Challenges and lessons learned in using small area estimation for official statistics – how do we help countries?

Francesca Grum, Assistant Director

United Nations Statistics Division

unstats.un.org

Outline

❑Briefly: ISWGHS and IAEG-SDGs

❑Why the UN Small Area Estimation (SAE) Toolkit?

❑From experiment to production: challenges

❑Lessons learnt from countries that did it successfully

❑What is next?

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 + 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.

ISWGHS

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)

OLDER PERSONS   PERSONS WITH DISABILITIES   CHILDREN   WOMEN   MIGRANTS AND REFUGEES

ISWGHS

Why SAE Toolkit (https://unstats.un.org/wiki/display/SAE4SDG/)

❑Many countries have experimented with SAE in the past but very 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: SAE is very resource intensive therefore needs to be focused on a smaller number of indicators. The SDG indicators are therefore a perfect framework to start with if resource is limited.

- Provides hands-on exercise, including "semi-synthetic" data (national data + noises) and programing guide. The data sets are used to illustrate, for 3 SDG indicators, on how to work with a complete cycle of SAE estimation such as data preparation, selection of methods, analysis and adaptation and evaluation/benchmarking.

- 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 a long discussion, based on our discussion with countries, on the challenges and enabling environment for countries to move from SAE experiment to official production. We hope this discussion can help countries build appropriate measures, in additional to technical capacity, in using SAE for official production.

- 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 example

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

SAE practices

Dashboard / SAE4SDG · 38 views

SAE practices

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

Asian Development Bank

FAO

UNICEF

US Census Bureau

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 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 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 f 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 Educat

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.

ministrative data:

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

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

urce: 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

uality 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 re 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

djustment made on the model and estimates

provements 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 estim

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 th

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 wit 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 comp 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 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...

ISWGHS

Challenges in using SAE for official statistics

- Lack of interest and support from the top management, hence lack of resource
- Lack of dedicated resources for SAE research and implementation
- While one household survey can produce a large number of indicators, great efforts are necessary for SAE to derive just one indicator for small domains
- 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

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

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

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

ISWGHS

Lessons learnt: Input data in countries

❑Chile:
- CASEN survey (cross-sectional multipurpose household survey)
- Comuna level administrative data

❑Colombia:
- Integrated household survey (GEIH)
- Population census

❑Indonesia:
- Indonesian National Socioeconomic Survey
- Village Potential Statistics (PODES)

❑Jamaica:
- Jamaica Survey for Living Condition
- 2011 census

❑US:
- American Community survey, Current Population Survey (Annual)
- Administrative data: Income tax; Supplemental Nutrition Assistance Program participants data; Supplemental Security Income Recipiency rate

ISWGHS

Lessons learnt: maintaining a high and fit-for-purpose quality standard

❑ Internal assessment to evaluate the models, the estimation procedure and corresponding results

- Compare with direct estimates, at national, urban/rural, principal cities and state level (Colombia)
- Coefficient of variation requirement (CV):
  - Colombia: CV < 30% for publishing
  - ISTAT: CV ≤15% for domains; CV ≤18% for small domains
  - Statistics Canada: CV ≤16.5% no release restriction; 16.5%<CV ≤33.3% add warnings; >33.3% not recommended for release

❑ External/independent evaluation:

- Public consultation: consultation is carried out with local government (Indonesia)
- Review by experts:
  - A National Academy of Sciences panel was funded to provide advise on the suitability of the Census Bureau estimates for use in allocating funds (United States)

ISWGHS

Lessons learnt: effective capacity building

❑Step 1: Organising broad training on SAE methods and why SAE outcomes are important to inform policy.

❑Step 2: Providing technical training for staff working on SAE: covering basic foundations and introducing different methods and models.

❑Step 3: customized hands-on training specific to the outcome indicator

- A good understanding of data needs: outcome indicator and the level of disaggregation
- Assessing input data availability/quality/timeliness/auxiliary variables
- All exercises should be carried out by national staff, ideally also using country data

ISWGHS

Consultation/ next steps

❑Consultation with key SAE experts

❑Working with countries through focus-group discussions

- Australia, Canada, Chile, Colombia, Indonesia, Italy, Jamaica, Republic of Moldova, Philippines, South Africa, US, UK, Viet Nam

❑Next steps:

- Continue to add case studies and national experiences
- Provide training to countries though an eLearning course currently being developed by ECLAC-UNSD-UNFPA
- Organise small technical group discussion (countries + academic) to address specific questions from countries
- Explore potential of using non-traditional data sources such as remote sensing and mobile phone data for SAE

ISWGHS

Thank you