 2006. Equity and Development. Available online at: https://openknowledge.worldbank.org/handle/10986/6964
[11] Leaving No One Behind: Equality and Non-Discrimination at the Heart of Sustainable Development, The United Nations System Shared Framework for Action: https://www.unsystem.org/CEBPublicFiles/CEB%20equality%20framework-A4-web-rev3.pdf
[12] United Nations Development Assistance Framework Guidance, https://unsdg.un.org/sites/default/files/2017-UNDAF_Guidance_01-May-2017.pdf
[13] Leaving No One Behind: A UNSDG Operational Guide for UN Country Teams (Interim Draft), https://unsdg.un.org/resources/leaving-no-one-behind-unsdg-operational-guide-un-country-teams-interim-draft

system and policymakers in better responding to the call for leaving no one behind.

The following sections (section 4 and 5) describe the data and variables used to undertake the analysis, establishing concrete links with commitments in the 2030 Agenda for Sustainable Development.

Section 6 describes the new methodological tools in detail. Sections 6.1 and 6.2 present a methodology for measuring and understanding inequality of opportunity: the Dissimilarity Index (D-index). The D-index is a simple formula that determines inequality in accessing key services and opportunities. Governments can use this information to identify which opportunities are particularly unequally distributed. Section 6.3 presents an innovative and visually appealing method to identify those furthest behind in each of these services or opportunities, taking into account intersectionality and multiple layers of disadvantage. Through the classification tree analysis (CTA), researchers and policymakers can explore the circumstances shared by those most disadvantaged and the most advantaged groups in each country.

4. The data

In practice, leaving no one behind means moving beyond assessing average and aggregate progress, towards ensuring progress for all population groups at a disaggregated level. This requires disaggregating data to identify groups being excluded or discriminated against, as well those experiencing multiple and intersecting forms of discrimination and inequalities.

The methodological tools presented in this paper (Dissimilarity Index in sections 7.1 and 7.2 and Classification Tree Analysis in section 7.3) use the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS). DHS and MICS are publicly available for 26 Asian and Pacific countries. The datasets are selected because of a) comparability across countries; b) accessibility of the data; and c) the rich set of questions on health, demographic and basic socioeconomic data that refer both to the household (e.g. water and sanitation, financial inclusion, electricity and clean fuels, as well as ownership of mobile phones/bank cards) and to individuals (e.g. level of education, nutrition status, access to basic healthcare services for children, household head, other household members.) 17 countries have surveys representing two different points in time. The full list of 26 countries and survey years (latest and earliest) is provided in Table 1.

Despite their many advantages, DHS and MICS also have shortcomings. For example, because some questions are answered at the household level, they do not allow for calculation of sex-disaggregated data at the household level. Furthermore, men are not always asked the same sets of questions as women. Lastly, the surveys do not capture people least likely to be counted and reflected in national statistics, such as the homeless, slum dwellers, irregular migrants, nomadic or displaced populations, stateless persons, criminalized populations (e.g., people who use drugs, sex workers) and people in temporary shelters or institutions.

Table 1: List of countries and survey years

	Country	Earliest Year	Earliest Survey	Latest Year	Latest Survey
1	Afghanistan	2010	MICS	2015	DHS
2	Armenia	2000	DHS	2016	DHS
3	Azerbaijan	2000	MICS	2006	DHS
4	Bangladesh	2000	DHS	2014	DHS
5	Bhutan	n/a	n/a	2010	MICS
6	Cambodia	2000	DHS	2014	DHS
7	India	2006	DHS	2016	DHS
8	Indonesia	2000	MICS	2017	DHS
9	Kazakhstan	2006	MICS	2015	MICS
10	Kyrgyzstan	1997	DHS	2018	MICS
11	Lao People's Democratic Republic	2000	MICS	2018	MICS
12	Maldives	n/a	n/a	2017	DHS
13	Mongolia	2000	MICS	2018	MICS
14	Myanmar	2000	MICS	2016	MICS
15	Nepal	2001	DHS	2016	DHS
16	Pakistan	1991	DHS	2018	DHS
17	Papua New Guinea	n/a	n/a	2018	DHS
18	Philippines	1998	DHS	2017	DHS
19	Tajikistan	2000	MICS	2017	DHS
20	Thailand	2005	MICS	2016	MICS
21	Turkmenistan	2006	MICS	2016	MICS
22	Timor-Leste	2010	DHS	2016	DHS
23	Turkey	2003	DHS	2013	DHS
24	Uzbekistan	2000	MICS	2006	MICS
25	Vanuatu	n/a	n/a	2007	MICS
26	Viet Nam	2000	MICS	2014	MICS

5. Basic opportunities

5.1. The indicators

The indicators measuring household and individual opportunities have been identified as areas where inequality jeopardizes a person's life prospects. Each of these opportunities are covered by specific commitments outlined in the 2030 Agenda for Sustainable Development.

The indicators used to uncover inequality in access to opportunities can be grouped into two groups, core and auxiliary indicators. Core indicators include 12 opportunities covering secondary and higher education, children's nutrition (stunting, overweight and wasting), women's health (professional help during childbirth and use of modern contraceptive methods), basic drinking water, basic sanitation facilities, electricity, clean fuels, and bank account. Auxiliary indicators can be constructed using core ones, being either their inverse values or a combination of more than two core indicators.

As reported by the Interagency Group on SDG Indicators (IAEG-SDGs), the connection between these indicators and the Sustainable Development Goals (SDGs) was the main criterion for their selection.[14] (Table 2)

Table 2: The links between opportunities and the SDGs

Opportunities / Indicators			Closest SDG indicator reference
Core opportunities	1	Secondary education	4.1.1 Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex
	2	Higher education	4.3.1 Participation rate of youth and adults in formal and non-formal education and training in the previous 12 months, by sex
	3	Stunting	2.2.1 Prevalence of stunting (height for age <-2 standard deviation from the median of the World Health Organization (WHO) Child Growth Standards) among children under 5 years of age
	4	Overweight	2.2.2 Prevalence of malnutrition (weight for height >+2 or <-2 standard deviation from the median of the WHO Child Growth Standards) among children under 5 years of age, by type (wasting and overweight)
	5	Wasting	2.2.2 Prevalence of malnutrition (weight for height >+2 or <-2 standard deviation from the median of the WHO Child Growth Standards) among children under 5 years of age, by type (wasting and overweight)
	6	Use of modern contraceptive	3.7.1 Proportion of women aged 15-49 years who have their need for family planning satisfied with modern methods
	7	Professional help in childbirth	3.1.2 Proportion of births attended by skilled health personnel
	8	Basic drinking water	6.1.1 Proportion of population using safely managed* drinking water services
	9	Basic sanitation services	6.2.1 Proportion of population using safely managed* sanitation services, including a hand-washing facility with soap and water

[14] The latest indicators to be used for monitoring the SDGs can be found at: https://unstats.un.org/sdgs/iaeg-sdgs/.

Opportunities / Indicators			Closest SDG indicator reference
	10	Electricity	7.1.1 Proportion of population with access to electricity
	11	Clean fuels	7.1.2 Proportion of population with primary reliance on clean fuels and technology
	12	Bank account	8.10.2 Proportion of adults (15 years and older) with an account at a bank or other financial institution or with a mobile money-service provider

The exact questions from DHS and MICS questionnaires with a brief description can be found in Table 3.

Table 3: Survey questions from DHS/MIC for each indicator and their description

Survey Reference				
Indicator	**Variable name**	**Survey Question (in DHS/ MICS)**	**Description**	**Survey Recode**
Secondary education	DHS: HV109 MICS: ED4A, ED4B	What is the highest level of school you attended: primary, secondary, or higher?		PR
Higher education	DHS: HV109 MICS: ED4A, ED4B	What is the highest level of school you attended: primary, secondary, or higher?		PR
Stunting	DHS: HC70 MICS:HAZ	Height in centimeters for children age 0-5	If the height of the child is two standard deviations below the average of children of the same age, he/she is considered stunted	PR
Overweight	DHS: HC71 MICS: WAZ	Weight in kilograms and height in centimeters for children age 0-5	If the ratio of the weight over height of the child is more than two standard deviations of the average of children of the same age, he/she is considered overweight	PR
Wasting	DHS: HC72 MICS: WHZ	Weight in kilograms and height in centimeters for children age 0-5	If the ratio of the weight over height of the child is below two standard deviations of the average of children of the same age, he/she is considered wasted	PR
Use of modern contraceptive	DHS: V313 MICS: CP2, CP3A, CP3B, CP3C, CP3D, CP3E, CP3F, CP3G, CP3H, CP3I, CP3J, CP3K	What are you doing/which method are you using to delay or avoid a pregnancy?	Modern contraceptive methods include pills, UID, foam, condom, etc.	IR
Professional help in childbirth	DHSM3A,M3B,M3C,M3D,M3E,M3F MICS: MN17A, MN17B, MN17C, MN17D, MN17E, MN17I, MN17J, MN17K	Who assisted with the delivery of (name)?	Professional help includes doctor, nurse, and midwife	IR
Basic drinking water	DHS: HV201 MICS (4&5): WS1	What is the main source of drinking water for members of your household?	Population using improved drinking water sources such as piped household water connection, public standpipe, borehole, protected dug well, protected spring, rainwater collection. (*instead of "safely managed", using the "basic services" definition so as to cover more countries)	HH

Survey Reference				
Indicator	Variable name	Survey Question (in DHS/ MICS)	Description	Survey Recode
Basic sanitation services	DHS: HV205, HV225 MICS (4&5): WS8, WS9	What kind of toilet facility do members of your household usually use?	Population using improved, non-shared sanitation facilities such as those with: sewer connections, septic system connections, pour-flush latrines, ventilated improved pit latrines, pit latrines with a slab or covered pit (*instead of "safely managed", using the "basic services" definition so as to cover more countries)	HH
Electricity	DHS: HV206 MICS: HC8A	Does your household have electricity?		HH
Clean fuel	DHS: HV226 MICS: HC6	What type of fuel/energy does your household mainly use for cooking?	Clean fuel includes natural fuel (e.g. compressed natural gas or liquified petroleum gas) or a blend (e.g. gasohol) used as a substitute for fossil fuels and which produces less pollution than the alternatives	HH
Bank Account	DHS: HV247 MICS: HC15	Does any member of this household have a bank account?		HH

Note: The variable names here are for the version 5 and 6 of MICS database.

5.2. The determinant factors (circumstances)

Inequality of opportunity in recent literature has been used to denote the extent of inequality of outcome (income or consumption) that can be attributed to circumstances over which individuals have no control, such as race and sex, as opposed to effort. In literature, this analysis is usually conducted through regression analysis, where the identified circumstances explain a share of the inequality of outcome. The inequality of opportunity resulting from this regression is usually understood to be the lower bound of the total (unobserved) inequality of opportunity since available datasets cannot and do not include all possible circumstances that may impact outcomes. This approach to measuring inequality of opportunity is deemed to be 'indirect'.

Given that the DHS and MICS datasets do not include information on income or consumption (both classified as outcomes), the approach proposed in this paper does not include such regressions. However, future analysis might use the wealth index of the DHS and MICS as a proxy of an 'outcome' and regress it against the set of circumstances used in these reports.

A more 'direct' approach, used here, is to identify a set of 'opportunities' and to measure the gaps among groups in access to these opportunities. To do so, a set of 'circumstances' is selected from available variables in the DHS and MICS datasets to define the groups. These circumstances are usually a set of conditions that the individuals or the households have little control over.

The selection of variables is consistent across all surveys to maintain comparability of inequality across countries. Ultimately, these circumstances (determinant factors) define the composition of the groups. However, circumstances should not be interpreted as 'causes' of inequality. Furthermore, there are many other factors that these models cannot consider, given the limited variables available in the datasets.

Ideally, it would have been preferred to include only circumstances over which a household member had almost no control, such as dominant religion in a household where a respondent is born, ethnicity, existence of a disability, education of the mother or father of the respondent. The

majority of the DHS did not include these questions. Some MICS, however, did ask questions related to ethnicity, language and religion.[15] In the cases where these questions were included, the analysis can be repeated using these additional determinant factors. Additional potentially useful factors that could have been of interest for the study are geographical variables, such as province or region of a given country, but that would have affected comparability across countries. These geographic variables are analysed in the work that focuses on one country only.

Table 4: Circumstances used to determine groups, per indicator

Opportunities			Circumstances used to determine groups of the furthest behind/ ahead					
No.	Indicators	Reference group in survey	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1	Secondary education	Household member age 20-35	Wealth	Residence	n/a	Woman/ Man	n/a	n/a
2	Higher education	Household member age 25-35	Wealth	Residence	n/a	Woman/ Man	n/a	n/a
3	Stunting	Child aged 0-5 who has been measured	Wealth	Residence	Mother's Education	Boy/ Girl	Number of children <5	n/a
4	Overweight	Child aged 0-5 who has been measured	Wealth	Residence	Mother's Education	Boy/ Girl	Number of children <5	n/a
5	Wasting	Child aged 0-5 who has been measured	Wealth	Residence	Mother's Education	Boy/ Girl	Number of children <5	n/a
6	Use of modern contraceptive	Women between 15-49 currently in union	Wealth	Residence	Responder's education	n/a	Number of children <5	Age: 15-24, 25-49
7	Professional help in childbirth	Women between 15-49 ever given birth in the last 5 years	Wealth	Residence	Responder's education		Number of children <5	Age: 15-24, 25-49

[15] Newer MICS and DHS also have started including more questions on disability, migration status, etc.

Opportunities			Circumstances used to determine groups of the furthest behind/ ahead					
No.	Indicators	Reference group in survey	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
8	Basic drinking water	All households	Wealth	Residence	Highest Education in household (hh)	n/a	n/a	n/a
9	Basic sanitation services	All households	Wealth	Residence	Highest Education in hh	n/a	n/a	n/a
10	Electricity	All households	Wealth	Residence	Highest Education in hh	n/a	n/a	n/a
11	Clean fuels	All households	Wealth	Residence	Highest Education in hh	n/a	n/a	n/a
12	Bank account	All households	Wealth	Residence	Highest Education in hh	n/a	n/a	n/a

6. A set of practical tools

6.1. Dissimilarity Index (D-Index)

In order to assess how countries are faring in terms of inequality of opportunity, the analysis uses the Dissimilarity Index, or the D-index, a measure of how unequally access to an opportunity is distributed among population groups.

The Dissimilarity Index (D-Index) measures how different groups - such as women, poorer households, or ethnic minorities - fare in terms of access to a certain opportunity. Like the Gini coefficient, the D-Index ranges from 0 to 1, where 0 indicates no inequality, and 1 that the entire access to a service is reserved to a specific group of people with shared circumstances (e.g. men from urban areas). For example, two countries that have identical average access rates may have a different D-Index if the distribution of the opportunity in one country excludes certain groups (such as females, poorer groups, or ethnic minorities).

To obtain the D-Index, inequalities in access among all possible groups generated by the tree classification are calculated using the formula (1) below:

$$D = \frac{1}{2\bar{p}} \sum_{i=1}^{n} \beta_i \, |p_i - \bar{p}|$$

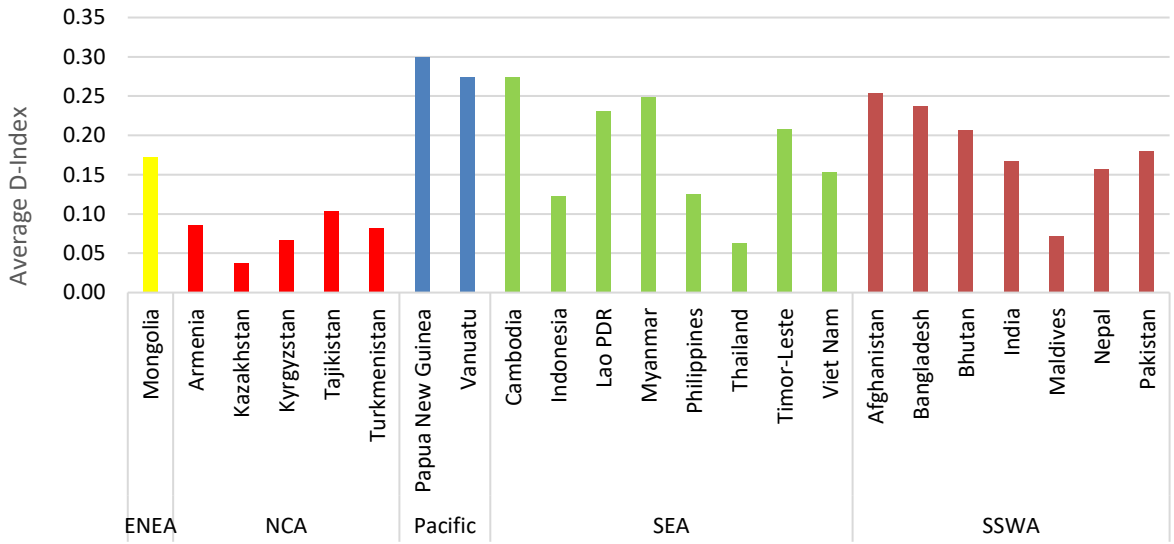
(1)

where β_i is the weighted sampling proportion of the group i, (sum of β_i equals 1), $\bar{p}$ is the average access rate in the country and p_i is the level of access of population group i, and takes values from 0 to 1. There are n groups, which are defined using the interactions of the circumstances selected

for the analysis.

In the case of an opportunity where three circumstances were considered: wealth (2 groups), residence (2 groups) and education (4 groups), covering the entire sample population, this calculation produces 16 groups (2x2x4). The interactions between these groups form each population group i and are used to calculate the D-index. This index is therefore a weighted average of the absolute difference of most and least advantaged population groups from the average access rate in the country ($\bar{p}$).

Figure 1: Average D-indices in Asia-Pacific countries, grouped by subregion



Source: ESCAP calculations using data from the latest DHS and MICS surveys for countries in Asia-Pacific. Note: ESCAP subregions are: East and North-East Asia (ENEA), North and Central Asia (NCA), Pacific, South-East Asia (SEA), and South and South-West Asia (SSWA).

Averaging the D-indices for individuals and households by country highlights which countries have relatively higher inequality across all opportunities (Figure 1). The highest D-indices are found in the Pacific, followed closely by countries in South-East Asia and South and South-West Asia. Afghanistan, Cambodia, Myanmar, Papua New Guinea and Vanuatu appear as particularly unequal across the board of opportunities, with average D-indices above 0.25.At the other end of the scale, Maldives and Thailand, together with several North and Central Asian countries seem to have achieved a relatively equal distribution of opportunities across various population groups for most opportunities. The subregion of North and Central-Asia stands out as the most equal in terms of access to opportunities, thanks to a tradition of a large state that ensures universal provision of basic services. In the middle of the distribution are some of the region's most rapidly developing countries, including India, Indonesia, Philippines and Viet Nam.

6.2. Shapley decomposition

Building on the calculation of the D-Index, it is also possible to estimate the contribution of each of the circumstances to inequality or dissimilarity. The results of this analysis can reveal possible drivers of inequality – and even rank those factors by importance of their contribution. From a

policymaking perspective, knowing the most important contributors to inequality can hint to which inequality gaps countries could tackle first in their efforts to 'leave no one behind'.

The decomposition method used in this analysis, called Shapley decomposition, estimates the marginal contribution of each circumstance to inequality in access to an opportunity. The basic idea behind this decomposition, taken from cooperative game theory, is to measure how much the estimated D index would change when a circumstance was added to the pre-existing set of circumstances. The change in inequality because of the addition of a new circumstance would be a reasonable indicator of its "contribution" to inequality. [16]

The impact of adding a circumstance A is given by the following formula:

$$D_A = \sum_{S \subseteq N\{A\}} \frac{|S|!(n-|S|-1)!}{n!} [D(S \cup \{A\}) - D(S)]$$

where N is the set of all the n circumstances; and S is the subset of N circumstances obtained after omitting the circumstance A. D(S) is the dissimilarity index estimated with the set of circumstances S. D(SU{A}) is the dissimilarity index calculated with set of circumstances S and the circumstance A.

The contribution of characteristic A to the dissimilarity index is then formula:

$$M_A = \frac{D_A}{D(N)}$$

The critical property satisfied by the Shapley decomposition is that the sum of contributions of all characteristics adds up to 1 (100%).

6.3. The Classification Tree methodology

Knowing that inequality of opportunity is broadly associated with specific circumstances opens the door to deeper exploration of the data to see exactly which groups the most marginalized and which groups are have benefitted most from development. Identifying these groups could help policymakers better focus policy and programmes to tackle inequality of opportunity.

The primary goal of using the classification tree analysis is therefore to identify the groups with the lowest and highest access to the opportunities, using the selected indicators. The indicators used are the "response variables", while the factors that characterize these groups are defined as "circumstances". A classification tree is an analytical structure that represents groups of the sample population that have significantly different response values, or different levels of access to a certain opportunity.

To identify the groups with the highest or lowest access to a specific opportunity, a classification tree is constructed for each country, using R, an open source statistical software. The root node of the tree is the entire population sample. The tree method algorithm starts by searching for the first split (or branch) of the tree. It does so by looking at each circumstance and separating the sample in two groups, so that it achieves the most "information reduction". This information metric can be defined in a few ways, while the most common one – and the one used in this

[16] Shorrocks, A. F. (2013). Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. Journal of Economic Inequality, March 2013, Volume 11, Issue 1, pp 99–126.

analysis is the "entropy".[17]

For example, the algorithm estimates access to professional help during childbirth by partitioning the household into different groups based on the household circumstances chosen. The formula that represents the core of the algorithm is the following:

$$p(Y_i = 1 || X_{1i}, X_{2i}, ..., X_{li}) = \sum_{j=1}^{m} p_j \times I((X_{1i}, X_{i2}, ...X_{li}) \in A_j)$$

where Yi is the observed opportunity for the i-th household in the sample, X1i, ..., Xli are the circumstances for the household. In the example of access to professional help during childbirth, Y is the rate of professional help during childbirth, X1, X2, X3 (where l = 4) are residence, household wealth level, mother's education and number of children in the household. A1, A2, ... Am are the different partitions of the sample, also called end nodes, where:

$$A_i \cap A_j = \emptyset$$

and

$$\cup_{i=1}^{m} A_i = \Omega.$$

This means the end nodes are mutually exclusive and complementary, and every woman who has recently given birth belongs to one and only one of the end nodes. I () only takes value 1 when the i-th household belongs to j-th end node, otherwise, I () takes value 0. The tree algorithm generates the end nodes, according to metrics that measure the effectiveness of the partition that gives to different levels of access to clean energy.

Information theory and entropy is a very common choice for the metrics. Entropy for j-th end node can be calculated according to the definition:

$$I_E(p_j) = -(p_j \times log_2 p_j + (1 - p_j) \times log_2(1 - p_j))$$

The aggregated entropy for the tree is calculated by:

$$H(T) = \sum_{j=1}^{m} q_j \times I_E(p_j)$$

where q_j is the sample proportion of A_j. The actual algorithm that generates the end-nodes is step-by-step, starting from the entire sample. Each time the sample is partitioned new end-nodes are generated and the entropy is calculated and compared to the entropy before the new partition. Each partition (and hence the new end nodes) is kept when the increment of entropy is bigger than a preset threshold. The algorithm stops when no more increment of entropy can be made by new partition, or a set of preset conditions can't be satisfied.

In addition to finding groups that have significant differences in their access to an opportunity, the classification tree algorithm also operates under the limitation that each group should have enough group members. To avoid a too small sub-sample size, the analysis has set the tree nodes

[17] Kelleher, John D., Brian Mac Namee, Aoife D'Arcy. Fundamentals of Machine Learning for Predictive Data.

to have a minimum size of at least 10 per cent of the total population and the split of tree is only made when an "information reduction" criterion is satisfied.

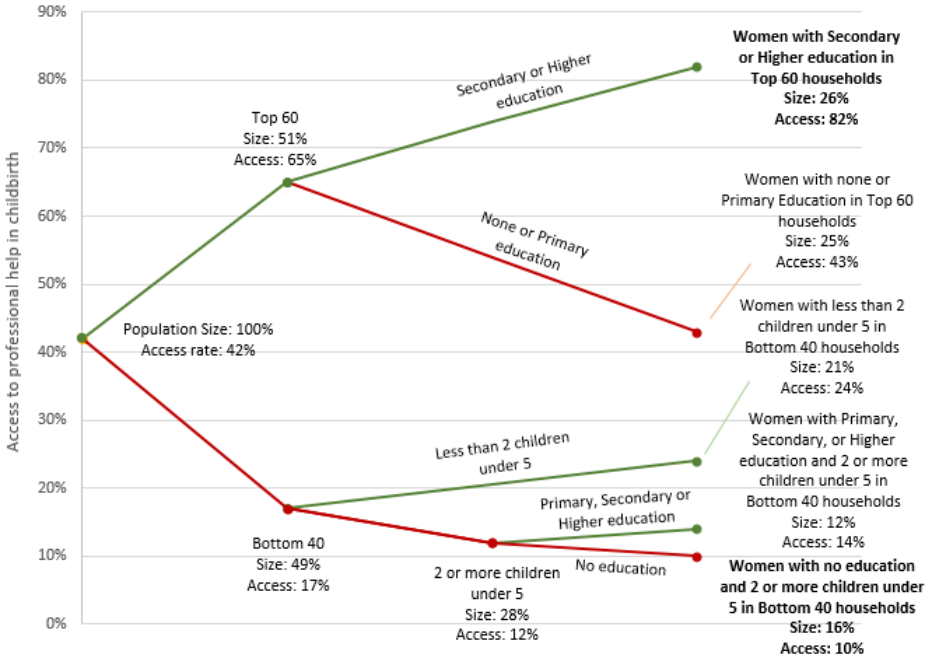
An example: To illustrate how the classification tree identifies the most disadvantaged or advantaged groups, the example of access to professional help during childbirth in Lao People's Democratic Republic is used.

Indicator ('response variable'): Access to professional help during childbirth in Lao People's Democratic Republic

Factors ('circumstances): The circumstances being considered are the following:

1. Household wealth (Bottom 40 or Top 60),
2. Education (Secondary or Higher vs. Primary or None)
3. Number of children aged less than 5
4. Residence (Rural or Urban)

Figure 2: Classification tree highlighting differences in women's access to professional help in childbirth in Lao People's Democratic Republic, 2011 (women 15-49 years of age)



Source: ESCAP calculations, using data from the latest DHS and MICS surveys for countries in Asia-Pacific.

The classification tree starts at the average access rate of 42 per cent. The algorithm determines that the first split into branches should be wealth, specifically where in the wealth distribution a woman belongs: the top 60 per cent or the bottom 40 per cent. Women belonging to the top 60 per cent group have 65 per cent access rate to professional help in childbirth, compared with only 17 per cent for those in the bottom 40 group.

In the same example, the algorithm determines a second split for the less advantaged (bottom 40 group) around the number of children a woman has had. For their first childbirth, one in four

women in the bottom 40 group uses professional help. That rate falls to one in nine for subsequent childbirths or for women with more than one child. The rate of access to professional help also varies for women with more than one child: only one in ten women with no education get professional help, while one in eight of those with completed primary, secondary or higher education do. Among the women belonging to the top 60 group, the only further split is based on education. Half of the women with primary or no education access professional help, compared with eight out of ten of those with secondary or higher education.

The group with the highest access to professional help in childbirth is women with secondary or higher education in households belonging to the top 60 of the wealth distribution. They have an access rate of 82 per cent and represent 26 per cent of Laotian women in union who have given birth in the past five years. Conversely, only one in ten women in the bottom 40 group with no education and two or more children under 5 years of age use professional help during childbirth. The total gap between the groups with the highest and the lowest access is a staggering 72 percentage points. Living in an urban or rural residence does not appear to influence level of access.

The uniqueness of the classification tree approach is that it becomes very clear where policies should, or should not, be focused to reach those furthest behind.

7. Gaps and limitations

The methodologies presented in this short paper have several advantages, but also some limitations:

Firstly, the available datasets (DHS and MICS) limit the scope of the analysis to only those indicators for which data are collected. In reality, there are many variables shaping access to different opportunities. For example, the quality and reliability of a water connection is an important factor that might affect the access to basic drinking water. Similarly, distance from a health-care provider is an important circumstance that might shape a women's access to skilled birth attendance. These variables are not measured in existent DHS and MICS surveys, so results have to be understood with this caveat.

Consistent with similar studies on inequalities among groups, this analysis also does not consider inequality within groups. Even with homogeneous groups (e.g. women from poorer households and with lower education), additional unobserved circumstances affect outcomes. This analysis only calculates the observable average access to an opportunity for each group, and thus draw conclusions on gaps and inequality based in these averages.

For assessing completion of secondary and higher education, the sample has been restricted to those 20-35 (for secondary education) and 25-35 (for higher education). The reason is to avoid: (1) skewing the results because of an older population with significantly lower education levels; and (2) including individuals that, because of their young age, could not have completed their education.

An important limitation is the lack of information on income of individuals or households, as it is not collected by DHS and MICS. Instead, the analysis uses the wealth index, a composite index

reflecting a household's cumulative living standard, developed by the DHS and MICS researchers. The wealth index combines a range of household circumstances including: a) ownership of household assets, such as TVs, radios and bicycles; b) materials used for housing; and c) type of water and sanitation facilities. The wealth index is calculated using the Principal Component Analysis and thus allows a relative ranking of households based on their assets.[18] The wealth index is not comparable across countries, as it may consist of different assets in each country. As a result, any cross-country comparison of household access based on "wealth" should be understood with that caveat.

Finally, the results are limited by available indicators. The classification tree analysis only presents circumstances in the tree branches if they are found to reduce "entropy". Ultimately, these circumstances define the composition of the groups, but should not be interpreted as "causes" of a lower access. There are also many other factors that could potentially impact the analysis, but because of the limitation of the datasets, have not been included.

Annex

To bolster the analytical findings, logistic regressions were conducted to observe the effects of circumstance variables (household wealth, residence, sex, etc.) on a specific indicator, for example individual's secondary or higher education attainment.

The logistic regression model for each country is given by:

$$logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_{2+} \beta_3 x_3$$

Where p_i stands for P(y=1), and y is a binary response variable which assumes two values:

$$y = \begin{cases} 1, & if\ the\ individual\ has\ access\ to\ a\ specific\ opportunity \\ 0, & if\ the\ individual\ does\ not\ have\ access\ to\ a\ specific\ opportunity \end{cases}$$

and

where β0..n are logit model coefficients and X1 ..n are circumstance variables, i.e. X1 is household wealth, X2 is residence, and X3 is the sex of the individual, either male or female and so forth.

[18] Wealth Index Construction, https://www.dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm

About the Economic and Social Commission for Asia and the Pacific (ESCAP)

The Economic and Social Commission for Asia and the Pacific (ESCAP) serves as the United Nations' regional hub promoting cooperation among countries to achieve inclusive and sustainable development. The largest intergovernamental platform with 53 member States and 9 associate members, ESCAP has emerged as a strong regional think-tank offering countries sound analytical products that shed insight into the evolving economic, social and enviromental dynamics of the region. The Commission's strategic focus is to deliver on the 2030 Agenda for Sustainable Development, which it does by reinforcing and deepening regional cooperation and integration to advance connectivity, financial cooperation and market integration. ESCAP's research and analysis coupled with its policy advisory services, capacity building and technical assistance to governments aim to support countries' sustainable and inclusive development ambitions.

Social Development Division
United Nations Economic and Social Commission for Asia and the Pacific
United Nations Building
Rajadamnern Nok Avenue
Bangkok 10200, Thailand
Email: escap-sdd@un.org
Website: www.unescap.org

UNITED NATIONS
ESCAP
Economic and Social Commission for Asia and the Pacific