HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Small Area Estimation

Spring 2015

Topic 1: Introduction to SAE

Risto Lehtonen, University of Helsinki

Lecture topics

- **Practicalities**
- **Topic 1: Introduction to SAE**
- **Topic 2: Basic concepts and approaches**
- **Topic 3: Direct estimators for domains**
- **Topic 4: GREG and calibration estimators**
  - Part I: Direct GREG and calibration
  - Part II: Indirect GREG estimators
  - Part III: Extended family of GREG estimators
- **Topic 5: EBLUP estimators for domains**
- **CASE STUDY-1: SAE in SILC data**
- **CASE STUDY-2: EBLUP example**

Lecture topic 1

- **Introduction to small area estimation (SAE)**
  - Motivation: why SAE?
  - Key definitions
  - Estimation tasks in SAE
  - Software
  - Examples
  - Main literature

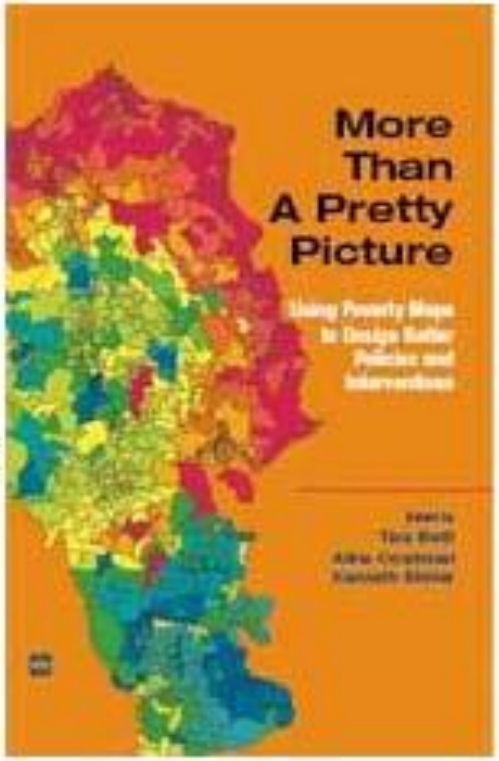
Small area estimation: World-wide trend

- An increasing need in society for reliable statistics for regional and other population subgroups or domains
  - Why? Let us discuss this for a while...
- SAE: Challenge for scientific research
  - What kind of a challenge? Let us discuss this point...
- SAE: Challenge for statistics describing the society
  - Why?

EXAMPLE: Poverty mapping

More than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions

The allocation of resources and the design of policies tailored to local-level conditions require highly disaggregated information. Data on poverty at the local level is typically not available because most household surveys are not representative past the regional level. This volume aims to promote the effective use of Small Area Estimation poverty maps in policy making. It presents the range of policies and interventions which have been informed by poverty maps, focusing on the political economy of poverty maps and the key elements to their effective use by policy makers. The volume also looks at the future of poverty maps in terms of new techniques and new areas of application.



(See Terms & Conditions for use of these materials)

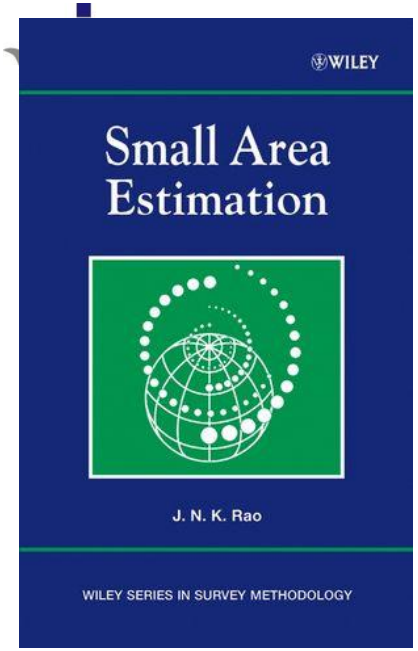
Full Document (4.3mb PDF)

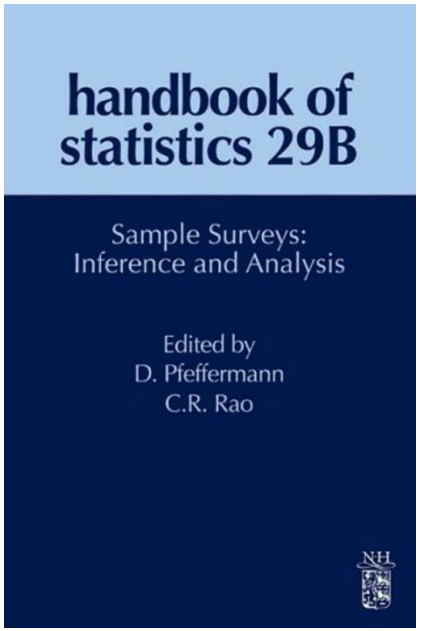
Lively SAE Research under EU's Framework Programmes

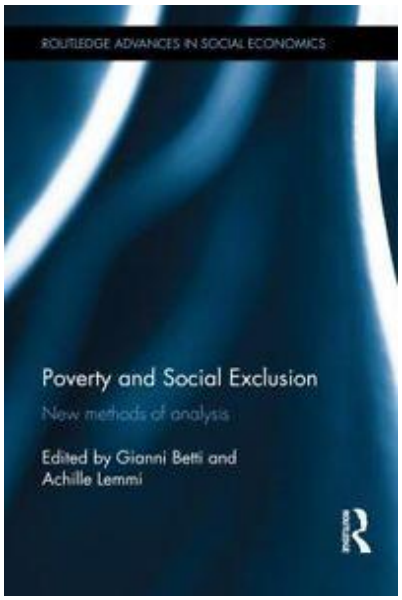
- Small area estimation research under Framework Programmes (FPs) in Europe
- Actors: Universities & NSIs
  - EURAREA Project (2001-2004), EU / FP5
  - AMELI Project (2008-2011), EU / FP7
  - SAMPLE Project (2008-2011), EU / FP7

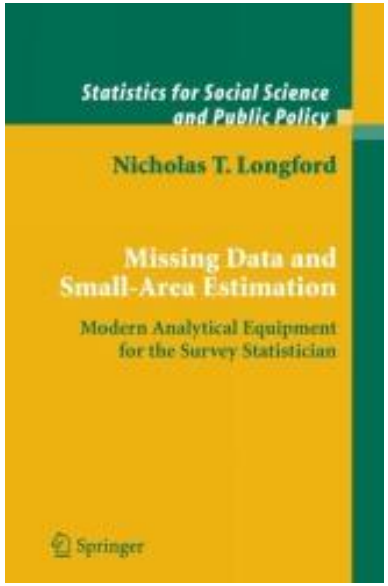
Series of SAE Conferences

- EWORSAE European Working Group on Small Area Estimation http://sae.wzr.pl/
  - SAE1993 (Warsaw, Poland)
  - SAE2000 (Riga, Latvia)
  - SAE2005 (University of Jyväskylä)
  - SAE2007 (University of Pisa)
  - Recent Advances in SAE (RRC09) (Univ. of Trier)
  - SAE2009 (University of M. Hernandez, Elche)
  - SAE2011 (University of Trier)
  - SAE2013 (Bangkok)
  - SAE2014 (Poznan University of Economics, Poland)
  - SAE2015 (Santiago, Chile)
  - SAE2016 ...


WILEY
Small Area Estimation
J. N. K. Rao
WILEY SERIES IN SURVEY METHODOLOGY


handbook of statistics 29B
Sample Surveys: Inference and Analysis
Edited by
D. Pfeffermann
C.R. Rao


ROUTLEDGE ADVANCES IN SOCIAL ECONOMICS
Poverty and Social Exclusion
New methods of analysis
Edited by Gianni Betti and Achille Lemmi


Statistics for Social Science and Public Policy
Nicholas T. Longford
Missing Data and Small-Area Estimation
Modern Analytical Equipment for the Survey Statistician
Springer

Main materials for this course
see course website

- Rao J.N.K. (2003). *Small Area Estimation*. New York: John Wiley & Sons.
- Lehtonen R. and Pahkinen E. (2004). *Practical Methods for Design and Analysis of Complex Surveys.* Second Edition. Chichester: John Wiley & Sons. Chapter 6.
  Web extension: VLISS-virtual laboratory in survey sampling
  http://vliss.helsinki.fi

- Lehtonen R. and Veijanen A. (2009). Design-based methods of estimation for domains and small areas. Chapter 31 in Rao C.R. and Pfeffermann D. (Eds.). *Handbook of Statistics. Sample Surveys: Inference and Analysis. Vol. 29B.* New York: Elsevier.

- Lehtonen R. and Veijanen A. Model-assisted methods to small area estimation of poverty indicators. In: Pratesi M. (Ed.) (2015). Analysis of Poverty Data by Small Area Estimation. Chichester: Wiley. (Forthcoming, to be distributed to participants)

Additional SAE materials

- EURAREA
- Downloads

- SAMPLE
- Downloads

- BIAS
- Downloads

Useful background materials

- Survey sampling
  - Lehtonen R. and Djerf K. (2008). *Survey sampling reference guidelines.* Luxembourg: Eurostat Methodologies and Working papers
    see course website

What is estimation for domains and small areas - 1?

- Domains of interest: Well-defined population subgroups
- **Examples, social survey**
  - Regional areas constructed by administrative criteria: county, municipality,…
  - Demographic criteria: sex and age grouping
  - Demographic breakdown within regional areas
- **Examples, business survey**
  - Grouping of enterprises into domains according to the type of industry

What is estimation for domains and small areas - 2?

- *Estimation for domains*, or *domain estimation* for short, refers to the estimation of population quantities, such as:
  - Totals – the total number of people in poverty for the given regions
  - Means – Mean disposable income across regions
  - Proportions – Proportion of ILO unemployed across regions
  - Medians, Quantiles, Percentiles…

  for the desired population subgroups called **domains** (small or large)

Special case - SAE

- *Small area estimation,* SAE
  - Estimation for domains whose **sample size** is small or very small (even zero)
  - Alternative definition (Partha Lahiri):

    Small area = Domain of interest, for which the sample size is not adequate to produce reliable **direct estimates**

More general framework: Small area statistics

- Production of statistics for domains and small areas based on sample survey data and/or population data from registers
- **SAE approach**
  - The target variable can only be measured using sample survey data (e.g. ILO employment/unemployment)
  - Statistics are produced using sample survey data and auxiliary data & models to improve accuracy of estimates
- **Register-based small area statistics approach**
  - The data for the target variable can be obtained from registers (e.g. Income based on tax data)
  - Production of statistics based solely on population data from statistical registers

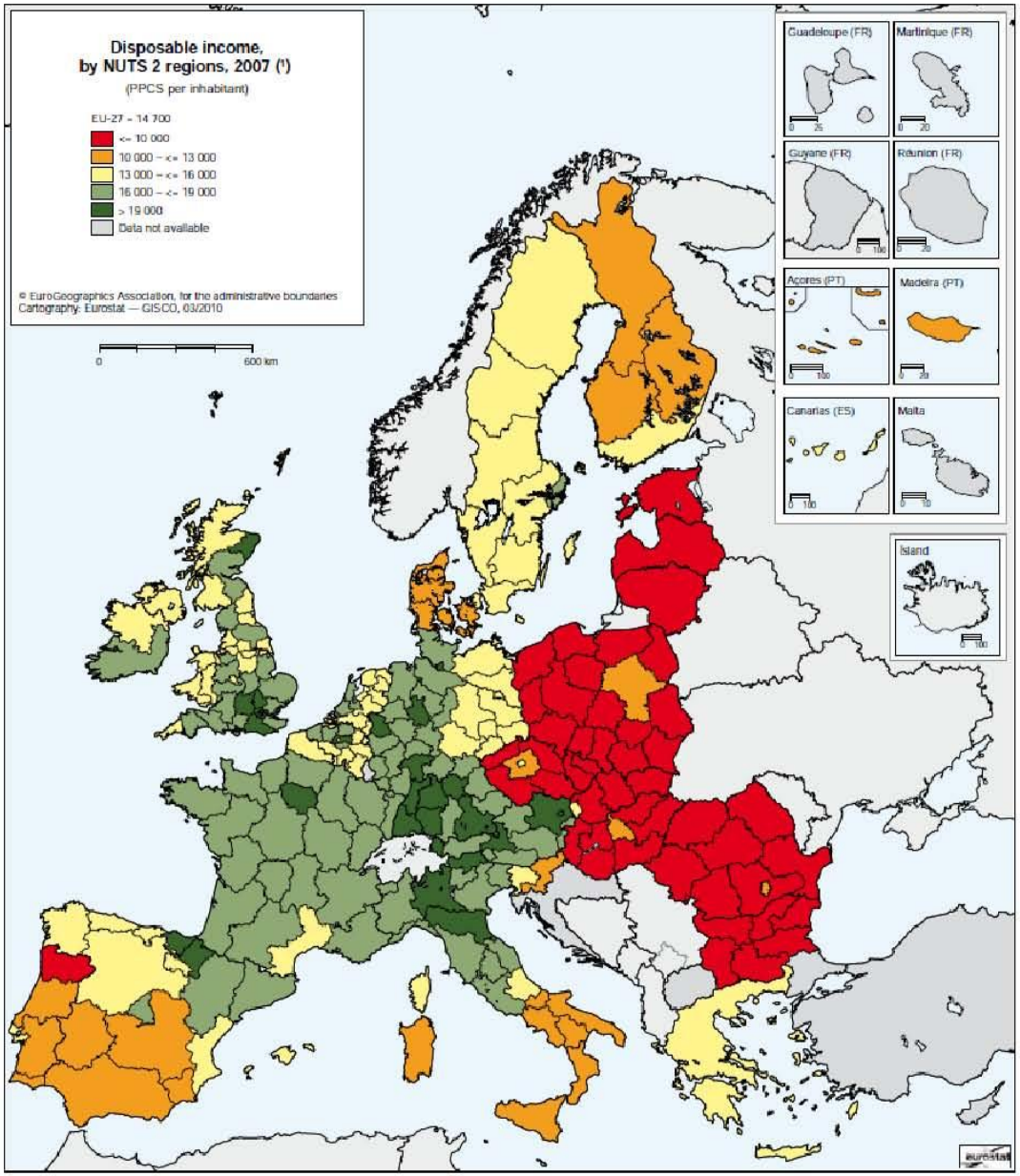
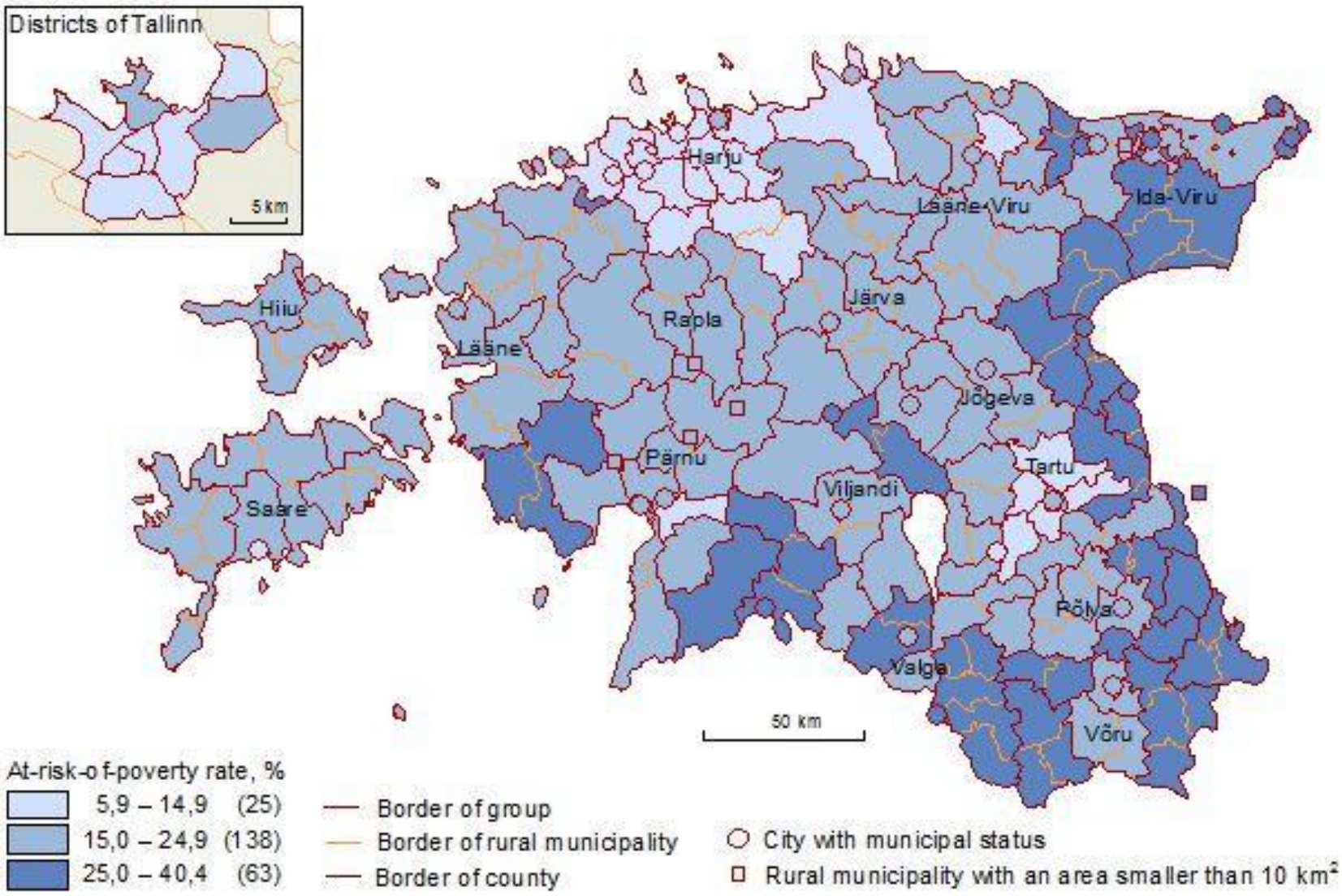


Figure 1. Disposable income by NUTS 2 regions in 2007 in the European Union

Source: Eurostat Regional Yearbook 2010, p.93, Section on Household Accounts. Information about the metadata is available at http://epp.eurostat.ec.europa.eu/cache/ITY_SDDS/EN/reg_ecohh_esms.htm

Poverty map: Estonia

World Bank 2014 – Regional poverty rates based on SILC data



Typical estimation task - 1

- Specify and identify the domains of interest
  - Breakdown of population into sub-populations
  - The number of domains can be large
- Specify study variable *y*
- Specify target parameters for the domains
  - Totals
  - Means
  - Ratios
  - Percentiles, Medians, ...

Typical estimation task - 2

- Specify data sources
  - Sample survey data
  - Auxiliary data: Census, Admin. Registers, Statistical registers
- Specify estimators of domain parameters
- Specify variance and MSE estimators
- Computation, graphical illustration
- Quality assurance
- Publication

Software

- SAS procedures
  - PROC SURVEYMEANS
  - PROC SURVEYREG
  - PROC MIXED
- SAS macro programs
  - SAS Macro EBLUPGREG (Ari Veijanen)
- R tools
  - Package SAE
- Stand-alone programs
  - Program DOMEST (Ari Veijanen)

Examples

- Estimation of regional number of ILO unemployed by sex and age group, based on Labour Force Survey LFS
- Estimation of median household disposable income by municipality, based on sample survey data such as EU SILC
- Estimation of regional poverty indicators, such as regional poverty rate, based on sample survey data such as EU SILC
- UK: Child poverty

Discussion topics

- Let us discuss possible application areas of domain / small area estimation in YOUR environment!
- Let us discuss the two types of statistical infrastructures in EU countries – “Survey” countries and “Register” countries!
  - Consequences to domain estimation and SAE

Selected literature

- Battese, G.E., Harter, R.M., and Fuller, W.A. (1988), An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data, *JASA* 80, 28–36.
- Betti G. and Lemmi A. (Eds.) (2013). *Poverty and Social*
- *Exclusion: New Methods of Analysis*. Routledge.
- Datta G. (2009). Model-based approach to small area estimation. Chapter 32 in Rao C.R. and Pfeffermann D. (Eds.). *Handbook of Statistics. Sample Surveys: Inference and Analysis. Vol. 29B.* New York: Elsevier.
- Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *JASA* 87, 376-382.
- Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedure to census data. *JASA* 74, 269–277.

Selected literature (contd.)

- Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: an appraisal. *Statistical Science* 9, 55–93.
- Jiang J. and Lahiri P. (2006). Mixed model prediction and small area estimation. *TEST* 15, 1–96.
- Lehtonen R., Särndal C.-E. and Veijanen, A. (2003). The effect of model choice in estimation for domains, including small domains. *Survey Methodology*, 29, 33–44.
- Lehtonen R., Särndal C.-E. and Veijanen A. (2005). Does the model matter? Comparing model-assisted and model-dependent estimators of class frequencies for domains. *Statistics in Transition,* 7, 649–673.

Selected literature (contd.)

- Lehtonen R. and Pahkinen E. (2004). *Practical Methods for Design and Analysis of Complex Surveys.* Second Edition. Chichester: John Wiley & Sons. Chapter 6.
- Lehtonen R. and Veijanen A. (2009). Design-based methods of estimation for domains and small areas. Chapter 31 in Rao C.R. and Pfeffermann D. (Eds.). *Handbook of Statistics. Sample Surveys: Inference and Analysis. Vol. 29B.* New York: Elsevier.
- Longford N. (2005). *Missing Data and Small-area Estimation: Modern Analytical Equipment for the Survey Statistician*. New York: Springer.

Selected literature (contd.)

- Molina, I. and Rao, J.N.K. (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistics*, Volume 38, Issue 3, 369–385.
- Münnich, R., Zins, S., Alfons, A., Bruch, C., Filzmoser, P., Graf, M., Hulliger, B., Kolb, J.-P., Lehtonen, R., Lussman, D., Meraner, A., Myrskylä, M., Nedyalkova, D., Shoch, T., Templ, M., Valaste, M. and Veijanen, A. (2011): Policy Recommendations and Methodological Report. Research Project Report WP10 of the EU/FP7 AMELI Project.
- Münnich, R., Burgard J.P. and Vogt M. (2009). Small area estimation for population counts in the German Census 2011. JSM 2009, Section on Survey Research Methods.

Selected literature (contd.)

- Pfeffermann D. (2013). New important developments in small area estimation. *Statistical Science* 28, 40–68.
- Rao J.N.K. (2003). *Small Area Estimation*. New York: John Wiley & Sons.
- Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology* 33, 99–119.
- Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer.
- Torabi, M. and J.N.K. Rao (2008). Small area estimation under a two-level model. *Survey Methodology* 34, 11-17.