A Framework for Producing Small Area Statistics

Nikos Tzavidis
Professor of Statistical Methodology & Head of School
Southampton Statistical Sciences Research Institute & Social Statistics & Demography
University of Southampton

UN World Data Forum
Bern, Oct. 3 - Oct 6, 2021
Background

- Interest in SAE due to its relevance to policy decisions
- Area of growing methodological developments
- Increasing complexity of targets of estimation
- Significant progress with uptake of SAE methods
- Increased availability of regularly updated, publicly available forms of data
- However, SAE not always part of a framework for producing official statistics
SAE Stages: Specification

- User needs
- Data availability
- SAE methods
- Targets of estimation & geography
- Initial choice of methods (& software)
**SAE Stages: Specification**

- **Data availability & Targets of estimation**
  - Linear statistics (averages/proportions), access to area level auxiliary data is sufficient
  - Non-linear statistics (inequality measures), access to auxiliary (Census/admin) micro-data preferable but not necessary

- **Data availability & Target geography**
  - Study sample coverage at different levels of geography. Top-to-bottom approach
  - Can we obtain adequate estimates when most of the areas are out of sample? High dependency on model assumptions

- **SAE methods & Targets of estimation**
  - Increasing interest in complex parameters
  - Availability of SAE methods for different problems
  - Software that integrates data processing & estimation stages
SAE Stages: Analysis/Adaptation

Problem-specific but as a starting point

- **Initial triplet of estimates**: Set of estimators obtained as by-product of the existing estimation processes:
  1. Direct estimator
  2. Synthetic estimator
  3. Composite estimator

  Accompany with estimates of precision. Produce sets of estimates at different levels of geographic/domain aggregation.

- Consider use model-based SAE methods following the principle of parsimony
SAE Stages: Analysis/Adaptation - Use of models in SAE

- Initial triplet may offer required precision
- Even if it does, consider the use of models. It may help with improving precision
- Model selection, evaluation & diagnostics important steps
- Use of more complex e.g. machine learning methods can capture complex data relationships
- Software availability makes this attractive. But use cautiously - Avoid 'Black-Box' approach
SAE Stages: Evaluation

Uncertainty assessment
- Process of obtaining a measure of uncertainty of SA estimates
- Particularly important in the context of official statistics

Method evaluation
- Comparison of estimators under alternative methods
- Assessment of an estimator’s performance under departures from model assumptions
Thoughts on topics of interest

- Reduce reliance on Census micro-data
- Increased availability of public, frequently updated data e.g. remote sensing data
- Use of area-level models - Data access advantages
- Use of area-level models - Naturally accommodate e.g. remote sensing data
- Open source software widely available *
- Need for systems integrating data harnessing & processing with statistical processing
- Need to continue building capacity

Thank You

Prof. Nikos Tzavidis
n.tzavidis@soton.ac.uk
https://www.southampton.ac.uk/demography/about/staff/ntzav1.page

Work supported by the Data and Evidence to End Extreme Poverty (DEEP) research programme. DEEP is a consortium of the Universities of Cornell, Copenhagen, and Southampton led by Oxford Policy Management, in partnership with the World Bank - Development Data Group and funded by the UK Foreign, Commonwealth & Development Office.
https://povertyevidence.org