SUSTAINABLE DEVELOPMENT GOALS

THE INTER-SECRETARIAT WORKING GROUP ON HOUSEHOLD SURVEYS

Building national capacity on small area estimation

Haoyi Chen, Coordinator

Inter-Secretariat Working Group on Household Surveys

JPSM Session on Small Area Estimation, 15 March 2024

unstats.un.org/iswghs

Outline

❑ Inter-Secretariat Working Group on Household Surveys (ISWGHS) & IAEG-SDGs

❑ Toolkit on using SAE for SDG indicators

❑ Capacity building activities on SAE: challenges and opportunities

❑ Next steps

ISWGHS

The ISWGHS: a primer

- □ Established in 2015 under the aegis of the UNSC
- □ Objectives:
    - □ Improve coordination of household surveys
    - □ Advance cross-cutting survey methodology
    - □ Enhance communication and advocacy
- □ Governance
    - Membership: 11 international agencies + 10 (rotating) member states
    - Secretariat: UN Statistics Division
    - Current co-chairs: WB and UNW
- □ Work through time-bound Task Forces, led by and with contribution from members and non-member experts.

ISWGHS

Inter-agency and Expert Group on Sustainable Development Goal Indicators (IAEG-SDGs)

1 NO POVERTY
2 ZERO HUNGER
3 GOOD HEALTH AND WELL-BEING
4 QUALITY EDUCATION
5 GENDER EQUALITY
6 CLEAN WATER AND SANITATION
7 AFFORDABLE AND CLEAN ENERGY
8 DECENT WORK AND ECONOMIC GROWTH
9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
10 REDUCED INEQUALITIES
11 SUSTAINABLE CITIES AND COMMUNITIES
12 RESPONSIBLE CONSUMPTION AND PRODUCTION
13 CLIMATE ACTION
14 LIFE BELOW WATER
15 LIFE ON LAND
16 PEACE, JUSTICE AND STRONG INSTITUTIONS
17 PARTNERSHIPS FOR THE GOALS
THE GLOBAL GOALS

The 2030 Agenda for Sustainable Development

- □ A global blueprint for people, planet, prosperity , peace and partnerships, now and in the future
- □ 17 Goals, 169 targets and "Leaving no one behind" principle

The IAEG-SDGs :

- □ Composed of 28 Member States (and representatives of regional commissions, regional and international agencies and CSOs are observers)
- □ Developed the global indicator framework for SDGs (**231 indicators)**

OLDER PERSONS
PERSONS WITH DISABILITIES
CHILDREN
WOMEN
MIGRANTS AND REFUGEES

IAEG-SDGs workstream on data disaggregation:

- □ Compilation of existing guidelines and methodologies on data disaggregation
- □ Preparation of Handbook on data disaggregation for SDGs
- □ Task Force on Small Area Estimation (joint with ISWGHS)

ISWGHS

Positioning household survey for the next decade

Organized around **8 technical priorities**:

1. Enhancing the interoperability and integration of household surveys
2. Designing and implementing more inclusive, respondent-centric surveys
3. Improving sampling efficiency and coverage
4. Scaling up the use of objective measurement technologies
5. Building capacity for CAPI, phone, web, and mixed-mode surveys
6. Systematizing the collection, storage, and use of paradata and metadata
7. Incorporating machine learning and artificial intelligence for data quality control and analysis
8. Improving data access, discoverability, and dissemination.

Plus:

Foster stronger **enabling environment**:
at national and global level

https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji220042

ISWGHS

The SAE4SDG Toolkit

❑**The Toolkit on Using Small Area Estimation for SDGs** (https://unstats.un.org/wiki/display/SAE4SDG/) in Wiki is a space to provide information on methods to produce disaggregated data through small area estimation.

❑**Goal:** To provide practical tools with accompanying case studies for countries to use SAE for SDG monitoring.

❑**Objectives:**

- Using SAE methods to improve SDG data availability for vulnerable population groups
- Offering practical guidance and country case studies
- Guiding on the enabling environment for using SAE for official data production
- Providing a space for partners to document and disseminate their SAE methodologies

ISWGHS

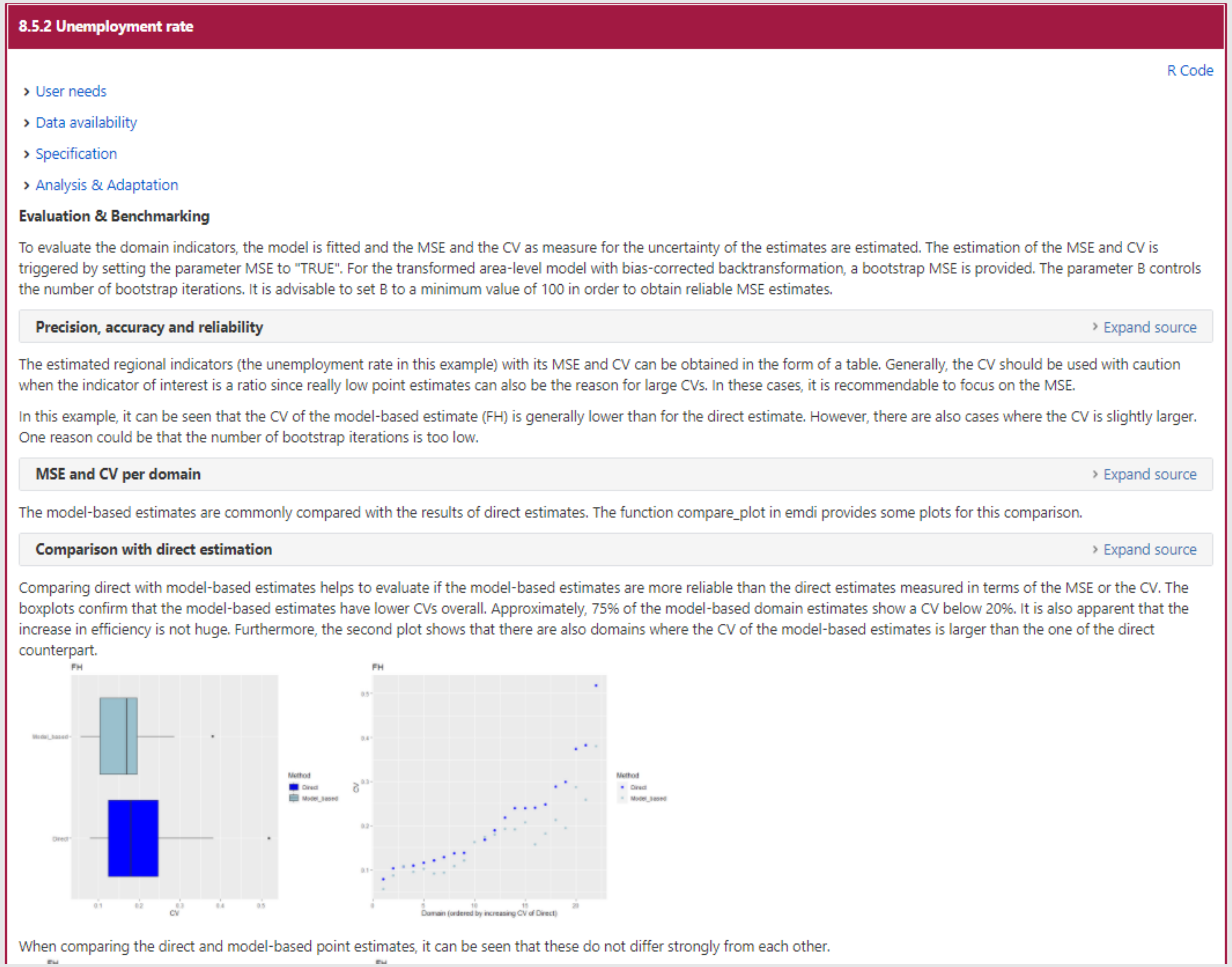
What the SAE Toolkit Offer

❑ Many countries have experimented with SAE in the past but few were able to transform from experiment to official production. The Toolkit:

- Finds out why this is happening?
- Establishes a close link of SAE to SDG monitoring
- Provides hands-on exercise, including “semi-synthetic” data (national data + noises) and programming guide.
- Incorporates national examples and case studies through two angles: (a) documenting the lessons learnt and challenges of countries in using SAE for official data production; and (b) illustrating SAE practices for indicators under different SDG goals.
- Includes main challenges and enabling environment to move from SAE experiment to official production, based on our discussion with national statistical offices.
- Provides an up-to-date and comprehensive list of SAE software packages in major languages (R/Stata/SAS/Python).

ISWGHS

Guiding through steps with practical examples



8.5.2 Unemployment rate

R Code

› User needs

› Data availability

› Specification

› Analysis & Adaptation

Evaluation & Benchmarking

To evaluate the domain indicators, the model is fitted and the MSE and the CV as measure for the uncertainty of the estimates are estimated. The estimation of the MSE and CV is triggered by setting the parameter MSE to "TRUE". For the transformed area-level model with bias-corrected backtransformation, a bootstrap MSE is provided. The parameter B controls the number of bootstrap iterations. It is advisable to set B to a minimum value of 100 in order to obtain reliable MSE estimates.

Precision, accuracy and reliability › Expand source

The estimated regional indicators (the unemployment rate in this example) with its MSE and CV can be obtained in the form of a table. Generally, the CV should be used with caution when the indicator of interest is a ratio since really low point estimates can also be the reason for large CVs. In these cases, it is recommendable to focus on the MSE.

In this example, it can be seen that the CV of the model-based estimate (FH) is generally lower than for the direct estimate. However, there are also cases where the CV is slightly larger. One reason could be that the number of bootstrap iterations is too low.

MSE and CV per domain › Expand source

The model-based estimates are commonly compared with the results of direct estimates. The function compare_plot in emdi provides some plots for this comparison.

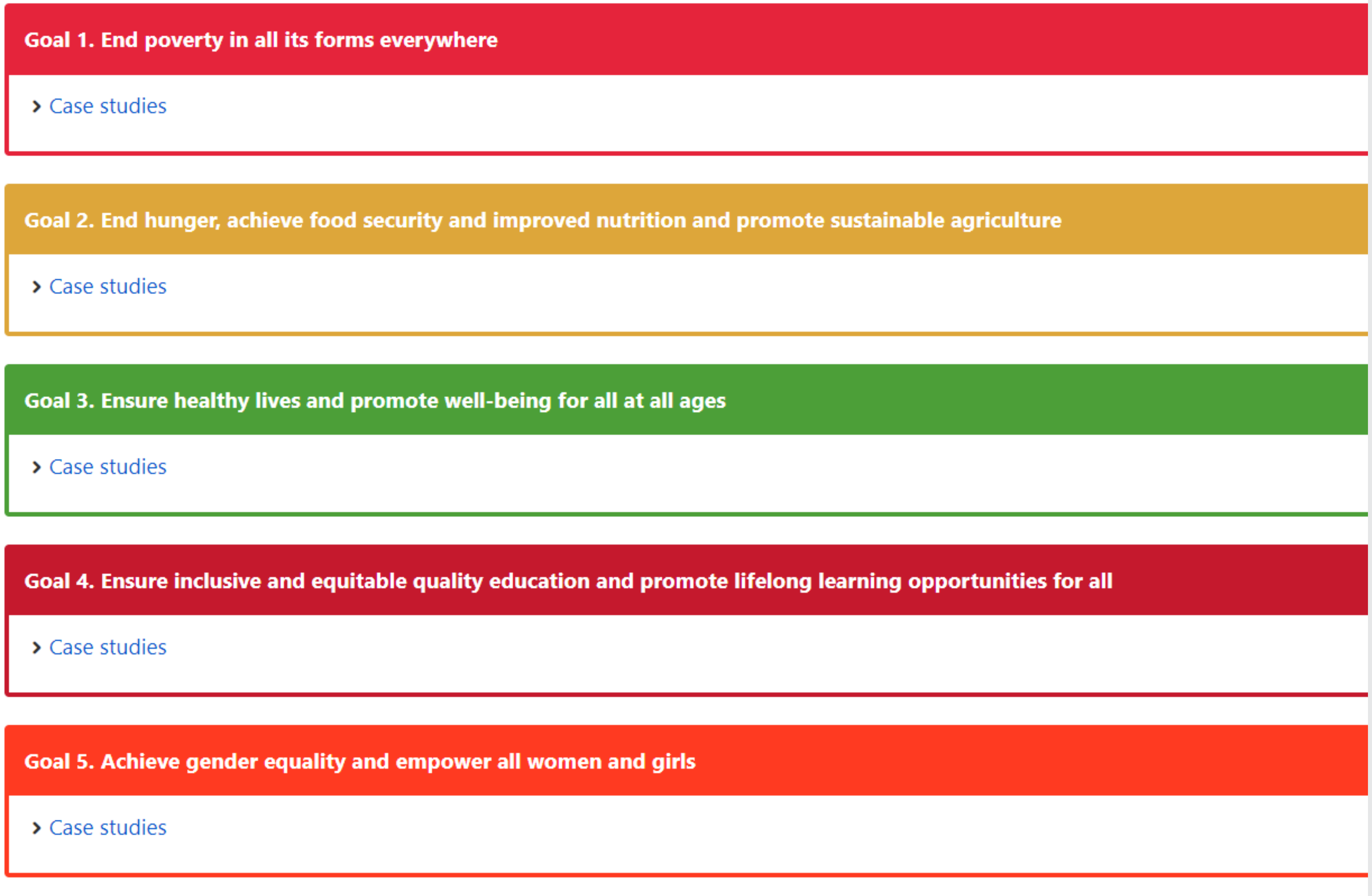
Comparison with direct estimation › Expand source

Comparing direct with model-based estimates helps to evaluate if the model-based estimates are more reliable than the direct estimates measured in terms of the MSE or the CV. The boxplots confirm that the model-based estimates have lower CVs overall. Approximately, 75% of the model-based domain estimates show a CV below 20%. It is also apparent that the increase in efficiency is not huge. Furthermore, the second plot shows that there are also domains where the CV of the model-based estimates is larger than the one of the direct counterpart.

When comparing the direct and model-based point estimates, it can be seen that these do not differ strongly from each other.

ISWGHS

Case studies covering different SDG goals/indicators



Goal 1. End poverty in all its forms everywhere

› Case studies

Goal 2. End hunger, achieve food security and improved nutrition and promote sustainable agriculture

› Case studies

Goal 3. Ensure healthy lives and promote well-being for all at all ages

› Case studies

Goal 4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all

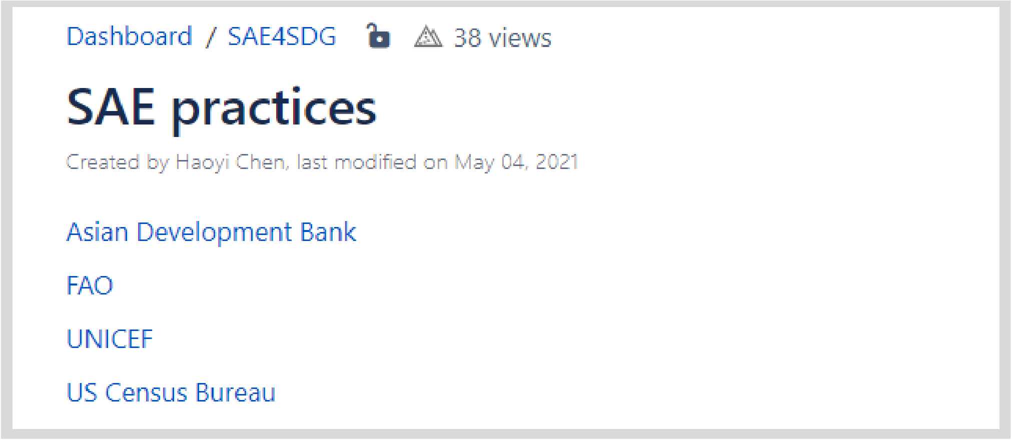
› Case studies

Goal 5. Achieve gender equality and empower all women and girls

› Case studies

ISWGHS

SAE methodologies used by countries and international agencies



Dashboard / SAE4SDG 38 views

SAE practices

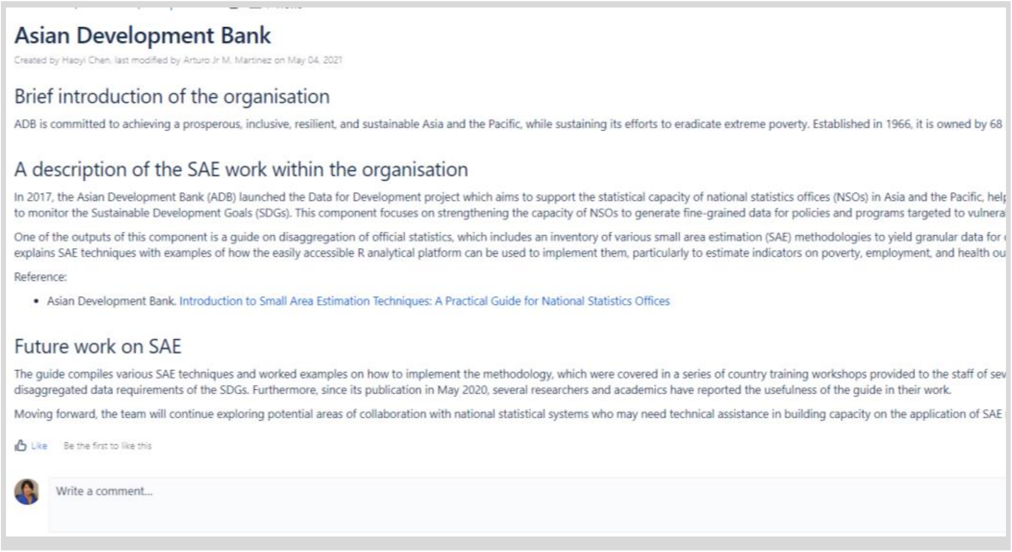
Created by Haoyi Chen, last modified on May 04, 2021

Asian Development Bank

FAO

UNICEF

US Census Bureau



Asian Development Bank

Created by Haoyi Chen, last modified by Arturo Jr M. Martinez on May 04, 2021

Brief introduction of the organisation

ADB is committed to achieving a prosperous, inclusive, resilient, and sustainable Asia and the Pacific, while sustaining its efforts to eradicate extreme poverty. Established in 1966, it is owned by 68

A description of the SAE work within the organisation

In 2017, the Asian Development Bank (ADB) launched the Data for Development project which aims to support the statistical capacity of national statistics offices (NSOs) in Asia and the Pacific, help
to monitor the Sustainable Development Goals (SDGs). This component focuses on strengthening the capacity of NSOs to generate fine-grained data for policies and programs targeted to vulnera

One of the outputs of this component is a guide on disaggregation of official statistics, which includes an inventory of various small area estimation (SAE) methodologies to yield granular data for
explains SAE techniques with examples of how the easily accessible R analytical platform can be used to implement them, particularly to estimate indicators on poverty, employment, and health ou

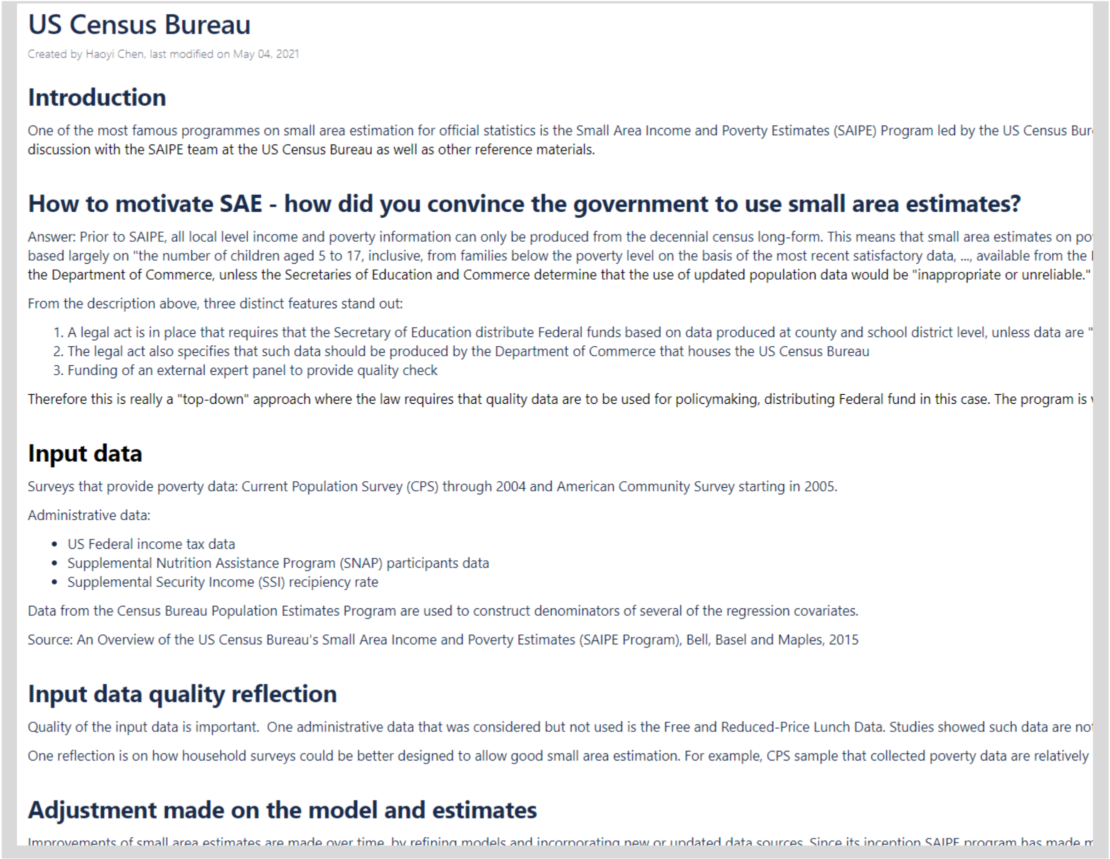
Reference:

- Asian Development Bank. Introduction to Small Area Estimation Techniques: A Practical Guide for National Statistics Offices

Future work on SAE

The guide compiles various SAE techniques and worked examples on how to implement the methodology, which were covered in a series of country training workshops provided to the staff of sev
disaggregated data requirements of the SDGs. Furthermore, since its publication in May 2020, several researchers and academics have reported the usefulness of the guide in their work.

Moving forward, the team will continue exploring potential areas of collaboration with national statistical systems who may need technical assistance in building capacity on the application of SAE



US Census Bureau

Created by Haoyi Chen, last modified on May 04, 2021

Introduction

One of the most famous programmes on small area estimation for official statistics is the Small Area Income and Poverty Estimates (SAIPE) Program led by the US Census Bur
discussion with the SAIPE team at the US Census Bureau as well as other reference materials.

How to motivate SAE - how did you convince the government to use small area estimates?

Answer: Prior to SAIPE, all local level income and poverty information can only be produced from the decennial census long-form. This means that small area estimates on po
based largely on "the number of children aged 5 to 17, inclusive, from families below the poverty level on the basis of the most recent satisfactory data, ..., available from the
the Department of Commerce, unless the Secretaries of Education and Commerce determine that the use of updated population data would be "inappropriate or unreliable."

From the description above, three distinct features stand out:

1. A legal act is in place that requires that the Secretary of Education distribute Federal funds based on data produced at county and school district level, unless data are "
2. The legal act also specifies that such data should be produced by the Department of Commerce that houses the US Census Bureau
3. Funding of an external expert panel to provide quality check

Therefore this is really a "top-down" approach where the law requires that quality data are to be used for policymaking, distributing Federal fund in this case. The program is

Input data

Surveys that provide poverty data: Current Population Survey (CPS) through 2004 and American Community Survey starting in 2005.

Administrative data:

- US Federal income tax data
- Supplemental Nutrition Assistance Program (SNAP) participants data
- Supplemental Security Income (SSI) recipiency rate

Data from the Census Bureau Population Estimates Program are used to construct denominators of several of the regression covariates.

Source: An Overview of the US Census Bureau's Small Area Income and Poverty Estimates (SAIPE Program), Bell, Basel and Maples, 2015

Input data quality reflection

Quality of the input data is important. One administrative data that was considered but not used is the Free and Reduced-Price Lunch Data. Studies showed such data are no

One reflection is on how household surveys could be better designed to allow good small area estimation. For example, CPS sample that collected poverty data are relatively

Adjustment made on the model and estimates

Improvements of small area estimates are made over time, by refining models and incorporating new or updated data sources. Since its inception SAIPE program has made m

ISWGHS

Challenges in using SAE for official statistics

- Lack of interest and support from the top management
- Lack of dedicated resources for SAE research and implementation
- Lack of in-house technical capacity
- Lack of proper input data (access to/poor quality of admin data source)
- Reluctance about the use of model-based estimates (vs. survey estimates that are design-based/model-assisted)
- Difficulties in communicating the technical aspects to users

ISWGHS

Challenges in using SAE for official statistics (cont.)

- "*We did an experiment using small area estimation method for poverty but the results were not consistent with our own estimates so we did not pursue it again*."
- "*We do not have good input data source for SAE - census data are outdated, and administrative data sources do not have good coverage and lack proper auxiliary variables."*
- *"SAE method is complicated and we are not comfortable with independently developing the method."*
- *"It is very difficult to convince the managers to use model-based estimates."*
- *"Producing SAE requires a lengthy period of looking for input data, finding the right auxiliary variables, testing different models and their assumptions and validating the estimates."*

Source: UNSD conversations with NSOs

Enabling environment for SAE

- *Establishing a clear and focused objective that links SAE to data use for policymaking*
- *Building the legal foundation for using SAE for official data production*
- *Fostering an environment for research and development*
- *Design-based versus model-based estimates: a changing culture in the national statistical offices*
- *Input data for SAE*
- *Maintaining a high and fit-for-purpose quality standard*
- *Collaboration*
- *Capacity building*
- *Transparency in releasing methodology and communicating quality*

ISWGHS

Lessons learnt: driven by needs for key policies and funding decisions

❑ *Colombian National Development Plan 2018-22 made it mandatory to redesign the national monetary transfer programs* (Jóvenes en Acción and Familias en Acción), for population in poverty and in extreme poverty. This needs poverty data at municipal level. (Colombia)

❑ *In 2009, the law of the Fondo Común Municipal (FCM) required the Ministry to provide poverty rate estimates every 2 years for all comunas in the country. Funding to all comunas will be allocated based on such data.* (Chile)

❑ *The 2005-2009 BPS Strategic Plan for Statistical Development defined "the development of an efficient and low-cost methodology, which allows for the creation of small area and local specific statistics data" as one of the main activities to support government decentralization (*Indonesia*)*

❑ *The Cabinet of the Government of Jamaica made a request for the Statistical Institute of Jamaica to use small-area estimation for poverty mapping, to produce poverty data for smaller geographical areas within the country. (*Jamaica*)*

❑ Improving America's Schools Act: "*the number of children aged 5 to 17, inclusive, from families below the poverty level on the basis of the most recent satisfactory data, ..., available from the Department of Commerce"* (US)

ISWGHS

Lessons learnt: access to good quality input data

- ❑ Access to auxiliary data sources (e.g., administrative data), regularly
- ❑ Input data are of good quality:
  - Coverage, accuracy and timeliness
  - Availability of auxiliary variables that have good prediction power for the outcome indicator

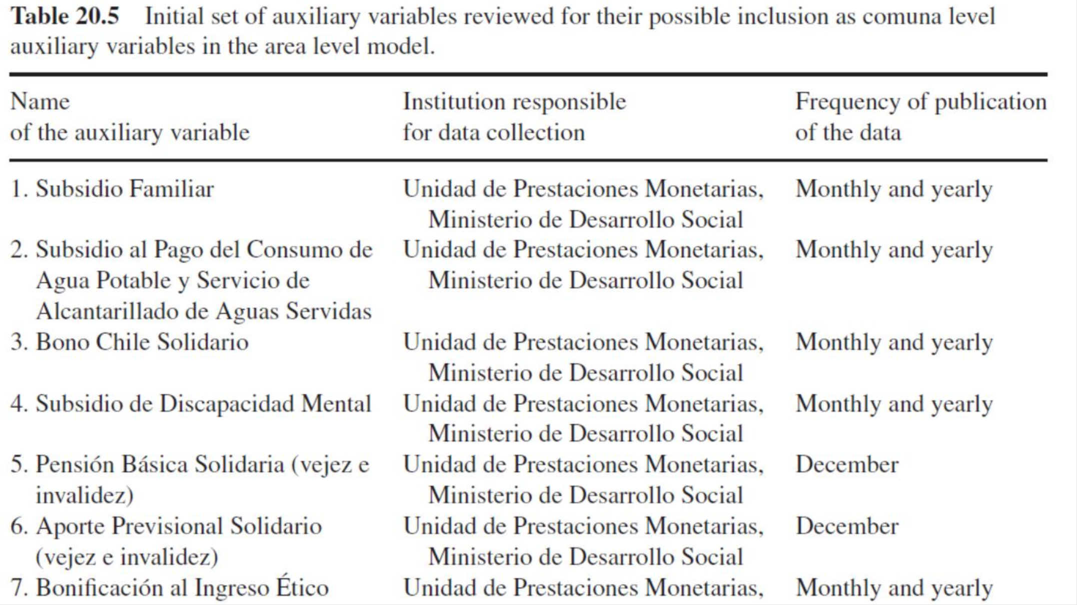


Table 20.5 Initial set of auxiliary variables reviewed for their possible inclusion as comuna level auxiliary variables in the area level model.

Name of the auxiliary variable	Institution responsible for data collection	Frequency of publication of the data
1. Subsidio Familiar	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social	Monthly and yearly
2. Subsidio al Pago del Consumo de Agua Potable y Servicio de Alcantarillado de Aguas Servidas	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social	Monthly and yearly
3. Bono Chile Solidario	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social	Monthly and yearly
4. Subsidio de Discapacidad Mental	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social	Monthly and yearly
5. Pensión Básica Solidaria (vejez e invalidez)	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social	December
6. Aporte Previsional Solidario (vejez e invalidez)	Unidad de Prestaciones Monetarias, Ministerio de Desarrollo Social	December
7. Bonificación al Ingreso Ético	Unidad de Prestaciones Monetarias,	Monthly and yearly

Source: Example from Chile, Casas-Cordero, Encina and Lahiri (2016)

Capacity building on SAE

❑A joint effort of ECLAC-UNSD-UNFPA: https://learning.officialstatistics.org/user/index.php?id=103

- Reading materials
- Recorded videos (50 videos with about 10-15 minutes for each video), organized in 10 modules
- Evaluation materials including weekly computer-graded assessments, two mid-term projects, and a final project
- R program language code that can be used for SAE modelling

❑Opened in August 2023

- Self-paced students on the platform: 460
- Guided learning sessions with an extra 1.5-hour per week to provide guidance: 200 students registered and we are currently supporting around 120 students from Asia, Africa and Latin America (with ECLAC, ESCAP and ECA)

ISWGHS

Offering more and better training

❑High demand: continuing the eLearning course guided training 2024:

- SIAP will be offering one session for Asia and the Pacific
- One for English-speaking African countries and one for Latin America and the Caribbean
- French translation soon to be available, for Francophone African countries (self-paced)

❑Reflecting on the learning experiences: R skills, linear model, busy schedules, sometimes the interested students do not really work on the area, course material very intense

❑Improving the training experiences:

- Reducing the complexity of the project assignments, to cater to different levels of students
- Doing more intensive follow-ups/reminders with students on homework assignments/video watching
- Making certain modules elective for more advanced students
- Preparing Syllabus that has specific grading/marking requirements
- Extending the course completion period by 1-2 more weeks to allow extra time for projects

ISWGHS

Geospatial data for SAE: a review of its potential, limitations and effectiveness

1. An overview of SAE method, why and the audience of the review
2. Input data: geospatial data and training data
3. Geospatial SAE methods
4. Skills and tools to apply the methods
5. Future research and work

❑A draft available: here; will finalise end 2024

❑Partners: World Bank, SAE expert, IAEG-SDGs, GGIM-ISGI

ISWGHS

Geospatial data for SAE: hands-on guidance

- ❑To develop a step-by-step guidance on:
  1. Accessing geospatial data for SAE
  2. Selecting the types of data to use
  3. Illustrating with datasets
- ❑Regional training: Asia/Pacific and Africa

ISWGHS

Thank you

Haoyi Chen
chen9@un.org

ISWGHS