Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Statistical Office FSO

Work for the Swiss Federal Statistical Office

Small Area Estimation in the Structural Survey

Executive Summary

EXPERIMENTAL STATISTICS

Neuchâtel, 2018

Published by:	Federal Statistical Office (FSO)	**Layout concept:**	Section DIAM
Information:	info.pop@bfs.admin.ch, tel. +41 58 467 25 25	**Downloads:**	www.statistics.ch
Editors:	Section METH, FSO	**Copyright:**	FSO, Neuchâtel 2018 Reproduction with mention of source authorised (except for commercial purposes)
Contents:	Ewa Strzalkowska and Isabel Molina Department of Statistics, Universidad Carlos III de Madrid		
Topic:	00 Statistical Basis and Overviews		
Original text:	English		

Small Area Estimation in the Structural Survey:

Ewa Strzalkowska and Isabel Molina
Department of Statistics, Universidad Carlos III de Madrid

December 21, 2015

EXECUTIVE SUMMARY

OVERVIEW: When estimating at a detailed geographical level (districts, communes), direct estimators that make use only of the available area-specific data, such as the currently used generalized regression (GREG) estimators, are unreliable for areas with small sample sizes. Confidence intervals based on these estimates become too wide and hence not very informative for those areas. Small area estimation (SAE) methods make use of auxiliary information available in registers that may be related with the study variable measured in a survey. The relationships between the study variable and the auxiliary variables are represented in the form of regression models that link all the areas. The small area estimators derived from those models (model-based estimators) "borrow strength" from the other areas through the model, increasing considerably the effective sample size, which results in increased efficiency.

GLOBAL AIMS: The aims of this project were to study and apply SAE techniques for estimation of the proportions of active people in the Swiss communes. For caution, in a first stage SAE methods were studied for estimation in districts, and in a second stage some studies were repeated for estimation in communes.

PROJECT OUTPUT: The main products of this project are:

- **Regression models:** Regression models that explain the fact of being active or not, both in Swiss districts or communes have been constructed: This entailed a process of selection of powerful regressors and their adequate shapes in the model. The model was validated by various means including residual diagnostics. Districts or communes that appeared poorly explained by the model were represented in the model by specific parameters that lead to GREG-type estimators for those districts or communes, avoiding a bias in their model-based estimates.

- **Comparison of estimators:** Different types of model-based estimators were compared in terms of bias and efficiency with direct estimators in simulation experiments of different types: assuming that the model holds or not, when varying the population values or when varying the samples.

- **Selected estimator:** A model-based estimator (called LMM along the report) was selected due to its good performance together with its feasibility of implementation.
- **Adjusted estimator:** Model-based estimates of totals for each area (district/commune) do not add up to the GREG estimates of the national total. Similarly, they do not add up to the GREG estimates of the canton totals. Ratio-adjusted estimates that add up to the national total or to the canton totals were also studied. These estimates are called benchmarked estimates (at national or cantonal level) along the report.
- **Mean squared error:** Mean squared error (MSE) estimators based on several bootstrap procedures were proposed: parametric, nonparametric, a mixture of the two called mixed bootstrap, and a specific bootstrap approach for estimation of design-based MSEs called parametric design bootstrap (PDB). The last two are the first decent procedures for estimation of the design MSE of model-based estimators.
- **Confidence intervals:** Confidence intervals based on the selected model-based estimates and the developed MSE estimates were constructed.
- **Properties of MSE estimators:** Properties of MSE estimators, together with coverage and length of confidence intervals were analyzed in simulation experiments.
- **Applications with data:** Model-based estimators of the area proportions of active people and developed MSE estimators were computed for districts and communes.

MAIN PROJECT CONCLUSIONS

- **Good model fit:** The obtained model fits quite well the Structural Survey data. Resulting estimates of regression coefficients are meaningful and help to understand the factors that affect the activity/inactivity in Switzerland.
- **Consistent with GREG:** The obtained model-based estimates approach the current GREG estimates for areas (districts, communes) with larger sample sizes.
- **Drastic CV reductions:** Model-based estimators have CVs that are 49% below the CVs of GREG estimators when estimating in districts and approximately 90% below when estimating in communes. These figures are medians of CV reductions over all districts/communes. Model-based estimators reduce the CV in practically all districts and communes, and reductions are specially large for those districts and communes with smaller sample sizes. Lengths of confidence intervals are reduced accordingly.
- **Limited design-bias:** Model-based estimates keep the design bias below 5% for the areas with at least 20 observations in the survey and below 10% for the areas of at least 10 observations.
- **Decent design MSE estimators:** For estimation of the MSE of the selected LMM estimators, we proposed two bootstrap procedures, called parametric design bootstrap (PDB) and mixed bootstrap (MB), and described in the report on Phase I of

contract num. 150011928. Simulation results indicate that the PDB MSE estimator tracks well the peaks of the design MSE (largest design MSE values) without underestimation for areas with sample size larger than 100, although the coverage of 95% confidence intervals obtained using this MSE estimator can be underestimated by about 5% on average. On the other hand, the MB procedure, when used to obtain 95% confidence intervals, preserves the nominal 95% level when averaging over groups of communes. Similarly, it estimates the design MSE adequately on average but tends to smooth the peaks of the design MSE. These procedures can be still improved with further research, but they are the first decent procedures for estimation of the design MSE of model-based estimators.