



THE INTER-SECRETARIAT
**WORKING GROUP ON
HOUSEHOLD SURVEYS**

Small area estimation for SDGs: from experiment to production

Haoyi Chen

Coordinator, Inter-Secretariat Working Group
on Household Surveys



unstats.un.org/iswghs

Outline

- ❑ Inter-Secretariat Working Group on Household Surveys (ISWGHS) & IAEG-SDGs
- ❑ ISWGHS and IAEG-SDGs collaboration: Toolkit on using SAE for SDG indicators
- ❑ Capacity building activities on SAE

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 + 8 (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.

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

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**)

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)



Improve national capacity on using SAE for regular production

SAE Toolkit for SDGs, UNSD working with experts on SAE, under the guidance of the IAEG-SDGs and ISWGHS

- Using SAE methods to improve SDG data availability for vulnerable population groups – requested by IAEG-SDGs
- Offering practical guidance and country case studies
- Providing a space for partners to document and disseminate their SAE methodologies: transparency
- Hosting on Wiki-platform: updating & collaborating

ISWGHS-IAEG-SDGs collaboration: work modality

UN Statistics Wiki Spaces People Analytics Cockpit Create ... Search

SAE4SDG ☆

Dashboard 440 views Edit Save for later Watching Share ...

SAE4SDG

Created by UNSD Clarence Lio, last modified by Haoyi Chen on Mar 03, 2021

Welcome to the Toolkit for using Small Area Estimation for the SDGs!

In committing to the realization of the 2030 Agenda for Sustainable Development, Member States recognized that the dignity of the individuals is fundamental and that the Agenda's Goals and targets should be met for all nations and people and for all segments of society. Ensuring that these commitments are translated into effective action requires a precise understanding of the target populations and progress made in addressing their particular priorities.

To properly measure this, statistics need to be presented for different population groups and geographical areas. The Sustainable Development Goal (SDG) indicator framework has included an overarching principle of data disaggregation: SDG indicators should be disaggregated, where relevant, by income, sex, age, race, ethnicity, migratory status, disability and geographic location, or other characteristics, in accordance with the Fundamental Principles of Official Statistics.

To enable national statistical offices to estimate disaggregated indicators, guidelines are needed to support the process. The idea of writing guidelines on how to use statistical methods and, in particular small area estimation (SAE), to receive disaggregated statistical indicators is not new. Some focus on methodological aspects, others provide methodology in a specific program language or focus on a specific topic as poverty mapping. Several statistical institutions conducted projects on the evaluation of the usability of SAE for official statistics. In 2020, the Asian Development Bank even published practical guidelines especially focusing on the monitoring process of the Sustainable Development Goal with SAE. So how do these guidelines differ from the existing work?

The idea of the **SAE4SDG Toolkit** in Wiki is to complement and use the existing methodological work and case studies to encourage and enable national statistical offices

SUSTAINABLE DEVELOPMENT GOALS

Table of contents

1. Why is SAE important for SDG data disaggregation
2. Practical guidance on producing SAE
3. Communicating SAE estimates
4. SAE by Sustainable Development Goals
5. From SAE experiment to production: the enabling environment
6. SAE practices
7. Software packages for SAE
8. Key readings on SAE
9. International and national SAE projects
10. Frequently asked questions

Guiding through steps with practical example

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

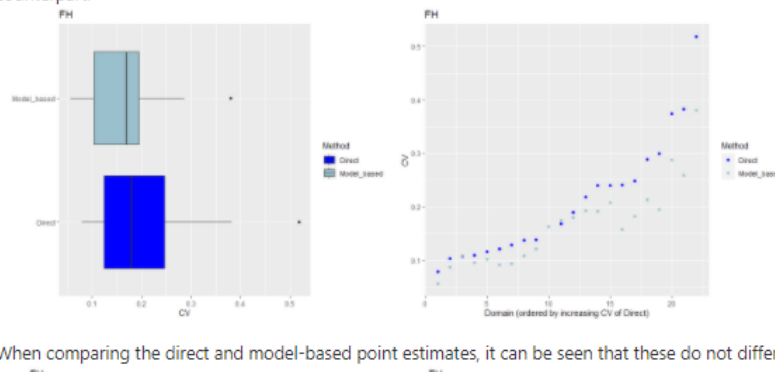
> Expand source

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.

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

No case studies yet

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

No case studies yet

Goal 6. Ensure availability and sustainable management of water and sanitation for all

No case studies yet

Goal 7. Ensure access to affordable, reliable, sustainable and modern energy for all

› [Case studies](#)

Goal 8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all

SAE for official statistics: lessons learnt

Dashboard / SAE4SDG   29 views

From SAE experiment to production: the enabling environment

Created by Haoyi Chen, last modified on Apr 21, 2021

Small area estimation has been in the field for many years but using it for official data production is still uncommon. It is important to understand the underlying reasons for the slow onset of SAE in the official data arena and identify "non-tech" areas that should be emphasized as creating an "enabling environment" for small area estimation.

- Challenges in using SAE for official data production
- Enabling environment to enable the use of SAE for official data production
 - Establishing a clear and focused objective that links SAE to data use for policymaking
 - Fostering an environment for research and development
 - Government commitment and sustainable financial support to SAE experimentation and production
 - Design-based versus model-based estimates: a changing culture in the national statistical offices
 - Usable input data for SAE
 - [Maintaining a high and fit-for-purpose quality standard](#)
 - Collaboration
 - Capacity building
 - Disclosure control
 - Transparency in releasing methodology and communicating quality
- Practical way forward: from experimental statistics to official statistics

Challenges

- Lack of support from upper management (resources)
- Lack of technical capacity
- Lack of proper input data
- Unsure about the use of model-based estimates
- Difficult to communicate the method and results

Government commitment and legal mandate

- Requirement of disaggregated data by law, to distribute funding
- Building a team

Input data

- Data access
- Data quality

Collaboration

- Researchers
- Other government agencies and private sector
- Other data community: IT/cloud infrastructure, processing and technical capacity
- Within NSO:
 - Subject-matter experts
 - Geospatial experts

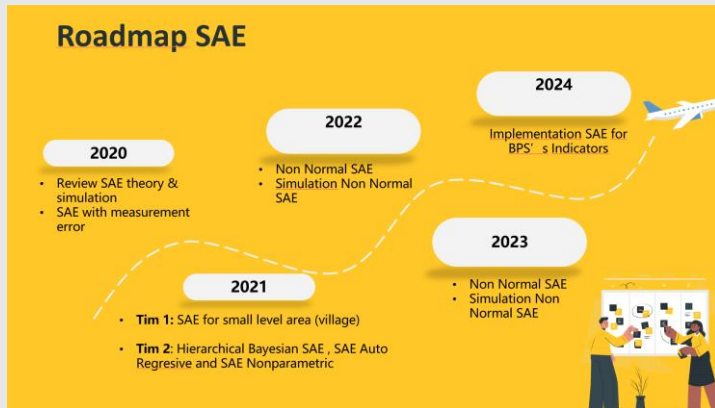
Capacity building

- What is the most effective way?

Quality standard

- Quality assurance
- External evaluation

SAE for official statistics – national examples



Example: United States SAIPE Program

In September 1994, the Congress passed the Improving America's Schools Act and signed it into law (PL 103-382). Title I of the law specifies the distribution of Federal funds to school districts 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 Department of Commerce."

This law further requires that in Fiscal Year 1997, the Secretary of Education use updated data on poor children for counties and, beginning in Fiscal Year 1999, updated data for school districts, published by the Department of Commerce, unless the Secretaries of Education and Commerce determine that the use of updated population data would be "inappropriate or unreliable."

It also directs the Secretary of Education to fund a National Academy of Sciences panel to provide advice on the suitability of the Census Bureau estimates for use in allocating funds.

Source: [Small Area Income and Poverty Estimates \(SAIPE\) Program, Origins of the Project](#)

Challenges in using SAE for official data production From National Statistical Offices

- "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 are lack of 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."

Model-based estimates at Statistics Netherlands

In a more recent paper from Statistics Netherlands (Buelens, Wolf and Zeelenberg, 2016), a set of guidelines were provided that can be used to evaluate. Those interested in more details should refer to the original paper.

1. General principle. The general principle when using model based estimation in official statistics, is the principle that official statistics give a de
a. Objectivity: data used to estimate the model should be related to the subject of the statistic of interest. The model should only be used to estimate the model, but estimation should not exceed the present.
b. Reliability: failure of the model should not lead to changes in the (conclusions based on the) estimate of the statistical phenomenon. The model should be used to estimate the model, but estimation should not exceed the present.
2. The use of models.
a. Goal. The goal of using model based estimation should be to estimate data that is not available, and as such to improve the overall estimation.
b. Data. Models are used to estimate missing data. Both for fitting the model as well as for the final estimation procedure, only data that is available should be used.
c. Standard. Model based methods that are used at Statistics Netherlands should follow any general consensus in the literature on similar methods.
d. Model selection. Alternative models should be considered, in order to find the most appropriate model. With model selection, the aim is to find the model that best fits the data.

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 members—49 from the

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, helping them comply with 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 vulnerable sectors of society.

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 official statistics compilation. The guide 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 outcomes.


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 several national statistics offices. The guide also addresses the 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 methods.

 Like Be the first to like this



Write a comment...

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 Bureau. SAIPE provides annual estimates of income and poverty for small areas. For more information, please contact 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 poverty is only available every 10 years, 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 Department of Commerce." This law from the Department of Commerce, unless the Secretaries of Education and Commerce determine that the use of updated population data would be "inappropriate or unreliable." It also directs the Secretary of Education to

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 "inappropriate or unreliable".
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 well-funded because of the legislative

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) reciprocity rate

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

Reference: 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

The 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 not sufficiently precise for formal use in

The 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 small and for some small geographic

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 many changes in its models and estimates



Consultations

- ❑ Key SAE experts: consultation meeting organized by JPSM Technical Group on SAE, May 2021
- ❑ Emails and focus-group discussions
 - Australia, Canada, Chile, Colombia, Indonesia, Jamaica, Republic Moldova, Philippines, South Africa, US, UK, Viet Nam
- ❑ Next steps:
 - Release the first stage of the Toolkit (**this week!!**)
 - Advertise the Toolkit so more countries can use it
 - Approaching more countries and document the challenges/lessons learned
 - Organise small technical group discussion (countries + academic)
 - Capacity building

eLearning course (UNSD, ECLAC, UNFPA)

- Course format
 - 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
- Self-paced learners:
 - Learn at their own pace
 - Access all the above learning materials
 - Machine graded weekly assessments
 - Access to projects, R script and data – not graded
- Guided learners
 - Guided learning and need to follow a fixed schedule that entails about 1-hour of work per day for ten weeks, reading assigned materials, watching course videos, and completing the assigned projects
 - 2-hour interactive workshop per week for ten weeks that will cover a summary of the weekly learning materials and instructions of R code that can be used for SAE modeling
 - Feedback and grading for all three projects

Thank you

Haoyi Chen
chen9@un.org

