

0 00 00



# PISA, 8-10 MAY 2018

## WORKSHOP "SMALL AREA METHODS AND LIVING CONDITIONS INDICATORS IN EUROPEAN POVERTY STUDIES IN THE ERA OF DATA DELUGE AND BIG DATA"

## FINAL EVENT OF THE JEAN MONNET CHAIR SAMPLEU

## Some remarks on

- the production process of small area estimates,
- external validation of the estimates and
- dissemination

Tomasz Klimanek, Statistical Office in Poznan t.Klimanek@stat.gov.pl









<b>Table 3.</b> Problems and fisks of apprying SAE in orneral statistics (%)			
Problems and risks	NSIs	Universities	Total
Small sample size	42	39	40
Bias of indirect estimators	47	68	62
Model dependent approach	42	54	50
Difficulty in estimating the variance	53	32	38
There are no risks	5	2	3
Other	11	2	5

**Table 3** Problems and risks of applying SAE in official statistics (%)

*Notice:* Percentage of the number of indications among a specified group of respondents. There was a possibility of choosing more than one issue and therefore the percentages do not sum to 100%.

Source: Survey on teaching, use and/or development of SAE methods, July 2014.

Source: Gołata E., 2015, SAE Education Challenges to Academics and NSI

"Statistics Netherlands is generally rather reserved in the application of model-based estimation procedures for the production of official statistics. The most important reasons are that design-based approaches are robust against model-misspecification and are convenient to apply in production а environment to produce timely official statistics."

*Source:* Jan van den Brakel and Jelke Bethlehem, 2008, Model-Based Estimation for Official Statistics









- The principle that official statistics give a description of society as it is (or as it has been).
- Law on Official Statistics Act, 1995: "art. 3. Official statistics shall provide <u>reliable, objective</u> and systematic information for the society, State and public administration bodies, and business entities of the national economy on the economic, demographic and social situation, and the natural environment."
- Moreover, the European regulation on statistics (Regulation (EC) No 223/2009) includes the principle: "reliability: ... statistics must measure as faithfully, accurately and consistently as possible the reality that they are designed to represent"
- European Code of Practice includes the principle: *accuracy and reliability*: "European Statistics accurately and reliably portray reality."

Based on B. Buelens , P.P. Wolf , K. Zeelenberg, 2016, Model based estimation at Statistics Netherlands









### GSBPM - Generic Statistical Business Process Model – phases:

- 1. Specify Needs;
- 2. Design;
- 3. Build;
- 4. Collect;
- 5. Process;
- 6. Analyse;
- 7. Disseminate;
- 8. Evaluate.

CBS recommendations for the use of model-based estimation in official statistics:

- 1. Goal (cases: data are not available, improve the overall estimation process);
- 2. Data (only data that are available and relevant shoud be used);
- 3. Standard (models should follow the general consensus in the literature in similar situations);
- 4. Model selection (to choose between families of models and the estimation methods);
- 5. Model fit (model diagnostics);
- 6. Robustness (models should be robust against unexpected changes);
- 7. Stability (methods for statistics published on a periodical basis to be stable over time);
- 8. Mean Square Error (should be estimated, distinction between variance and bias, comparison with design-based estimators);
- 9. Assumptions (to be stated explicitly)
- 10. Publication (the use of the model-based approach to be explicitly mentioned and the model selection, assumptions and diagnostics to be trasparent to the reader)

Based on B. Buelens , P.P. Wolf , K. Zeelenberg, 2016, Model based estimation at Statistics Netherlands









Eurostat project **ESSnet on Small Area** Work package 6 – *Guidelines* (ISTAT, CBS, GUS, INE, ONS, SSB), 2012

" ... we shall outline a standardized approach, or process, to small area estimation based on existing sample survey data."

STAGES:

- 1. Clarification: needs, data, evaluation criteria
- 2. Basic smoothing (triplet of basic, synthetic and composite estimates)
- 3. Enhancement: modeling to address outstanding issues









### From start to finish: a framework for the production of small area official statistics

J. R. Statist. Soc. A, 2018, University of Southampton Freie Universitat Berlin

### STAGES:

- 1. Specification
- 2. Analysis & Adaptation (parsimony principle)
- 3. Evaluation









#### External validation is a great challenge

#### **Approaches:**

- model validation (with international accademic community: conferences, scientific bodies, individual experts),

",The user must remember that SAE is concerned with prediction and not with discovering associations and causal mechanisms between the explanatory variables and the outcome."

- validation of estimates with users (small area estimates local authorities, experts in specific area)
- validation with alternative data sources (census), proxy variables (LFS unemployment, registered unemployment)
- validation with alternative methods (triangulation method, Denzin 1978 the idea is that one can be more confident with a result if different methods lead to the same result)









**External validation – criteria** (Paradysz J., 2009): (to be met for a better public understandig of SAE estimation)

**formal internal** (level – small areas should sum up to the larger area (benchmarking) order – order of SAE estimates should reflect their order in the population order and distance – reflect not only the order of estimates but also distances between them)

formal external (largest area, predominating area, equal chances) - how to take into account ,, the importance" of the specific small area in the estimation process? Is it acceptable to take into account additional weights for ,, important" areas?

 factual evaluation (resistance to outliers, in accordance with the state of the art, spatial linearity of the phenomenon, logical set of coefficients)









Model-based estimation tested by the Centre for Small Area Estimation in the Statistical Office in Poznan, Poland:

- Non-response correction (weighting/calibration assuming the Missing At Random pattern) – production phase
- Small area estimation (Poverty mapping cooperation with the World Bank) experimental statistics

and several projects are quite mature but still not official statistics:

- Modifications of the poverty mapping project,
- SPREE estimation of disability,
- Monthly unemployment rate (modelling time series)









### Diversification of a complex measure and poverty rate estimated using the EB model

- complex measure:
  - low: -0.4 0.2
  - medium: 0.2 0.4
  - high: 0.4 0.7
- poverty rate:
  - low: 7.2470 22.2151
  - medium: 22.2151 26.8920
  - high: 26.8920 39.0571
- classification consistency: 61.7%













Thank you for your attention









