

# International cooperation of the Center for Small Area Estimation at the Statistical Office in Poznan and its contribution in the development of Polish statistics

Tomasz Józefowski<sup>1</sup> Tomasz Klimanek<sup>2</sup>  
Andrzej Młodak<sup>1</sup> **Marcin Szymkowiak**<sup>3,1</sup>

<sup>1</sup>Center for Small Area Estimation  
Statistical Office in Poznan

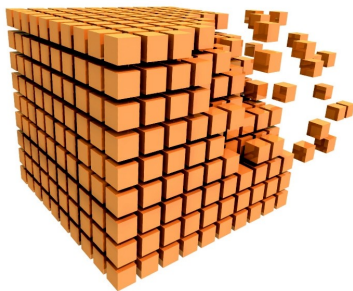
<sup>2</sup>Statistical Office in Poznan

<sup>3</sup>Department of Statistics  
Poznan University of Economics and Business  
e-mail: [m.szymkowiak@ue.poznan.pl](mailto:m.szymkowiak@ue.poznan.pl)

# Outline

- 1 Center for Small Area Estimation
  - Center for Small Area Estimation
  - Statutory tasks of the Center for Small Area Estimation
  - Main activities of the Center for Small Area Estimation
- 2 Main activities of the Center for Small Area Estimation
  - International cooperation (Eurostat, World Bank)
  - Cooperation with Central Statistical Office of Poland
  - Cooperation with local government
  - Cooperation with universities
- 3 Summary and references
  - Summary
  - References

# Center for Small Area Estimation



# OŚRODEK STATYSTYKI MAŁYCH OBSZARÓW

# Center for Small Area Estimation (CSAE)

- Research unit established in 2009 at the Statistical Office in Poznań,
- The main task of the Center for Small Area Estimation is to conduct work on the use of small area estimation methods for official statistics, and in particular, for local statistics.
- Other research activities: survey sampling, statistical data integration (statistical matching, probabilistic record linkage), weighting methods (calibration), trainings for data users etc.

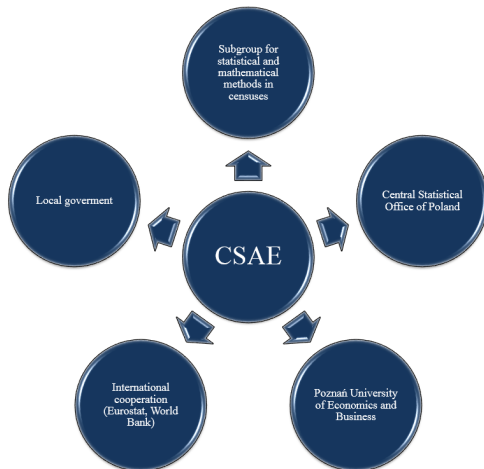
# Statutory tasks of the Center for Small Area Estimation

- 1 conducting works on the development of small area estimation methodology,
- 2 recognizing possibilities of using small area estimation methods in the context of administrative sources,
- 3 recognizing the actual and potential demand for information obtained using small area estimation methods,
- 4 conducting design and programming works concerning small area estimation methods,
- 5 developing indirect estimation methodology using statistical data integration methods,

# Statutory tasks of the Center for Small Area Estimation

- 6 cooperation with other public administration bodies, institutions and organizations aimed at integrating and coordinating statistical research and data acquisition, including participation in the work of international teams, inter-ministerial and expert groups,
- 7 preparing documents and materials for meetings and consultations for the purposes of international cooperation.

# Main activities of the Center for Small Area Estimation



## Estimation of poverty rate at NUTS 3 level

- In 2013 the Center for Small Area Estimation in cooperation with the Central Statistical Office and the World Bank prepared a poverty map of Poland at the level of subregions (NUTS 3) using the Fay-Herriot approach.
- By applying the Fay-Herriot area level model it was possible to produce estimates of the at-risk of poverty rate (ARPR) in Poland at the level of subregions, i.e. at a lower level of aggregation than the direct estimates published by official statistics so far. This has increased the scope of information about poverty: it is now available at the level of 66 subregions.



## Background on Poverty Mapping in Poland

- According to the Central Statistical Office (CSO), the poverty indicator for the whole country in 2011 based on EU-SILC survey amounts to 17.7%.
- CSO does not publish information about the poverty indicator for lower levels of spatial aggregation (NUTS 2 and below). This information is available only at the national and regional level (NUTS 1).

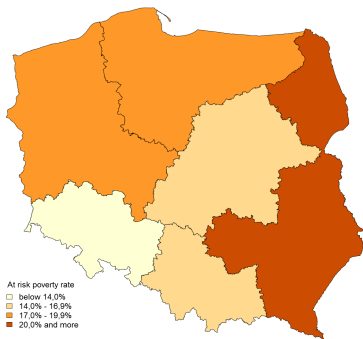


Figure: The poverty rate at the level of regions (NUTS 1)

## Data sources

- EU-SILC - the only variable from the EU-SILC survey included in the model was the poverty indicator.
- Many variables from the 2011 National Census of Population and Housing were considered as explanatory ones
- Data from the Local Data Bank were also considered.

## Explanatory variables

The final model included 6 explanatory variables.

- the percentage of single people over 25 — 2011 National Census of Population and Housing;
- the number of rooms per one household member — 2011 National Census of Population and Housing;
- the percentage of households with a bathroom — 2011 National Census of Population and Housing;
- the percentage of households with two persons aged over 25 with no more than vocational education — 2011 National Census of Population and Housing;
- population density — Local Data Bank;
- the ratio of people deregistered to the number of people registered for permanent residence in the subregion — Local Data Bank.

## Spatial variation in poverty in Poland based on SAE

- There is a strong territorial variation in the poverty indicator.
- There is a difference between Central and Eastern Poland (with a higher poverty rate) and Western Poland, characterized by a lower scope of poverty.

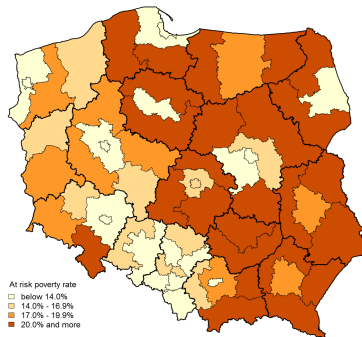


Figure: ARPR at subregions level (NUTS 3)

## Spatial variation in poverty in Poland based on SAE

- The lowest values of the poverty indicator can be observed in big cities.
- Most subregions surrounding big cities exhibit significantly lower levels of poverty (below 13%) than other subregions in the same province.

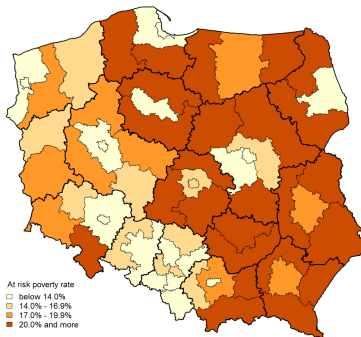


Figure: ARPR at subregions level (NUTS 3)

# Comparison of ARPR – direct versus indirect estimation

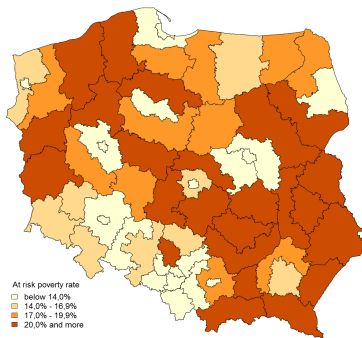


Figure: ARPR at subregions level (NUTS 3) – direct estimation

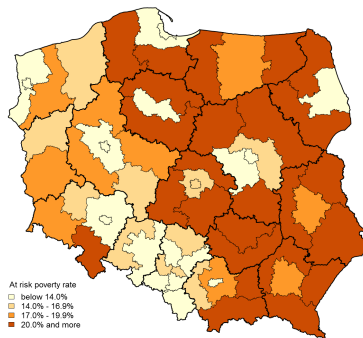


Figure: ARPR at subregions level (NUTS 3) – indirect estimation

## Precision of estimators

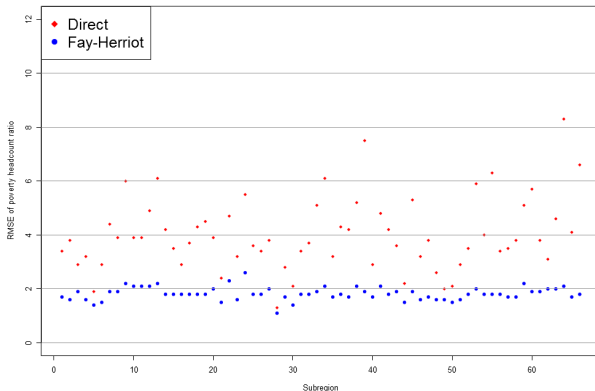


Figure: Precision of estimators of ARPR at subregions level (NUTS 3)

## Cooperation with the World Bank – summary

- Results obtained by using the Fay-Herriot model are characterized by better precision than direct estimates.
- These estimates are also more accurate than direct estimation.
- The study was a starting point for further work devoted to the use of small area estimation methods in official statistics at lower levels of spatial aggregation (NUTS 4/LAU 1).



## Cooperation with Eurostat

- ESSnet on Small Area Estimation
- ESSnet on Data Integration
- The MeMoBuSt project
- Meets project
- Monthly Unemployment Rate project
- VIP Admin project

## ESSnet on Small Area Estimation

**General objective:** development of a framework for the production of SAE for social surveys

- IT – Istituto Nazionale di Statistica (Coordinator) (ISTAT)
- FR – Institut National de la Statistique France et des Etudes Economiques (INSEE),
- GE – Statistisches Bundesamt (DESTATIS)
- NE – Centraal Bureau voor de Statistiek (CBS)
- NO – Statistisk Sentralbyrå (SSB)
- PO – Główny Urząd Statystyczny (GUS)
- SP – Instituto Nacional de Estadística de España (INE)
- UK – Office for National Statistics (ONS)
- SW – Swiss Federal Statistical Office (FSO)

## Aims of the ESSnet on Small Area Estimation

- **Literature review of methods and applications** – update the state of the art: previous EU projects, Book on SAE, describe the applications in, UE and non-UE, NSIs create a common knowledge base on applications of SAE methods;
- **Literature review of quality indicators** – review and develop suitable criteria to assess the quality of SAE methods for the choice of the proper model and the evaluation of MSE;
- **Provide software tools** – make available software tools for SAE in the ESS;
- **Foster knowledge transfer** – development of case studies and associated recommendations on representative problems for SAE in the ESS;
- **Provide guidelines** – provide practical guidelines in ESS social surveys context;
- **Transfer knowledge** – transfer know-how to non-participating NSIs and disseminate results.

# Output

- Poland was the leader of the package **Literature review of methods and applications**.
- This package was aimed to provide a comprehensive overview of small area estimation in the ESS social surveys with respect to implementation, needs and expectations.
- The last ten years of SAE literature and other SAE project outcomes have been reviewed.
- Moreover a survey on NSIs' SAE experiences has been carried out.
- Final report about this package prepared by the Center for Small Area Estimation can be found at [https://ec.europa.eu/eurostat/cros/system/files/WP2\\_Final\\_Report.pdf](https://ec.europa.eu/eurostat/cros/system/files/WP2_Final_Report.pdf).

## ESSnet on Data Integration

**General objective:** to focus on methodologies for data integration (Record Linkage, Statistical Matching, Micro integration Processing) and on statistical aspects that need to be considered in order to facilitate practical applications of those methods by NSIs.

- IT – Istituto Nazionale di Statistica (Coordinator) (ISTAT)
- NE – Centraal Bureau voor de Statistiek (CBS)
- NO – Statistisk Sentralbyrå (SSB)
- PO – Główny Urząd Statystyczny (GUS)
- SP – Instituto Nacional de Estadística de España (INE)
- SW – Swiss Federal Statistical Office (FSO)

## Aims of the ESSnet on Data Integration

- **State of the art** – completing the state-of-the-art on record linkage, statistical matching and micro integration processing. The output consists of an updated reference containing sound methods.
- **Development of methods** – dealing with those methodological developments that should be obtained in order to make record linkage, statistical matching and micro integration processing applicable in the ESS.
- **Development of common software tools** – software on record linkage and statistical matching.
- **Case studies** – the main aim was to show how some real world problems can be tackled by the methods of record linkage and statistical matching.
- **Dissemination towards the ESS** – organizing of on-the-job training, courses, workshops and meetings with experts.
- **Management**

# Output

- Center for Small Area Estimation was involved in all Work Packages within ESSnet on Data Integration.
- Center for Small Area Estimation organized a workshop on statistical matching with experts from ISTAT.
- All materials may be found at [https://ec.europa.eu/eurostat/cros/content/data-integration\\_en](https://ec.europa.eu/eurostat/cros/content/data-integration_en).

# The MeMoBuSt project

- Main aims:
  - creation of a new methodological handbook on enterprise statistics taking into account various types of surveys and user needs
  - analysis of methods of data verification, response burden and data integration
  - easy online access to the handbook
  - use of the handbook as the basis for training within the European Statistical System and for possible new studies
- Participants: team of statisticians representing the statistical offices from the Netherlands (coordinator), Greece (only in the Phase 1), Hungary, Italy, Norway, Poland, Sweden and Switzerland.



## The MeMoBuSt project

- Structure of the handbook
  - 22 chapters
  - theme modules
  - method modules
  - glossary
- Polish contribution:
  - leader and author: *Questionnaire design and Response*
  - author: Different Types of Surveys (chapter *General Observations*), Calibration, Outlier Treatment, Synthetic Estimators for Small Area Estimation, Composite Estimators for Small Area Estimation (modules in the chapter *Weighting and Estimation*).
  - reviewer: *User Needs*, Balanced Sampling for Multi-Way Stratification (chapter *Sample Selection*), *Data Collection, Dissemination*.
- Publications: [Willenborg et al. (2014)] - handbook, [Młodak(2013)]

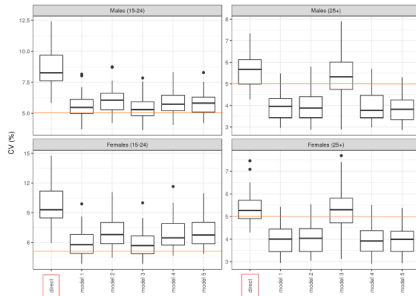
# Quality improvement of the Monthly Unemployment Rate

Center for Small Area Estimation took part in the EU funded project: "**Quality improvement of the Monthly Unemployment Rate**" – MUR. Project goals:

- review methods of measuring unemployment,
- compile LFS microdata for indirect estimation of MUR,
- assess data on registered unemployment in terms of their usefulness for MUR estimation,
- select the best method for weight calibration in LFS,
- select the best method for indirect estimation of MUR (Structural Time Series).

# Quality improvement of the Monthly Unemployment Rate

Ditect vs. indirect estimation



- Structural time series models significantly improve precision compared with direct estimation.
- Calibration of design weights could improve the precision of indirect estimation of MUR.

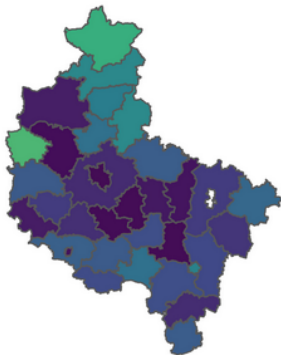
## VIP ADMIN

The main project task carried out by the Statistical Office in Poznań (Center for Small Area Estimation and Center for Urban Statistics) was to estimate the actual and de jure marital status based on existing available data sources.

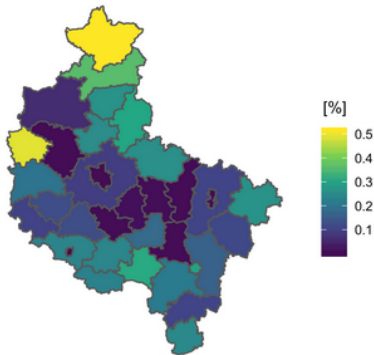
- Different data sources were taken into account: administrative registers and chosen surveys (EU-SILC, LFS, HBS, etc.).
- Different data integration techniques were used.
- Deterministic (using a common identifier in the case of administrative registers with the PESEL variable).
- Deterministic (using pseudo-identifier based on the date of birth and full address details of persons in the case of administrative registers without the PESEL variable).
- Probabilistic record linkage - in the case of surveys based on sampling.

# Estimation of marital status based on integrated dataset

Single (male)



Single (female)



## Cooperation with Central Statistical Office of Poland

- Calibration of weights in the National Census of Population and Housing 2011,
- Application of SAE methods in terms of labour market in the National Census of Population and Housing 2011 (methodological paper),
- Application of integrated calibration and repeated weighting methods in censuses based on the mixed approach,
- Application of small area estimation methods to poverty in Poland at LAU 1 level,
- Foreigners in the national labour market – regional approach,
- Methodological paper "Disclosure and Elaboration Control of Output Data".

## Statistical Disclosure Control

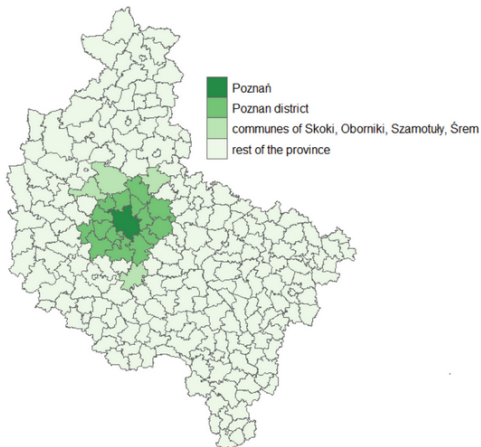
- Methodological paper "Disclosure and Elaboration Control of Output Data" being the result of international cooperation (The MeMoBuSt project, ESTP trainings in Eurostat, etc.) and recognition of the needs of Polish official statistics (growing demand for detailed data),
- Presentation of the most important aspects and methods of the SDC with many practical examples.
  - general description (concepts and definitions, legal issues),
  - statistical disclosure control (methods for microdata and tabular data protection, assessment of disclosure risk and information loss, software:  $\tau$ -Argus,  $\mu$ -Argus, R packages: `sdcTable` and `sdcMicro`),
  - general principles of access to data (rules of establishing access to microdata for scientific purposes, organization of such access, combining data files, safe presentation of output data: descriptive statistics, results of analyses, illustrations).

## Poznań agglomeration

- created in 2007 as a way of fostering cooperation between the city of Poznan and **17 surrounding** municipalities (the Poznan district),
- later on extended by including **4 municipalities** from outside the Poznan district,
- 11% of the province's area,
- 30% of the total population (appr. 1 million),
- 40% of all companies registered in the province,
- provide employment to 40% of all employees.



# Poznań agglomeration



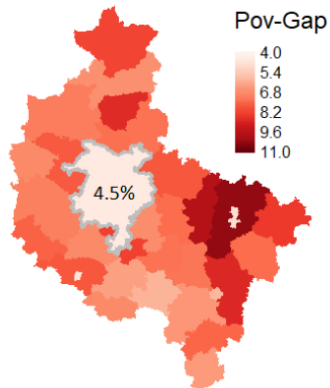
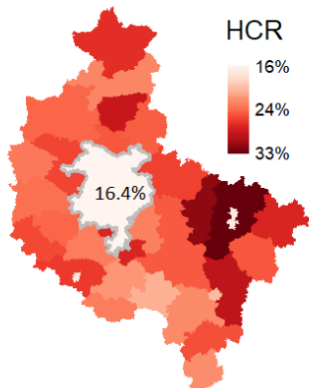
## Poverty indicators

- **At-risk-of-poverty rate (head count ratio (HCR), poverty incidence)** - the share of people with an equivalised disposable income (after social transfers) below the poverty threshold, (60 % of the national median EDI)
- **The poverty gap index** - sometimes referred to as poverty gap ratio or PG index, is defined as average of the ratio of the poverty gap to the poverty line. It is expressed as a percentage of the poverty line for a country or region.

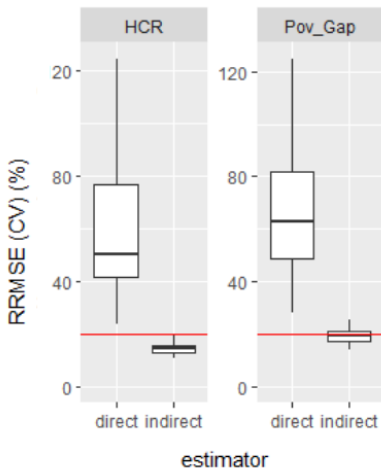
## Datasets and variables used in the EB unit-level model

- **Datasets:** EU-SILC and Census 2011
- equivalised disposable income
- share of men
- share of people aged 65 or older
- share of unemployed persons
- share of disabled persons
- share of people with primary education
- share of people with higher education
- ratio of children (aged  $\leq 15$ ) to persons aged 16-65
- binary – HH in rural area or in a municipality  $\geq 20000$
- binary – HH with 1 room
- binary – HH with 3 or more rooms

## Results: HCR (poverty rate) & Poverty gap



## Relative root mean square error (CV) (%)





## Cooperation with Universities

- participation of students as apprentices and interns in the work of the Center for Small Area Estimation,
- participation of employees of the Center for Small Area Estimation in scientific conferences and seminars organized by Department of Statistics at University of Economics and Business in Poznań (e.g. SAE2014),
- conducting statistical courses by employees of the Center for Small Area Estimation at post-graduate studies „Advanced analytical techniques in business. Postgraduate studies under the SAS Institute patronage” ,
- cooperation with the Poznań University of Technology.

## Summary

- Center for Small Area Estimation can be considered as a key research center which is using modern estimation techniques in practice.
- This is the only such unit in Europe in the structure of official statistics.
- **Conference on Current Trends in Survey Statistics: 13–16 August 2019 (Singapore)** - the conference will be devoted to small area estimation, statistical data integration, data confidentiality, disclosure methods and privacy assessment i.e. to topics under the main interest of the Center for Small Area Estimation.

## References

-  Willenborg L., Scholtus S., van de Laar R. (eds.) (2014) *Handbook on Methodology for Modern Business Statistics*, Collaboration in Research and Methodology for Official Statistics, [http://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics\\_en](http://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en).
-  Młodak A. (2013) *Coherence and comparability as criteria of quality assessment in business statistics*, *Statistics in Transition – new series*, vol. 14, pp. 287–318.



Thank you for your attention



Statistics Poland