SMALL AREA ESTIMATION WITH CALIBRATION METHODS

Risto Lehtonen[1] & Ari Veijanen[2]

[1] University of Helsinki, P.O. Box 18, 00014 University of Helsinki. Finland, risto.lehtonen@helsinki.fi
[2] Statistics Finland, P.O. Box 00022 Statistics Finland, ari.veijanen@stat.fi

Abstract. Calibration constitutes a flexible tool for design-based estimation for finite populations. The traditional model-free calibration has been routinely used for estimation for populations and sub-populations (domains) whose sample sizes are large. Model-free calibration can be unreliable for small area estimation because of the instability of direct calibrated estimates when domain sample sizes become small. We discuss model-assisted calibration (model calibration) that incorporates explicit modelling in the calibration procedure. We show that in the cases considered, the method outperforms model-free calibration in accuracy. In model-assisted calibration, the domain sums of predictions from model are produced, instead of the domain sums of the auxiliary variables as in model-free calibration. The built-in coherence property of model-free calibration is thus lost in model-assisted calibration. For retaining the coherence property, we propose a method called "hybrid" calibration. The method combines some preferred properties of model-free calibration (coherence property of a desired set of the auxiliary variables) and model-assisted calibration (flexible modelling and efficiency improvement). As a further extension, we introduce two-level hybrid calibration for situations where the model-free and model-assisted parts of a hybrid calibration procedure are executed at different hierarchical levels of the population. This approach might be useful in cases where the model-free part indicates instability because of small domain sample sizes and calibration at a higher level is expected to offer an improvement. We discuss the various calibration methods and compare their statistical properties (bias, accuracy) and weight distributions by using design-based simulation experiments with real data obtained from statistical registers of Statistics Finland and models from the family of generalized linear mixed models (GLMM). Our real world example is in poverty estimation.

Keywords. Model-free calibration, model-assisted calibration, "hybrid" calibration, design-based simulation experiments

(Draft working paper 30 May 2018)

Résumé. Estimation sur petits domaines par calage. Le calage est un outil flexible pour réaliser des estimations dans des populations finies. Le calage ordinaire, sans modèle, a été systématiquement utilisé pour des estimations dans des populations et des sous-populations (domaines) de grandes tailles. Il a la particularité de ne pas être fiable pour des estimations dans de petits domaines car l'estimateur calé est instable lorsque la taille des échantillons tirés dans des domaines est petite. Le calage assisté par un modèle («model-assisted») intègre la modélisation explicite dans la procédure de calage. Nous montrons que dans les cas considérés, la méthode «model-assisted» surpasse le calage ordinaire au niveau de la précision des estimations. Dans le calage assisté par un modèle, les sommes des prédictions du modèle sont utilisées dans les équations de calage au lieu des sommes des variables auxiliaires comme dans le calage ordinaire. La propriété de la cohérence intégrée du calage ordinaire (qui consiste dans l'utilisation d'un ensemble désiré de variables auxiliaires) est donc perdue dans le calage assisté par un modèle. Pour maintenir cette propriété, nous proposons une méthode appelée calage «hybride». La méthode combine certaines propriétés du calage ordinaire (propriété de la cohérence intégrée) et du calage assisté par un modèle (modélisation flexible et amélioration de l'efficacité). Comme extension supplémentaire, nous introduisons deux niveaux de calage hybride pour les cas où le calage ordinaire et celui assisté par un modèle sont exécutés à différents niveaux hiérarchiques de la population. Cette approche pourrait être utile dans les cas où la partie sans modèle indique une instabilité à cause de la taille des échantillons dans des petits domaines et le calage assisté par un modèle exécuté à un niveau plus élevé devraient apporter une amélioration. Pour illustrer les forces et faiblesses des différentes méthodes de calage, nous comparons leurs propriétés statistiques (biais, précision) et les distributions correspondantes de poids de calage à l'aide d'une simulation de type «design-based»; celle-ci utilise des données réelles obtenues à partir des registres statistiques provenant de Statistics Finland et des modèles mixtes linéaires généralisés (GLMM). Notre exemple basé sur des données réelles vise l'estimation de la pauvreté.

Mots-clés: calage sans modèle, calage assisté par un modèle, calage hybride, simulation de type «design-based».

1. Introduction

Model-free calibration (Deville and Särndal 1992, Särndal 2007) has been successfully used in official statistics production all over the world for reliable design-based estimation of totals and means for populations and their subgroups (domains) whose sample sizes are large enough. Important properties of model-free calibration are the ability to reproduce the known (published) official statistics of the auxiliary variables (coherence or benchmarking property) and the fact that aggregate-level auxiliary data only are needed for the calibration procedure. Calibrated weights do not depend on any particular study variable, which allows the use of the same weight system to the desired set of study variables. These properties of practical importance are appreciated in official statistics in particular. In official statistics practice, the coherence property is often considered as the main target, and efficiency improvement constitutes a "side product" that may realize if a given study variable correlates significantly with the auxiliaries in a domain of interest. Estevao and Särndal (2004) and Lehtonen and Veijanen (2009) discuss model-free calibration in the estimation for domains and small areas.

From the modelling point of view, model-free calibration is best justified for continuous study variables under an implicit linear model. A more recent design-based calibration method called model-assisted calibration (model calibration, Wu and Sitter 2001) allows flexible modelling with e.g. generalized linear mixed models (GLMM) and nonparametric methods (Montanari and Ranalli 2005) as assisting models for various study variable types including binary, polytomous and count variables, thus extending the scope of the calibration approach over a linear modeling framework. In model-assisted calibration, predicted values are computed for every population element by using the estimated model and the auxiliary variable values that are assumed available for all population elements. A model-assisted calibration procedure produces the domain sums of predictions in the population, instead of the domain sums of the auxiliary variables as in model-free calibration. A careful model choice together with powerful auxiliary x-variables can lead to efficiency improvement relative to model-free calibration with the same set of auxiliary variables. This can happen for small domains in particular, where the direct model-free calibration estimates can become instable (e.g. Hidiroglou and Estevao 2016). Lehtonen and Veijanen (2012) developed some new variants of model-assisted calibration intended for small area estimation, including semi-direct and semi-indirect model calibration estimators that are assisted by linear and logistic mixed models. They showed that model-assisted calibration can improve efficiency over direct model-free calibration for small domains in particular.

However, the built-in coherence property of model-free calibration is lost in model-assisted calibration. The so-called hybrid calibration method (Lehtonen and Veijanen 2015) was introduced to overcome this restriction by combining some of the favourable properties of model-free calibration and model-assisted calibration into a single calibration procedure. Hybrid calibration incorporates the coherence property of model-free calibration into the calibration procedure, still retaining flexibility in model choice and efficiency gain of model-assisted calibration. A multiple model calibration method of Montanari and Ranalli (2009) involves similar goals.

In hybrid calibration, a set of auxiliary x-variables in addition to the predictions from the model are inserted in the calibration vector, constituting the model-free calibration part and the model-assisted calibration part of the vector. In our research, hybrid calibration typically outperforms model-free calibration in efficiency but does not necessarily outperform model-assisted calibration in small domains. This is the price to be paid for incorporating the coherence property in the model-assisted calibration procedure.

A variant of the method, two-level hybrid calibration (Lehtonen and Veijanen 2017), offers some protection against the possible instability problems of the model-free calibration part of a hybrid calibration procedure in small domains. In this method, model-assisted calibration operates at the original domain level but the model-free calibration part is defined at a higher hierarchical level. In domain estimation, two-level hybrid calibration can be considered as a compromise between model-free calibration and model-assisted calibration.

The paper is organized as follows. The calibration methods are introduced in Section 2. In Section 3, the statistical properties (bias, accuracy) of the methods are compared by using design-based simulation experiments with real data obtained from statistical registers of Statistics Finland. Our empirical framework is in poverty estimation. We consider calibration estimation for the at-risk-of poverty rate with logistic mixed models as assisting models. Concluding remarks are in Section 4.

2. Calibration methods

2.1 Preliminaries

Consider a finite population $U=\{1,2,...,k,...,N\}$, where *k* refers to the label of population element. A sub-population or *domain* of *U* is denoted U_d, $d=1,...,D$. The size of a domain is N_d and the size of the corresponding subset $s_d = U_d \cap s$ in sample $s \subset U$ is n_d. The domain sample sizes are not controlled by the sampling design (domains are of *unplanned* type) and thus n_d is a random variate. Design weights are $a_k = 1/\pi_k$, where π_k is inclusion probability for $k \in U$ under a given sampling design $p(\cdot)$. We assume an access to auxiliary data at the unit level; let $\mathbf{x}_k = (x_{1k},...,x_{Jk})'$ denote a known vector value for every population element $k \in U$. We usually add in the vector a value $x_{0k}=1$ for all *k*. The study variable values y_k are obtained for the sample elements $k \in s$. The sample data set and the auxiliary data set are merged at the unit level by using unique identifiers. Under this set-up, we assume a complete data set without missingness.

In our empirical example we are working with a binary y-variable indicating whether a person is in poverty or not. A logistic model formulation is then a natural choice. A mixed model formulation is often preferred over a fixed-effects model for accounting for possible differences between domains, in particular in cases where the number of domains is large. A *logistic mixed model* incorporates domain-specific random intercepts $u_d \sim N(0,\sigma_u^2)$ for domain U_d and is given by

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

where $\mathbf{x}_k = (x_{0k}, x_{1k},...,x_{Jk})'$ with $x_{0k}=1$ for all *k*, $\boldsymbol{\beta} = (\beta_0, \beta_1,...,\beta_J)'$ is a vector of fixed effects common for all domains and *m* refers to the expectation under the model. The parameters $\boldsymbol{\beta}$ and σ_u^2 are first estimated by ML and estimates $\hat{u}_d$ are calculated. Predictions $\hat{y}_k = P\{y_k = 1 \mid \hat{u}_d; \hat{\boldsymbol{\beta}}\}$ are computed for every $k \in U_d$, $d=1,...,D$.

The predictions estimated from (1) are used as auxiliary information in model-assisted calibration estimation for domain totals of the study variable *y*. In model-free calibration estimation, the original x-variables (their domain totals) are used. An ordinary model-free calibration estimator is of *direct* type as it only involves observations from the domain of interest A model-assisted calibration estimator is of *indirect* type because it uses y-data from other domains as well (later on we will refine slightly the definition). A more detailed discussion is in Lehtonen and Veijanen (2012).

The domain total of our binary study variable *y* is given by

$$t_d = \sum_{k \in U_d} y_k, \quad d=1,...,D, \qquad (2)$$

where $y_k = 1$ if a person is in poverty, and $y_k = 0$ otherwise. Poverty rate in domain *d* is:

$$r_d = \frac{t_d}{N_d}. \qquad (3)$$

2.2 Calibration estimators

Calibration in domain estimation. A calibration weighting system is introduced here by using a distance measure and a set of calibration equations. In domain estimation, *calibration equations* are given by

$$\sum_{i\in s_d} w_{di}\mathbf{z}_i = \sum_{i\in U_d}\mathbf{z}_i \ , \ d=1,...,D, \tag{4}$$

where w_{di} is calibration weight for element i in domain d and $\mathbf{z}_i$ denotes a generic *calibration vector* whose structure depends on the chosen calibration method. We employ here the distance measure approach with a chi-square distance (Deville and Särndal 1990, 1992). Using Lagrange multipliers λ we minimize:

$$\sum_{k\in s_d}\frac{(w_{dk}-a_k)^2}{a_k} - \boldsymbol{\lambda}_d'\left(\sum_{i\in s_d} w_{di}\mathbf{z}_i - \sum_{i\in U_d}\mathbf{z}_i\right) \tag{5}$$

subject to the calibration conditions (4). The equation is minimized by weights

$$w_{dk} = a_k\left(1+\boldsymbol{\lambda}_d'\mathbf{z}_k\right), \tag{6}$$

where

$$\boldsymbol{\lambda}_d' = \left(\sum_{i\in U_d}\mathbf{z}_i - \sum_{i\in s_d} a_i\mathbf{z}_i\right)'\left(\sum_{i\in s_d} a_i\mathbf{z}_i\mathbf{z}_i'\right)^{-1} \ , \ d=1,...,D. \tag{7}$$

We assume $\sum_{i\in s_d} a_i\mathbf{z}_i\mathbf{z}_i'$ be invertible. In domain estimation, the weights (6) are applied over a domain.

Model-free calibration (MFC). In classical model-free calibration (Deville and Särndal 1992), a calibration equation is imposed: the weighted sample sums of auxiliary x-variable values reproduce the known population sums (coherence or benchmarking property). Calibration vector $\mathbf{z}_i$ for (4) contains the original auxiliary x-variables; it is of the form

$$\mathbf{z}_i = \mathbf{x}_i = (x_{0i}, x_{1i},..., x_{Ji})', \ i\in U_d \ , \ d=1,...,D \tag{8}$$

where $x_{0i}=1$ for $i\in U_d$. Calibration equations (4) are given by

$$\sum_{i\in s_d} w_{di}\mathbf{z}_i = \sum_{i\in s_d} w_{di}\mathbf{x}_i = \sum_{i\in U_d}\mathbf{x}_i = \left(\sum_{i\in U_d} x_{0i}, \sum_{i\in U_d} x_{1i},..., \sum_{i\in U_d} x_{Ji}\right)' \quad , \ d=1,...,D. \tag{9}$$

We minimize (5) subject to (9) and obtain MFC estimator of domain total t_d of the form

$$\hat{t}_{dMFC} = \sum_{k\in s_d} w_{dk} y_k \ , \ d=1,...,D, \tag{10}$$

where the weights w_{dk} are computed by (6) and (7) with (8). The estimator (10) is of direct type because it operates with y-data from the given domain only. Under the setting described in 2.1, it can happen that a small number of sample elements only fall in some domains, possibly causing instability problems for the MFC estimator.

Model-assisted calibration (MC). Model-assisted calibration for domain estimation is intended to extend the scope of model-free calibration by introducing flexible modelling in the calibration procedure (Lehtonen and Veijanen 2012). The method also aims to overcome the possible instability problems of MFC in small domains. The MC procedure involves two steps: the modelling step and the calibration step. In the modelling step, the chosen model for *y* is fitted to the sample data, and predictions $\hat{y}_k$ are computed for every $k \in U_d$, $d = 1,...,D$. In the calibration step, the predictions are incorporated in the calibration z-vector and calibration weights are determined by inserting the z-vector into formulas (4) to (7).

Model-assisted calibration can be defined at the population level, at the domain level or at an intermediate level, for example at a regional level (neighbourhood) that contains the domain of interest. In a "semi-direct" approach, predictions for the domain of interest only are used in the calibration step, whereas in a "semi-indirect" approach, predictions outside the domain of interest also are included. For both variants, the modelling step uses all available sample data. More detailed discussion is in Lehtonen and Veijanen (2012).

We discuss here the semi-direct method. Calibration vector $\mathbf{z}_i$ for (4) is:

$$\mathbf{z}_i = (x_{0i}, \hat{y}_i)', \ i \in U_d , \tag{11}$$

where $x_{0i} = 1$ for $i \in U_d$. Fitted values in (11) are $\hat{y}_k = f(\mathbf{x}_k'(\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_d))$ with $\mathbf{x}_k = (x_{0k}, x_{1k},..., x_{Jk})'$, $k \in U_d$. In poverty rate estimation, our study variable is binary and the fitted values are computed by using model (1). We obtain

$$\hat{y}_k = \frac{\exp(\mathbf{x}_k'\hat{\boldsymbol{\beta}} + \hat{u}_d)}{1 + \exp(\mathbf{x}_k'\hat{\boldsymbol{\beta}} + \hat{u}_d)}, \quad k \in U_d, \ d = 1,...,D. \tag{12}$$

Calibration equations are given by

$$\sum_{i \in s_d} w_{di}\mathbf{z}_i = \sum_{i \in U_d} \mathbf{z}_i = \left(\sum_{i \in U_d} x_{0i}, \sum_{i \in U_d} \hat{y}_i \right), \ d = 1,...,D. \tag{13}$$

We minimize (5) subject to (13) and obtain a model-assisted calibration estimator of (2) given by

$$\hat{t}_{dMC} = \sum_{k \in s_d} w_{dk} y_k , \ d = 1,...,D, \tag{14}$$

where the weights w_{dk} are computed by (6) and (7) with (11).

We call (14) semi-direct, as calibration is defined at the domain level but the weights are determined by a model that is fitted to the whole sample.

Hybrid calibration (HC). The coherence property (9) of model-free calibration for auxiliary data vector $\mathbf{x}_k = (x_{0k}, x_{1k},..., x_{Jk})'$ is lost in model-assisted calibration. In hybrid calibration (Lehtonen and Veijanen 2015), we impose the coherence property for a chosen subset of x-variables (the MFC part) and retain the MC calibration property (11) for another subset (the MC part). Hybrid calibration is discussed here for semi-direct MC.

Calibration vector for (4) is:

$$\mathbf{z}_i = (x_{0i}, \hat{y}_i, x_{1i},..., x_{ji})', \ i \in U_d , \ d = 1,...,D, \tag{15}$$

and calibration equations are given by

$$\sum_{i \in s_d} w_{di}\mathbf{z}_i = \sum_{i \in U_d} \mathbf{z}_i = \left(\sum_{i \in U_d} x_{0i}, \sum_{i \in U_d} \hat{y}_i, \sum_{i \in U_d} x_{1i},..., \sum_{i \in U_d} x_{ji} \right)', \tag{16}$$

where again $x_{0i} = 1$ for $i \in U_d$. We minimize (5) subject to (16) and obtain hybrid calibration estimator of (2) given by

$$\hat{t}_{dHC} = \sum_{k \in s_d} w_{dk} y_k \,,\ d = 1,...,D\,, \tag{17}$$

where w_{dk} are computed by (6) and (7) with (15). The fitted values for MC part are $\hat{y}_k = f(\mathbf{x}_k'(\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_d))$ with $\mathbf{x}_k = (x_{0k}, x_{j+1,k}, ..., x_{Jk})'$, $k \in U_d$, where the assisting model is (1) with the $\mathbf{x}_k$ above. Note that actually, we have split the original x-vector into two non-overlapping subsets: MFC part with $\mathbf{x}_k^{(1)} = (x_{1k}, ..., x_{jk})'$ and MC part with $\mathbf{x}_k^{(2)} = (x_{0k}, x_{j+1,k}, ..., x_{Jk})'$. It also is possible to make up overlapping decomposition of the x-vector.

Two-level hybrid calibration (HC2). The MFC part of hybrid calibration can involve instability for domains whose sample sizes are small. Two-level hybrid calibration (Lehtonen and Veijanen 2017) is intended to reduce the effects of the possible instability but still retain the option for efficiency improvement. The idea is to let the model-assisted calibration part to operate at the original domain level (e.g. NUTS4) and to define the model-free calibration part at a higher hierarchical level (e.g. NUTS3), where instability problems are hopefully not met.

In two-level hybrid calibration, we define two sets of calibration equations to be solved simultaneously:

$$\sum_{i \in r_d} w_{ri} \mathbf{z}_i^{(1)} = \sum_{i \in U_d} \mathbf{z}_i^{(1)} = \left(\sum_{i \in U_d} x_{0i}, \sum_{i \in U_d} \hat{y}_i \right)' \quad \text{MC part (lower level)} \tag{18}$$

$$\sum_{i \in r_d} w_{ri} \mathbf{z}_i^{(2)} = \sum_{i \in R_d} \mathbf{z}_i^{(2)} = \left(\sum_{i \in R_d} x_{1i}, ..., \sum_{i \in R_d} x_{ji} \right)' \quad \text{MFC part (higher level)} \tag{19}$$

where

$\mathbf{z}_i^{(1)} = (x_{0i}^{(1)}, \hat{y}_i^{(1)})'$, $r_d = s \cap R_d$, $R_d \supset U_d$, $d = 1,...,D$ (auxiliary data vector for MC part)

$x_{0i}^{(1)} = 1$, $i \in U_d$, 0 otherwise, (extended x_0-variable)

$\hat{y}_i^{(1)} = \hat{y}_i$, $i \in U_d$, 0 otherwise (extended predictions)

$\hat{y}_i = f(\mathbf{x}_i'(\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_d))$ (logistic mixed model)

$\mathbf{x}_i = (x_{0i}, x_{j+1,i}, ..., x_{Ji})'$, $k \in U_d$ (x-data vector for logistic model)

$\mathbf{z}_i^{(2)} = (x_{1i}, ..., x_{ji})'$ (auxiliary x-data vector for MFC part)

Using Lagrange multipliers λ we minimize:

$$\sum_{k \in r_d} \frac{(w_{rk} - a_k)^2}{a_k} - \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{pmatrix}' \left(\sum_{i \in r_d} w_{ri} \begin{pmatrix} \mathbf{z}_i^{(1)} \\ \mathbf{z}_i^{(2)} \end{pmatrix} - \begin{pmatrix} \sum_{i \in R_d} \mathbf{z}_i^{(1)} \\ \sum_{i \in R_d} \mathbf{z}_i^{(2)} \end{pmatrix} \right) \tag{20}$$

subject to calibration constraints (18) and (19). Writing $\boldsymbol{\lambda}_r' = (\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2')$ and $\mathbf{z}_k = (\mathbf{z}_i^{(1)}, \mathbf{z}_i^{(2)})'$, equation (20) is minimized by weights

$$w_{rk} = a_k (1 + \boldsymbol{\lambda}_r' \mathbf{z}_k),$$

where

$$\boldsymbol{\lambda}_r' = \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{pmatrix}' = \left(\sum_{i \in R_d} \begin{pmatrix} \mathbf{z}_i^{(1)} \\ \mathbf{z}_i^{(2)} \end{pmatrix} - \sum_{i \in r_d} a_i \begin{pmatrix} \mathbf{z}_i^{(1)} \\ \mathbf{z}_i^{(2)} \end{pmatrix} \right)' \left(\sum_{i \in r_d} a_i \begin{pmatrix} \mathbf{z}_i^{(1)} \\ \mathbf{z}_i^{(2)} \end{pmatrix} \begin{pmatrix} \mathbf{z}_i^{(1)} \\ \mathbf{z}_i^{(2)} \end{pmatrix}' \right)^{-1}. \tag{21}$$

The resulting two-level HC estimator of (2) given by

$$\hat{t}_{dHC2} = \sum_{k \in s_d} w_{rk} y_k \,,\ d = 1,...,D \,. \qquad (22)$$

3. Simulation experiments

Accuracy comparison of calibration estimators. The binary poverty indicator shows when a person's equivalized income is smaller than or equal to the poverty threshold, 60% of the median equivalized income M in the population. The indicator for sample person k is defined as $y_k = I\{y_k \le 0.6\hat{M}\}$, where $y_k = 1$ if a person is in poverty and 0 otherwise. The quantity $0.6\hat{M}$ is the estimated poverty threshold, where $\hat{M}$ is estimated by HT from the estimated distribution function of equivalized income in the population (see details in Lehtonen and Veijanen 2012). The binary poverty indicator *y* acts as the study variable in the calibration exercise.

For design-based simulation experiments, an adult population of about 600,000 persons was constructed from real income data of Statistics Finland, containing seven NUTS level 3 regions and 36 NUTS level 4 regions in Western Finland. In addition to the equalized income variable, our population contained three auxiliary variables: gender, three-category age and three-category labor force status. We created indicators for each class of a qualitative variable. The complete auxiliary x-vector for $k \in U$ thus is $\mathbf{x}_k = (x_{0k}, x_{1k}, x_{2k}, x_{3k}, x_{4k}, x_5)'$ and was used in MFC and MC. In HC and HC2, $\mathbf{x}_k$ was divided into two subsets: $\mathbf{x}_k^{(1)} = (x_{1k}, x_{2k})'$ was used for the MFC part and $\mathbf{x}_k^{(2)} = (x_{0k}, x_{3k}, x_{4k}, x_{5k})'$ was used in the MC part. The x-variables showed some explanatory power: in logistic mixed models, the complete x-data explained about 15% of the variation of *y*.

As domains of interest we used the D = 36 NUTS4 regions, and the 7 NUTS3 regions were used as the higher level regions in two-level HC. The NUTS classification is hierarchical: each NUTS4 region is contained within a larger NUTS3 region. Sample sizes in NUTS3 regions obviously are larger than in NUTS4 regions, which is beneficial for two-level hybrid calibration.

In the simulations, K = 1000 samples of n = 2000 units were drawn with simple random sampling without replacement (SRSWOR) from the unit-level population. Design bias and accuracy were measured by absolute relative bias (ARB) and relative root mean squared error (RRMSE):

$$ARB(\hat{\theta}_d) = \left| (1/K)\sum_{j=1}^{K}(\hat{\theta}_{dj} - \theta_d) \right| / \theta_d \ \text{ and } \ RRMSE(\hat{\theta}_d) = \sqrt{(1/K)\sum_{j=1}^{K}(\hat{\theta}_{dj} - \theta_d)^2} \,/\, \theta_d \,.$$

Poverty rate (3) for domain d was estimated by $\hat{r}_d = \hat{t}_d / N_d$, $d = 1,...,D$, where $\hat{t}_d$ is obtained by MFC (10), MC (14), HC (17) and HC2 (22). We present the averages of RRMSE figures over three domain classes defined by expected domain sample size. ARB figures are not presented because all methods appeared nearly design unbiased (results are available from authors). Our logistic mixed models contained regional random intercepts associated with NUTS4 regions.

Several conclusions can be drawn from Table 1. Over all domain sample size classes, semi-direct model-assisted calibration that incorporates the auxiliary information in the calibration procedure by a model indicates best accuracy, and direct model-free calibration was the worst. This holds for all domain sample sizes; the difference in accuracy declines with increasing domain sample size, as expected. The logistic mixed model in MC thus tends to improve accuracy over MFC, whose implicit assisting model is a linear fixed-effects model fitted separately in each domain. The figures also indicate that MFC can suffer from instability in the smallest domains.

When incorporating by HC the coherence property for a subset of x-variables, we lose accuracy relative to MC, but still HC outperforms MFC in all domain classes. Because of the MFC part, HC might suffer from instability in small domains and we try to clean this out by two-level HC. This appears successful and the improvement is best visible in the group of smallest domains, where the accuracy of two-level hybrid calibration is close to that of model-assisted calibration.

Table 1. Average relative root mean squared error (RRMSE) (%) of calibration estimators of domain poverty rate in three domain sample size classes in a simulation experiment of 1000 SRSWOR samples of 2000 units.

Estimator (formula number in parentheses)	Assisting model & calibration scheme		Expected domain sample size: Minor <25 10 domains	Medium 25-50 16 domains	Major >50 10 domains	All 36 domains
Direct estimator						
Model-free calibration (10)	NUTS4	$\mathbf{z}_k = (1, x_{1k}, x_{2k}, x_{3k}, x_{4k}, x_{5k})'$	61.1	40.4	20.1	47.3
Semi-direct estimators						
Model: $E_m(y_k \| u_d) = \exp(\mathbf{x}_k'\boldsymbol{\beta} + u_d)/(1+\exp(\mathbf{x}_k'\boldsymbol{\beta} + u_d))$, $k \in U_d$, $\mathbf{x}_k = (1, x_{1k}, x_{2k}, x_{3k}, x_{4k}, x_{5k})'$						
Model calibration (13)	NUTS4	$\mathbf{z}_k = (1, \hat{y}_k)'$	54.1	37.6	19.8	43.0
Model: $E_m(y_k \| u_d) = \exp(\mathbf{x}_k'\boldsymbol{\beta} + u_d)/(1+\exp(\mathbf{x}_k'\boldsymbol{\beta} + u_d))$, $k \in U_d$, $\mathbf{x}_k = (1, x_{3k}, x_{4k}, x_{5k})'$						
Hybrid calibration (17)	NUTS4	$\mathbf{z}_k = (1, \hat{y}_k, x_{1k}, x_{2k})'$	58.0	39.1	20.1	45.4
Two-level hybrid calibration (22)	NUTS4	$\mathbf{z}_k^{(1)} = (1, \hat{y}_k)'$	54.2	38.1	20.2	43.3
	NUTS3	$\mathbf{z}_k^{(2)} = (x_{1k}, x_{2k})'$				

Distribution of weights. It is well known that the calibration approach chosen here can involve large variation of weights and negative weights that are often considered unfeasible in practical applications. Negative weights are expected in small domains in particular. We were interested in the following question: to what extent the semi-direct calibration methods can improve the weight distribution relative to that of the model-free calibration?

We executed a small simulation experiment for weight distribution with K = 100 SRSWOR samples from our population and computed the maximum interdecile range of calibrated weights for each sample. Medians of the statistic in domain size classes are in Table 2.

Table 2. Median of maximum interdecile range of calibrated weights in three domain sample size classes in a simulation experiment of 100 SRSWOR samples of 2000 units.

Method	Expected domain sample size: Minor <25	Medium 25-50	Major >50
Model-free calibration MFC	1620	673	210
Model assisted calibration MC	984	418	172
Hybrid calibration HC	1415	665	245
Two-level hybrid calibration HC2	780	430	214

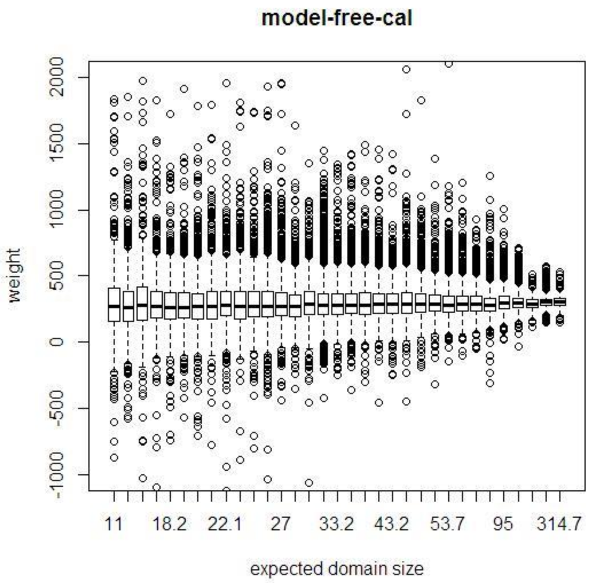
It is clearly seen that MFC and HC suffer from instability in the small domains group and MC and HC2 behave considerably better. Instability vanishes with increasing domain sample size.

More detailed light to weight distributions is thrown by Figure 1. It contains four panels, one for each calibration method. The upper left panel is for model-free calibration, the upper right is for model calibration, the lower left is for hybrid calibration and the lower right is for two-level hybrid calibration. The horizontal axis in each graph indicates the expected domain sample size, from smallest to largest, and the value of weight is on the vertical axis, varying from -1000 to 2000.

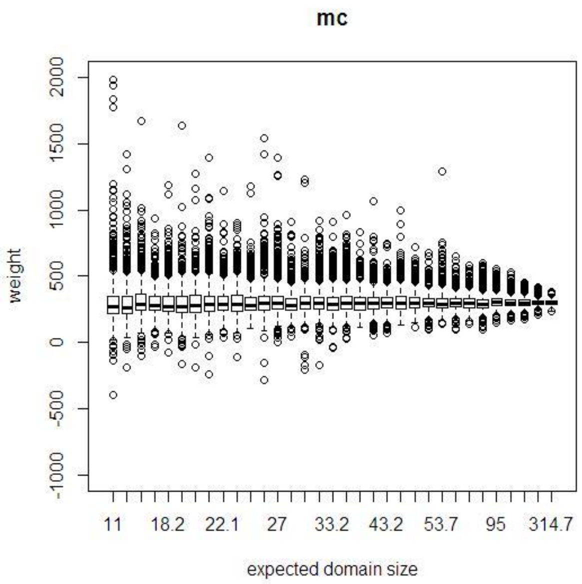
For each of the 36 domains, a graph shows a box plot containing the average weight and the distribution of the 100 weight values. The averages of weight values for each domain are pretty equal as expected. The variation of weights is clearly largest for model-free calibration. In this method, the presence of large positive and negative weights increases dramatically with decreasing domain sample size, and decreases only slowly when domain sample size increases. Model-assisted calibration shows best performance: the variation of weights is much smaller than in MFC and the variation decreases rapidly when domain sample size increases. Two-level hybrid calibration shows as good performance as MC, indicating successful stabilization of the weight distribution.

It is obvious that in the cases considered, semi-direct calibration can decrease the variation of weights and improve the distribution of weights, and can perform better relative to direct model-free calibration. Of the methods, model-assisted calibration and two-level hybrid calibration indicate best performance.

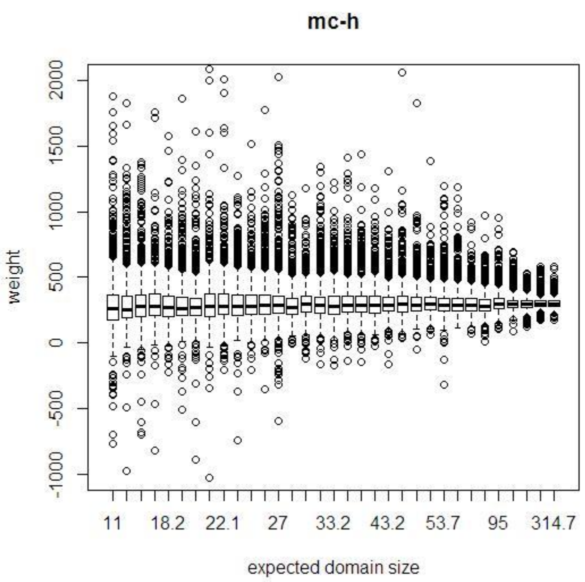
Panel 1. Model-free calibration MFC



Panel 2. Model-assisted calibration MC



Panel 3. Single-level hybrid calibration HC



Panel 4. Two-level hybrid calibration HC2

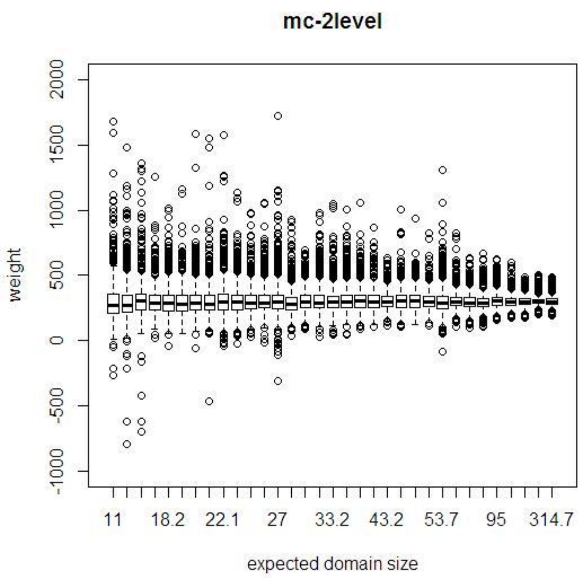


Figure 1. Distribution of calibrated weights by expected domain sample size in a simulation experiment of 100 SRSWOR samples of 2000 units.

Concluding remarks

We introduced three design-based calibration methods for domain and small area estimation that incorporated modelling as a part of the calibration procedure. Ordinary model-free calibration was presented as a reference method. In our simulation experiments, all methods were nearly design unbiased for the parameters of interest. The methods indicated nearly equal accuracy for domains whose sample sizes were large enough. In small domain estimation, accuracy properties differed considerably. The ordinary model-free calibration method of direct type failed in the smallest domains because of the instability and large variation of weights. The semi-direct methods that incorporated modelling provided clear improvement. Model-assisted calibration outperformed the other methods in accuracy, notably in small domains. Hybrid calibration improved accuracy over model-free calibration, but not much. Two-level hybrid calibration behaved better and may offer a compromise in situations where coherence properties are required for some x-variables and the possibility of efficiency improvement via flexible modelling is still desired.

We did not incorporate weight restriction in order to avoid subjective elements in the methods. Under the chosen calibration framework, weight distributions differed between methods and large negative and positive weights appeared, for the model-free method in particular. The distributions were improved considerably for model-assisted calibration and two-level hybrid calibration.

References

Deville, J.-C. and Särndal, C.-E. (1990) Calibration estimators and generalized raking techniques in survey sampling. Statistics Sweden. *R & D Report* 1990:1.

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

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

Hidiroglou, M. A. and Estevao, V. M. (2016) A comparison of small area and calibration estimators via simulation. *Joint Issue of Statistics in Transition and Survey Methodology*,17, 133–154.

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

Lehtonen, R. and Veijanen, A. (2017) A two-level hybrid calibration technique for small area estimation. SAE2017 Conference, Paris, June 2017.

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. *Survey Methodology*, 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.