HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Calibration methods for domain and small area estimation

Risto Lehtonen, University of Helsinki
Ari Veijanen, Statistics Finland

Summer School on Survey Statistics, Kyiv, 22-26 August 2016

Outline

- Calibration approaches: Overview
- Calibration methods for small area estimation
- Comparison of methods (bias & accuracy)
  - Simulation experiment with artificial data
  - Simulation experiment with real data
- References

Some design-based calibration approaches

- Model-free calibration MFC
  - Deville & Särndal (1992)
  - Lehtonen & Veijanen (2009)
- Model calibration MC
  - Wu & Sitter (2001)
  - Lehtonen & Veijanen (2012, 2016)
- Hybrid calibration HC
  - Combination of MFC and MC
  - Lehtonen & Veijanen (2014, 2015)
- Multiple and ridge model calibration.
  - Montanari & Ranalli (2009)

TAXONOMY: Statistical calibration methods in survey statistics

	Model-free calibration MFC	Model calibration MC	Hybrid calibration HC
Weight calibration	Calibration to reproduce known population totals of auxiliary x-variables	Calibration to population total of y-variable values predicted by a model	Combination of MC and MFC, depending on modeling and coherence requirements
Typical study variable	Continuous	Continuous, binary, polytomous, count	
Typical target parameters	Totals, means	Totals, means, proportions, cell frequencies More complex statistics e.g. poverty indicators	
Model specification	No explicit model statement Linear relationships	Explicit model statement. Non-linear relationships e.g. Nonlinear regression models Generalized linear (mixed) models	
Level of auxiliary data	Aggregate level	Unit level	Unit level (MC part) Aggregate (MFC part)
Main aims & properties	“Multi-purpose” weighting Coherence with published statistics Efficiency improvement Reduction of coverage and non-response biases	"Single-purpose" weighting Efficiency improvement Reduction of coverage and non-response biases	"Single-purpose" weighting Efficiency improvement Coherence with published statistics Reduction of coverage and non-response biases

Calibration estimators for domain totals

Population U of N elements $k \in U$

Sample s drawn from U

π_k inclusion probability for $k \in U$

D sub-populations (domains) of interest $U_d \subset U$

$s_d = s \cap U_d$ sample falling in domain d

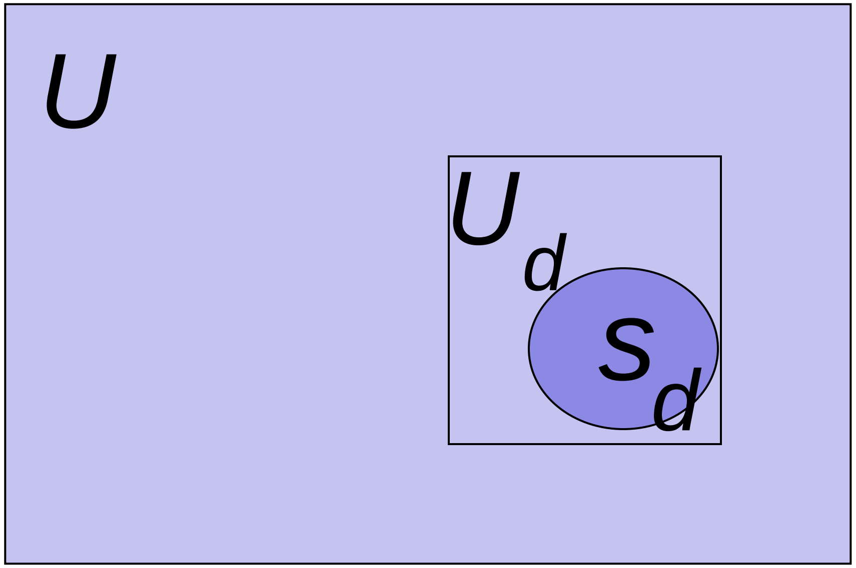
Domain totals $t_d = \sum_{k \in U_d} y_k,\ d = 1,\ldots,D$

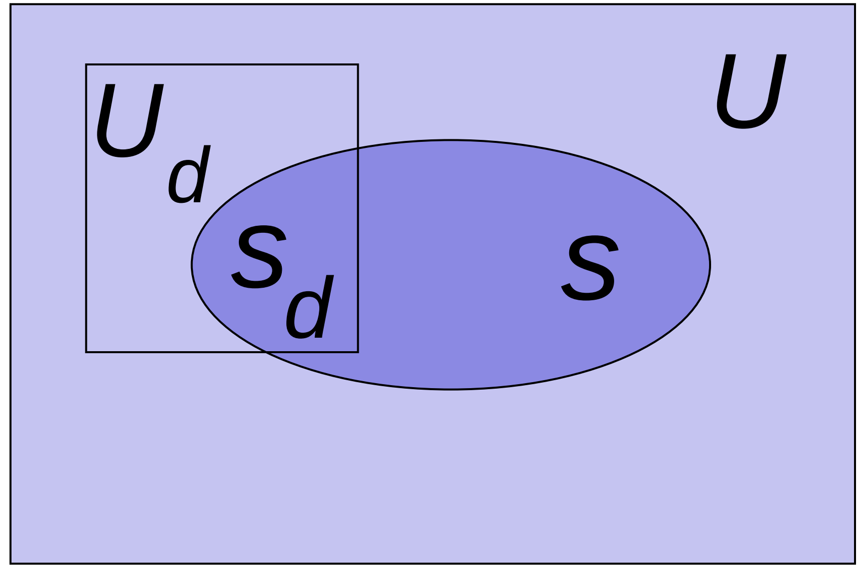
Calibration estimators $\hat{t}_d = \sum_{k \in s_d} w_k y_k = \sum_{k \in s_d} a_k g_k y_k$

$a_k = 1/\pi_k$ design weight

g_k method-specific g-weight for element $k \in s$

w_k method-specific **calibration weight** for element k





Planned domains

U Population

U_d Population domain d, $d = 1,\ldots,D$

Domains = Strata

$s_d \subset U_d$ Sample drawn in domain d

Sample size n_d is **fixed** by sampling design

Unplanned domains

U Population

s Sample

U_d Population domain d, $d = 1,\ldots,D$

$s_d = s \cap U_d$ Sample falling in domain d

Sample size n_d in domain d is **random**

THIS IS THE CASE FOR THIS EXERCISE

Calibration for model-free calibration

Calibration equation for model-free calibration

$$\sum_{k\in s_d} w_k \mathbf{x}_k = \sum_{k\in U_d} \mathbf{x}_k = \left(N_d, \sum_{k\in U_d} x_{1k}, \ldots, \sum_{k\in U_d} x_{pk} \right)'$$

$\mathbf{x}_k = (1, x_{1k}, \ldots, x_{pk})'$ known **calibration vector** for every $k \in U$

Minimize chi-square distance to design weights $a_k = 1/\pi_k$

$$\sum_{k\in s_d} \frac{(w_k - a_k)^2}{a_k} - \boldsymbol{\lambda}' \left(\sum_{k\in s_d} w_k \mathbf{x}_k - \sum_{k\in U_d} \mathbf{x}_k \right)$$

where $\boldsymbol{\lambda}$ denotes Lagrange coefficient

Calibration weights w_k for unit $k \in s_d$, $d = 1, \ldots, D$:

$$w_k = a_k \left(1 + \left(\sum_{i\in U_d} \mathbf{x}_i - \sum_{i\in s_d} a_i \mathbf{x}_i \right)' \left(\sum_{i\in s_d} a_i \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{x}_k \right)$$

General case

Calibration weights w_k minimize

$$\sum_{k \in s_d} \frac{\left(w_k - a_k\right)^2}{a_k} - \boldsymbol{\lambda}' \left(\sum_{k \in s_d} w_k \mathbf{z}_k - \sum_{k \in U_d} \mathbf{z}_k \right)$$

where $a_k = 1/\pi_k, \quad d = 1,\ldots,D$

$\mathbf{z}_k$ is **method-specific** vector of calibration variables

Calibrated weights are defined in:

$$w_k = a_k(1 + \boldsymbol{\lambda}'\mathbf{z}_k), \quad \text{where}$$

$$\boldsymbol{\lambda}' = \left(\sum_{i \in U_d} \mathbf{z}_i - \sum_{i \in s_d} a_i \mathbf{z}_i \right)' \left(\sum_{i \in s_d} a_i \mathbf{z}_i \mathbf{z}_i' \right)^{-1}$$

Model-free calibration equation

$$\sum_{k \in s_d} w_k \mathbf{z}_k = \sum_{k \in U_d} \mathbf{z}_k = \left(N_d, \sum_{k \in U_d} x_{1k}, \ldots, \sum_{k \in U_d} x_{pk} \right)'$$

where $\mathbf{z}_k = (1, x_{1k}, \ldots, x_{pk})'$, $s_d = s \cap U_d$, $d = 1, \ldots, D$

NOTE: Multi-purpose weighting

No y-variable involved

No explicit model statement

Coherence property for x-variable totals is met

Calibration of x-variable totals to the **domain level**

Direct MFC estimators of y-variable totals for domains

Model calibration equation: Semi-direct

$$\sum_{k\in s_d} w_k \mathbf{z}_k = \sum_{k\in U_d} \mathbf{z}_k = \left(N_d, \sum_{k\in U_d} \hat{y}_k \right)'$$

where $\mathbf{z}_k = (1, \hat{y}_k)'$, $s_d = s \cap U_d$, $d = 1,\ldots,D$

$\hat{y}_k = f(\mathbf{x}_k'(\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_d))$ with $\mathbf{x}_k = (1, x_{1k},\ldots,x_{pk})'$, $k \in U$

NOTE: Single-purpose weighting

Flexible modelling for y-variables of different types

Coherence property for x-variable totals is not met

Calibration of y-prediction totals to **domain level**

Semi-direct MC estimators of y-variable totals for domains

- modelling for the whole sample
- calibration at the domain level

EXAMPLE of assisting model in MC and HC: Linear mixed model

Linear mixed model for continuous study variable *y*

$$y_k = \mathbf{x}_k'\boldsymbol{\beta} + u_d + \varepsilon_k,\ k \in U_d,\ d = 1,\ldots,D$$

where $\mathbf{x}_k = (1, x_{1k},\ldots,x_{pk})'$, $\quad \boldsymbol{\beta} = (\beta_0, \beta_1,\ldots,\beta_p)'$

u_d are domain-level random intercepts

$u_d \sim N(0,\sigma_u^2)$, $\varepsilon_k \sim N(0,\sigma^2)$, u_d and ε_k independent

Estimate $\boldsymbol{\beta}$ and σ_u^2 from the data

Calculate estimates $\hat{u}_d$, $d = 1,\ldots,D$ and calculate fitted values

$$\hat{y}_k = \mathbf{x}_k'\hat{\boldsymbol{\beta}} + \hat{u}_d,\ \ k \in U_d,\ d = 1,\ldots,D$$

Can be used in linear mixed model assisted MC and HC (Lehtonen and Veijanen 2016a,b)

EXAMPLE of assisting model in MC and HC: Logistic mixed model

Logistic mixed model for binary response variable y

$$E_m(y_k \mid u_d) = \frac{\exp(\mathbf{x}_k'\boldsymbol{\beta} + u_d)}{1 + \exp(\mathbf{x}_k'\boldsymbol{\beta} + u_d)}, \quad k \in U_d,\ d = 1,\ldots,D$$

where u_d are domain-level random intercepts, $u_d \sim N(0, \sigma_u^2)$

Estimate $\boldsymbol{\beta}$ and σ_u^2 from the data

Calculate estimates $\hat{u}_d,\ d = 1,\ldots,D$ and calculate fitted values:

$$\hat{y}_k = \frac{\exp(\mathbf{x}_k'\hat{\boldsymbol{\beta}} + \hat{u}_d)}{1 + \exp(\mathbf{x}_k'\hat{\boldsymbol{\beta}} + \hat{u}_d)}, \quad k \in U_d,\ d = 1,\ldots,D$$

Can be used in logistic mixed model assisted MC and HC (Lehtonen and Veijanen 2016a,b)

Model calibration equation: Semi-indirect

$$\sum_{k \in c_d} w_k \mathbf{z}_k = \sum_{k \in U_d} \mathbf{z}_k = \left(N_d, \sum_{k \in U_d} \hat{y}_k \right)'$$

where $\mathbf{z}_k = (1, \hat{y}_k)'$, $c_d = s \cap C_d$, supersets $C_d \supset U_d$, $d = 1, \ldots, D$

NOTE: Weights $w_k^{(c)}$ are specific to area c_d, and $s_d \subset c_d$

$\hat{y}_k = f(\mathbf{x}_k'(\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_d))$ with $\mathbf{x}_k = (1, x_{1k}, \ldots, x_{pk})'$, $k \in U$

NOTE: Difference to semi-direct method

Calibration of y-prediction totals to **higher regional level**

Semi - indirect MC estimators of y-variable totals for domains

- modelling for the whole sample
- calibration at a regional level higher than the domain level

Hybrid calibration equation: Semi-direct

$$\sum_{k \in s_d} w_k \mathbf{z}_k = \sum_{k \in U_d} \mathbf{z}_k = \left(N_d, \sum_{k \in U_d} x_{1k}, \ldots, \sum_{k \in U_d} x_{jk}, \sum_{k \in U_d} \hat{y}_k \right)'$$

where $\mathbf{z}_k = (1, x_{1k}, \ldots, x_{jk}, \hat{y}_k)', \;\; s_d = s \cap U_d,\; d = 1, \ldots, D$

$\hat{y}_k = f(\mathbf{x}'_k(\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_d))$ with $\mathbf{x}_k = (1, x_{j+1,k}, \ldots, x_{pk})',\; k \in U$

NOTE: Combination of MFC and MC

x-variables in model part and calibration part are separate sets of variables or they can coincide or partially overlap

-Calibration of y-prediction totals to the domain level

-Calibration of selected x-variable totals to the domain level

-Coherence property for selected x-variable totals is met

Semi-direct HC estimators of y-variable totals for domains

Two-level hybrid calibration equation - 1

Basic idea: Two calibration equations, to be solved simultaneously for a single calibration weight variable

$$\sum_{k\in s_d} w_k \mathbf{z}_k^{(1)} = \sum_{k\in U_d} \mathbf{z}_k^{(1)} = \left(N_d, \sum_{k\in U_d} \hat{y}_k\right)' \quad \text{(MC part, lower level)}$$

$$\sum_{k\in r_d} w_k \mathbf{z}_k^{(2)} = \sum_{k\in R_d} \mathbf{z}_k^{(2)} = \left(\sum_{k\in R_d} x_{1k}, \ldots, \sum_{k\in R_d} x_{jk}\right)' \quad \text{(MFC part, higher level)}$$

where $\mathbf{z}_k^{(1)} = (1, \hat{y}_k)',\ s_d = s \cap U_d,\ d = 1, \ldots, D$

$$\mathbf{z}_k^{(2)} = (x_{1k}, \ldots, x_{jk})',\ \ r_d = s \cap R_d,\ R_d \supset U_d$$

$$\hat{y}_k = f(\mathbf{x}_k'(\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_d)) \text{ with } \mathbf{x}_k = (1, x_{j+1,k}, \ldots, x_{pk})',\ k \in U$$

NOTE: MFC part and MC part are calibrated at different levels

Two-level hybrid calibration equation - 2

MC part: Define extended predictions and x-variables

and new z-vector:

$$\hat{y}_k^* = \hat{y}_k \text{ and } x_{0k}^* = 1,\ k \in s_d,\ 0 \text{ otherwise}$$

$$\mathbf{z}_k^* = (x_{0k}^*, \hat{y}_k^*)'$$

Minimize function

$$f^*(w) = \sum_{k \in r_d} \frac{(w_k - a_k)^2}{a_k} - \boldsymbol{\lambda}_1' \left(\sum_{k \in r_d} w_k \mathbf{z}_k^* - \sum_{k \in U_d} \mathbf{z}_k^{(1)} \right) - \boldsymbol{\lambda}_2' \left(\sum_{k \in r_d} w_k \mathbf{z}_k^{(2)} - \sum_{k \in R_d} \mathbf{z}_k^{(2)} \right)$$

Two-level HC estimator of domain total $t_d = \sum_{k \in U_d} y_k$ is given by:

$$\hat{t}_d = \sum_{k \in r_d} w_k y_k^* = \sum_{k \in s_d} w_k y_k,\ d = 1, \ldots, D$$

EXAMPLE 1: Domain totals

Simulation experiment with synthetic population

Synthetic population U with one million elements

$D = 40$ domains of varying domain sample size

Auxiliary x-variables:

x_1, x_2 continuous variables

x_c categorical variable with 5 classes

Response variable y was created by linear mixed model:

$$y_k = 1 + (1.25 + u_{d1})x_{1k} + (0.75 + u_{d2})x_{2k} + (5 + u_{d3})x_3 + u_d + \varepsilon_k,$$

$k \in U_d,\ d = 1,\ldots,40$

Random effects u follow $N(0,0.04)$, errors $\varepsilon \sim N(0,5)$

Sampling: 1000 independent SRSWOR samples

$n = 4000$ elements

Assisting models in MC and HC

Linear mixed model with domain-level random intercepts u_d

$$y_k = \mathbf{x}_k'\boldsymbol{\beta} + u_d + \varepsilon_k \quad \text{for } k \in U_d,\ d = 1,\ldots,40$$

$u_d \sim N(0,\sigma_u^2)$, $\varepsilon_k \sim N(0,\sigma^2)$, u_d and ε_k independent

Special cases :

Model 1: $\mathbf{x}_k = (1, x_{1k}, x_{2k})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$

Model 2: $\mathbf{x}_k = (1, x_{1k}, x_{2k}, x_{3k})'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$

Estimate $\boldsymbol{\beta}$ and σ_u^2 from n element sample s (by ML or REML)

Calculate estimates $\hat{u}_d$, $d = 1,\ldots,40$

Calculate fitted values $\hat{y}_k = \mathbf{x}_k'\hat{\boldsymbol{\beta}} + \hat{u}_d$, for all $k \in U_d$, $d = 1,\ldots,40$

Estimators for domain totals $t_d = \sum_{k \in U_d} y_k, \; d = 1,\ldots,40$

Design-based estimators

Direct HT estimator

$$\hat{t}_{dHT} = \sum_{k \in s_d} a_k y_k \quad \text{where } a_k = 1/\pi_k$$

Direct model-free calibration estimator

$$\hat{t}_{dMFC} = \sum_{k \in s_d} w_k^{MFC} y_k$$

Design-based model-assisted calibration estimators

Semi-direct model calibration estimator

$$\hat{t}_{dMC} = \sum_{k \in s_d} w_k^{MC} y_k$$

Semi-direct hybrid calibration estimator

$$\hat{t}_{dHC} = \sum_{k \in s_d} w_k^{HC} y_k$$

Quality measure of estimators

- **Accuracy**
  - Relative root mean squared error RRMSE (%)

$$RRMSE(\hat{t}_d) = \sqrt{\frac{1}{1000}\sum_{i=1}^{1000}(\hat{t}_d(s_i) - t_d)^2} \,/\, t_d$$

$d = 1,...,40$

where $\hat{t}_d(s_i)$ is estimate from sample s_i for domain d

t_d is known domain total

- NOTE: All methods are nearly design unbiased

Comparison scenario 1

- Accuracy comparison of design-based direct estimators and semi-direct estimators
  - HT against calibration methods
  - Model-free calibration MFC against model calibration MC
  - Effects to hybrid calibration HC?
- NOTE: Information supply
  - MFC and MC: Supply of same auxiliary information
  - HC: Supply of more auxiliary information relative to MFC and MC

Table 1 Mean relative root mean squared error (RRMSE) (%) of design-based estimators of domain totals over domain sample size classes.

Estimator	Assisting model & domain-level calibration scheme	Expected domain sample size		
		Minor 13-20	Medium 20-50	Major >50
Direct estimators				
HT	None	24.00	13.23	7.59
Model-free calibration	Calibration: $\mathbf{z}_k = (1, x_{1k}, x_{2k})'$	5.90	2.96	1.70
Semi-direct estimators				
Model: $y_k = \beta_0 + \beta_1 x_{1k} + \beta_2 x_{2k} + u_d + \varepsilon_k,\ k \in U_d,\ d = 1,\ldots,40$				
Model calibration	Calibration: $\mathbf{z}_k = (1, \hat{y}_k)'$	5.66	2.94	1.70
Hybrid calibration	Calibration: $\mathbf{z}_k = (1, x_{3k}, \hat{y}_k)'$	4.19	2.08	1.22

Conclusions for Scenario 1

- Calibration improves accuracy substantially over HT
- Under same auxiliary information supply, semi-direct model calibration MC outperforms direct model-free calibration MFC in minor domains
- Under increased information supply for MFC part of semi-direct hybrid calibration, HC outperforms MFC and MC

Comparison scenario 2

- Does the model & information supply matter in calibration?
  - Increased auxiliary information for model-free calibration MFC (added one x-variable)
  - More powerful model for model calibration MC (all three x-variables in the model)
  - Less powerful model in MC part of HC and inclusion of MFC part with a single x-variable - Effects to hybrid calibration?
- NOTE: The same auxiliary information is supplied to each estimator
  - MFC: Via the calibration x-data
  - MC: Via the model
  - HC: Via the model and the calibration x-data

Table 2 Mean RRMSE (%) of design-based calibration estimators of domain totals over domain sample size classes.

Estimator	Assisting model & domain-level calibration scheme	Expected domain sample size		
		Minor 13-20	Medium 20-50	Major >50
Direct estimator				
Model-free calibration	Calibration: $\mathbf{z}_k = (1, x_{1k}, x_{2k}, x_{3k})'$	4.27	1.97	1.16
Semi-direct estimators				
Model: $y_k = \beta_0 + \beta_1 x_{1k} + \beta_2 x_{2k} + \beta_3 x_{3k} + u_d + \varepsilon_k,\ k \in U_d,\ d = 1,\ldots,40$				
Model calibration	Calibration: $\mathbf{z}_k = (1, \hat{y}_k)'$	3.86	1.95	1.15
Model: $y_k = \beta_0 + \beta_1 x_{1k} + \beta_2 x_{2k} + u_d + \varepsilon_k,\ k \in U_d,\ d = 1,\ldots,40$				
Hybrid calibration	Calibration: $\mathbf{z}_k = (1, x_{3k}, \hat{y}_k)'$	4.19	2.08	1.22

Conclusions for Scenario 2

- Direct model-free calibration MFC with increased auxiliary information supply outperforms MFC with less information supply (ref: Table 1)
- Semi-direct model calibration MC with stronger assisting model outperforms MC with less powerful model (ref: Table 1)
- Under same information supply, MC outperforms MFC
- MC can offer a safe choice over MFC
  - protection against instability of model-free calibration due to small domain sample size and implicit model misspecification
- Hybrid calibration HC offers a realistic compromise between MFC and MC especially under coherence requirements
- Efficiency gain w.r.t MFC but loss when compared with MC

EXAMPLE 2: Poverty rate for regions

- **Design-based simulation experiment with real data**
- Fixed finite population of 1,000,000 persons
  - Western Finland
  - Register data of Statistics Finland
- Regional hierarchy: NUTS4 regions within NUTS3 regions
  - **Domains of interest:** 36 NUTS4 regions
  - **Higher level regions:** 7 NUTS3 regions
- SRSWOR sampling
  - Sample size n = 2000 persons
- 1000 independent samples drawn from the population

Variables and models

- Binary indicator variable Y with values:

  1=in poverty

  0=not in poverty

  - European Union definition, one of the AROPE indicators
  - The poverty indicator shows when a person’s equivalized income is smaller than or equal to the poverty threshold, 60% of the median equivalized income in the population

- Model
  - Logistic mixed model with domain-level random intercepts

- Auxiliary data from register (known at the unit level)
  - Equivalized income (for construction of poverty variable)
  - X-variables

    Labour force status (3 classes) for MC part

    Gender and age group (5 classes) for MFC part

Estimators

- Target parameters: At-risk-of poverty rate in domains

  $r_d = t_d / N_d$ where $t_d = \sum_{k \in U_d} y_k$ and y_k is poverty indicator

- Estimators $\hat{t}_d = \sum_{k \in s_d} w_k y_k, \ d = 1,...,36$

  where calibration weights w are method specific

  - MC part in hybrid calibration and two-level HC estimators assisted by logistic mixed model

    $$E_m(y_k | u_d) = \frac{\exp(\mathbf{x}_k' \boldsymbol{\beta} + u_d)}{1 + \exp(\mathbf{x}_k' \boldsymbol{\beta} + u_d)}, \ k \in U_d, \ d = 1,...,D$$

    where x-variables are dummy variables generated from the original categorical variables
    x-vector for model fitting in MC: $\mathbf{x}_k = (1, x_{1k}, x_{2k})'$

    x-vector in MFC: $\mathbf{x}_k = (x_{3k}, x_{4k}, x_{5k}, x_{6k}, x_{7k})'$

Table 3 Mean RRMSE (%) of design-based hybrid calibration estimators of poverty rate by domain sample size class in an experiment of 1000 simulated SRSWOR samples of size 2000 elements from a real population

MC part: X-variable in logistic mixed model: labor force status indicator
MFC part: Calibration variables: gender and age class indicators

Method	Level of calibration		Expected domain sample size			
	MC part	MFC part	Minor <25	Medium 25-50	Major >50	All
Hybrid calibration	NUTS4	NUTS4	57.8	39.1	20.1	45.3
2-level hybrid calibration	NUTS4	NUTS3	**54.2**	38.1	20.3	43.3

Conclusions for Example 2

- Two-level hybrid calibration can outperform single-level HC in accuracy for small domains in particular
- This may happen if estimation in model-free part at the domain level is unstable in single-level HC because of small domain sample size
- Calibration to higher regional level in MFC part can provide better stability because of larger domain sample sizes
- In this case, the new two-level hybrid calibration method may offer a safe choice

Basic literature 1

- **Model-free calibration**
  No model statement
  Prevailing paradigm in official statistics

  Deville J.-C. & Särndal C.-E. (1992) Calibration estimators in survey sampling. *JASA* 87, 376–382.

  Särndal C.-E. (2007) The calibration approach in survey theory and practice. *Survey Methodology* 33, 99–119.

- **Powerful computational tools**

  CBS, SCB, Statistics Canada, INSEE

  ISTAT: R package ReGenesees http://www.istat.it/it/files/2014/05/Zardetto-jos-2015-0013.pdf

- **Model-free calibration for domain estimation**

- Estevao V.M. & Särndal C.-E. (1999) The use of auxiliary information in design-based estimation for domains. *Survey Methodology* 2, 213–221.

- Estevao V.M. & Särndal C.-E. (2004) Borrowing strength is not the best technique within a wide class of design-consistent domain estimators. *JOS* 20, 645–669

- Lehtonen R. & 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 Vol. 29B. Sample Surveys. Inference and Analysis*. Amsterdam: Elsevier, 219–249.

Basic literature 2

- **Model calibration**
  Explicit model specification

  Wu C. and Sitter R.R. (2001) A model-calibration approach to using complete auxiliary information from survey data. *JASA* 96, 185–193. (with corrigenda)

  Wu C. (2003) Optimal calibration estimators in survey sampling. *Biometrika* 90, 937–9

  Kim J.K. & Park M. (2009) Calibration estimation in survey sampling

  .

Extensions

Nonparametric model calibration:
Montanari & Ranalli (2005) JASA 100

Ridge calibration:
Beaumont & Bocci (2008) METRON LXVI

- **Model calibration for domain estimation**

- Lehtonen R. and Veijanen A. (2016) Design-based methods to small area estimation and calibration approach. In: Pratesi M. (Ed.) *Analysis of Poverty Data by Small Area Estimation*. Chichester: Wiley.

- Lehtonen R. and Veijanen A. (2016) Estimation of poverty rate and quintile share ratio for domains and small areas. In: Alleva G. and Giommi A. (Eds.) *Topics in Theoretical and Applied Statistics*. New York: Springer.

- **R tools for computation are available**

Basic literature 3

- **Hybrid calibration**
  Explicit model specification

- Combination of model-free calibration and model calibration

  Lehtonen & Veijanen (2014) Small area estimation of poverty rate by model calibration and "hybrid" calibration. NORDSTAT 2014, Turku.

  Lehtonen & Veijanen (2015) Small area estimation by calibration methods. WSC 2015 of the ISI, Rio de Janeiro.

- **Extension**

- Two-level hybrid calibration

- Related papers

- Montanari G.E. and Ranalli M.G. (2009) Multiple and ridge model calibration. Proceedings of Workshop on Calibration and Estimation in Surveys 2009. Statistics Canada.

References

- Deville J.-C. and Särndal C.-E. (1992) Calibration estimators in survey sampling. *JASA* 87, 376-382.
- Estevao V. M. and Särndal C.-E. (1999) The use of auxiliary information in design-based estimation for domains. *Survey Methodology* 2, 213-221.
- 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.
- 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 Vol. 29B. Sample Surveys. Inference and Analysis*. Amsterdam: Elsevier, 219–249.
- Lehtonen R. and Veijanen A. (2012) Small area poverty estimation by model calibration*. Journal of the Indian Society of Agricultural Statistics* 66, 125-133.
- Lehtonen R. and Veijanen A. (2016a) Design-based methods to small area estimation and calibration approach. In: Pratesi M. (Ed.) *Analysis of Poverty Data by Small Area Estimation*. Chichester: Wiley.
- Lehtonen R. and Veijanen A. (2016b) Estimation of poverty rate and quintile share ratio for domains and small areas. In: Alleva G. and Giommi A. (Eds.) *Topics in Theoretical and Applied Statistics*. New York: Springer.
- Montanari G. E. and Ranalli M. G. (2005) Nonparametric model calibration estimation in survey sampling. *JASA* 100, 1429–1442.
- Montanari G.E. and Ranalli M.G. (2009) Multiple and ridge model calibration. Proceedings of Workshop on Calibration and Estimation in Surveys 2009. Statistics Canada.
- Särndal C.-E. (2007) The calibration approach in survey theory and practice. *SMJ* 33, 99–119.
- Wu C. and Sitter R.R. (2001) A model-calibration approach to using complete auxiliary information from survey data. *JASA* 96, 185–193. (with corrigenda)
- Wu C. (2003) Optimal calibration estimators in survey sampling. *Biometrika* 90, 937–9

Thank you for your attention