US Census Bureau

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 statistics for all school districts, counties, and states. More information about what SAIPE produces is available here. The following information was compiled based on the 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. The real push for the creation of and the financial support for SAIPE came after the Improving America's Schools Act (PL 103-382) that 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 (NAS) panel to provide advice on the suitability of the Census Bureau estimates for use in allocating funds. (Source: SAIPE, original of the Project)

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 support.

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 not sufficiently precise for formal use in producing school district poverty estimates at that time. (Craig Cruse and David Powers, Estimating School District Poverty with Free and Reduced-Price Lunch Data, 2006)

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 small and for some small geographical areas (county or school-district level), there is no data point from the survey. Then producing small area estimates for these geographical areas would be challenging.

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 estimation procedures. Some are minor refinement while others have major impact. More information on major changes is available in Bell, Basel and Maples, 2015. Two considerations were covered in the discussion in making changes:

- Comparability of data over time. Any changes made in the methods will impact on comparability; and needs to be communicated with users. Therefore any major revisions need to be considered carefully before implementation.
- Major data source update such as replacing the 1990 Census data with data from Census 2000; and switch the use of CPS to ACS data. During this time it might be advisable to make major methodological revisions.

Quality standard

Small area estimation, when used for official purposes, needs to follow quality standards of official statistics. Relevant quality principle about the use of modelling is included in the UN National Quality Assurance Frameworks Manual for Official Statistics Use of modelling:

- Quality principle 10: Assuring methodological soundness. When statistical modelling is used in the statistical production process (e.g., for seasonal adjustment), the validity of model assumptions is carefully considered and the impact on final estimates is evaluated.

The following is a description of what the US Census Bureau has on quality standard for model estimates. The standard applies to models used to produce estimates such as small domain estimates, including small area estimates. Requirements that are related to SAE are detailed as below:

Requirement D2-1: Throughout all processes associated with estimation, unauthorized release of protected information or administratively restricted information must be prevented.

*Requirement D2-2***:** A plan must be developed that addresses:

1. Purpose and rationale for using a model (e.g., data to compute precise estimates are not available, or modeling with additional data will provide more accuracy).
2. Key estimates that will be generated and the domain of application for the model.
3. Methodologies and assumptions related to the model, such as the:
    1. Model structure (e.g., functional form, variables and parameters, error structure, and domain of interest).
    2. Model estimation procedure (e.g., least squares estimation, maximum likelihood estimation, and demographic estimation methods).
    3. Data source and how the data will be used in the model, including key modifications to the data.
4. Criteria for assessing the model fit (e.g., goodness-of-fit statistics and R-squared) and the model specification (e.g., measures of multicollinearity).
5. Verification and testing of the systems for generating estimates.
6. Verification of the modeled estimates and evaluation of their quality

*Requirement D2-3***:** Models must be developed and implemented using statistically sound practices. Examples are then provided on what might be considered as "statistically sound" for model development practice, demographic estimates and projections, and seasonal adjustments.

Sub-Requirement D2-3.1: Model results must be evaluated and validated, and the results of the evaluation and validation must be documented.

Sub-Requirement D2-3.2: Specifications for the modeling and estimation systems must be developed and implemented.

Sub-Requirement D2-3.3: Estimation systems must be verified and tested to ensure that all components function as intended.

Sub-Requirement D2-3.4: Methods and systems must be developed and implemented to verify the modeled estimates and evaluate their quality.

Sub-Requirement D2-3.4.1: The seasonal adjustment process and results must be reviewed annually by the program manager (or the appropriate mathematical statistician) to identify needed changes in the X-12-ARIMA specification files.

Sub-Requirement D2-3.4.2: For indicator releases, any routine revisions to the annual review process, such as benchmarking and updating of seasonality factors, must be consolidated and released simultaneously.

Requirement D2-4: Documentation needed to replicate and evaluate the modeling activities must be produced.

Managing confidential microdata level data when building SAE estimates

During the process of building SAE estimates, individuals involved might include census bureau staff, academia and independent consultants, who could have different levels of access rights to confidential individual level data. How to manage the data access needs to be considered. There are in-house staff who are specialised in SAE. There are also collaborators from Academia. Sensitive data are not given to external collaborators. The collaborators will provide the methods and codes and the Census staff will run the program and share the results with the external collaborators.

Information provided by Wesley Basel, William R. Bell and Jerry J. Maples, Census Bureau, United States of America.