Communicating SAE methods and results

Fruitful communication and dissemination of small area estimation results requires:

- knowledge about the target audience (e.g. policy makers, researchers, users in statistical offices);
- identifying the best possible form of communication for the desired audience.

Two important aspects should be communicated with users:

- methodology, including input data, models used, assumptions, data validation process, data quality and comparability over time
- results, including how to interpret the results and ways to (or not to) use the results

- Communicating SAE method
- Communicating SAE results
  - Visualization
    - Regional disaggregation
    - Other domain disaggregation
  - Exporting estimation results
Communicating SAE method

Communication around the SAE method should be carried out for two broad groups of users:

1. Users who require some general information about the methods and how data can be used safely;
2. Users who are interested in specific models, assumptions made and the validation process.

For the first group, easy-to-understand language should be employed while links to documents describing technical details should be provided to those users interested in these. Some examples below are how countries communicate SAE with their users are presented below.

“There is a need for high-quality income statistics at the smallest possible geographical level.”

“The requirement for data on income was previously reflected by Census User Groups who made a strong case for a question on income to be included in the 2001 Census. Although this need was recognised by the government, concerns were raised about the public acceptability of asking people about their income, and the risks this could have on the overall number of census returns.”

“Alternative methods for obtaining data on income at the small area level were identified and implemented. One of the options identified was the use of small area estimation methodologies to produce small area income estimates.”

“The method for producing small area estimates combines survey data with auxiliary data that are correlated with the target variable. The approach is to create a model that relates the survey variable of interest (for example, income) to these auxiliary variables (covariates).”

“The survey sample is too small to provide reliable direct estimates for small areas or domains, but synthetic estimates can be made based upon the model parameters and values for the covariate data, which are available for all the small areas.”

“Synthetic estimation produces estimates for domains where survey data are insufficient, by borrowing strength from other data sources. The other data sources (known as auxiliary data or covariates) are available on an area basis and for all areas in the target population. At the level of these small areas, sample survey sources are not generally available, so the covariate data are usually from some administrative system or a previous census.”


“These model-based estimates of average household income in MSOAs are not calculated in the same way as the national and regional household income estimates published separately by the Office for National Statistics (ONS). The definitions of income and data sources used for these statistics are different. It is not possible, therefore, to aggregate the estimates up to match the regional and national estimates.”


The most recent UK small area estimates are accompanied by a technical report that covers (a) methodology (models); (b) input datasets for modelling; (c) developing the models; (d) quality of the estimates; (e) comparing results with earlier rounds of estimates; and (f) guidance on the use of the estimates.

Communicating SAE results

Different forms of communications are widely used for the dissemination of small area estimates. These include reports, presentations and infographics. Commonly all these different forms of communication make use of tables, plots and maps. Since it is rather unlikely to communicate small area estimates without any visualization, a couple of different visualization approaches are presented in the next section.

Visualization

In this section, some widely used visualization techniques are presented.

Regional disaggregation

⚠️ No interpretation of results

Please note that none of the results can be interpreted in any kind. The data is solely used to illustrate the kinds of plots that can be used, and are completely artificial, hence no analysis can be conducted.
The regional distribution of the estimated indicators of interest can most meaningfully be presented by using maps. The following map shows the regional distribution of the poverty HCR in Colombia (the estimates displayed here are based on synthetic data and cannot be interpreted in any sense). The map illustrates the regional variation of the HCR and should enable policy makers to target public interventions aimed at alleviating poverty more precisely. When the disaggregation dimension is a combination of geographic location and ethnic group, several maps must be prepared (one for each ethnic group). It can also be useful to produce maps for the MSE and/or the CV to visualize the distribution of the estimates’ uncertainty. In that way, areas or regions of high uncertainty can be identified and studied more closely.

Other domain disaggregation

If the focus is on non-geographical disaggregation dimensions, e.g., age and disability status, mapping is not feasible. To visualize the indicators of non-geographical domains, other visualization techniques, such as barplots can be used. The plot below shows the distribution of the unemployment rate over the defined domains - age groups and disability status. In this example, the unemployment rate is on average higher for the population with a disability compared to the population without disability. However, the differences in the unemployment rate differ strongly among the age groups.
Exporting estimation results

It is advisable to export the final estimation results to a csv-file for further processing and visualization in Excel or other tools. Exporting the results is easily feasible in R by saving the results as 'data frames' and writing them to a csv-file. The following R code exemplary shows the export of poverty estimates to a csv file. The data corresponds to the example data used in the production process.
Get estimates

# Load packages
library(emdi)
library(maptools)

# Set working directory
setwd("Add path")

# Import sample and census at household level
survey <- read.csv("syntheticSurvey1.csv")
# The census csv was too large for the upload, thus it is available as RData file
load("syntheticCensus.RData")

# Data preparation

# Convert categorical variables to factor variables
census$classwkd <- factor(census$classwkd)
census$sex <- factor(census$sex, levels = c(0,1), labels = c("m","f"))
survey$classwkd <- factor(survey$classwkd)
survey$urban <- factor(survey$urban)
survey$electric <- factor(survey$electric)
survey$sex <- factor(survey$sex, levels = c(0,1), labels = c("m","f"))

# Fit final model

# Please note that L and B are set to low values to reduce the computation time
# For real applications, L and B need to be higher
povEBP_final <- ebp(fixed = eqIncome ~ age + sex + yrschool + classwkd,
                     pop_data = census, pop_domains = "geolev2",
                     smp_data = survey,
                     smp_domains = "geolev2", MSE = TRUE, transformation = 'log',
                     L = 10, B = 5)
Export estimation results

# Export results to csv

# Save results as data frame

```r
df_estimation_results <- as.data.frame(estimators(povEBP_final,
        indicator = c
        ("Median",
        "Head_Count",
        "Gini"),
        MSE = TRUE, CV = TRUE))
class(df_estimation_results)
# "data.frame"

# Write data to a csv file

write.csv(df_estimation_results$ind, row.names = FALSE,
          file = "estimation_results.csv")
```

Afterwards the estimation results can be opened with Excel or other tools.

Figure: Extract of the estimation results exported opened in Excel