Universität Trier

A Generalized Calibration Approach Ensuring Coherent Estimates with Small Area Constraints

Jan Pablo Burgard
Ralf Münnich
Martin Rupp

Research Papers in Economics
No. 10/19

A generalized calibration approach ensuring coherent estimates with small area constraints

Jan Pablo Burgard,* Ralf Münnich,* and Martin Rupp*

Economic and Social Statistics, Trier University, 54286 Trier

Abstract

Within this article, a generalized calibration approach is presented, which provides coherent and efficient estimates considering a high number of constraints on different hierarchical levels. These constraints may be obtained from different sources such as survey data, register data, administrative data, or even other sources like big data derived using different estimation approaches, including small area techniques on different levels of interest. In order to incorporate a possible heterogeneous quality and the multitude of the constraints, a relaxation of selected constraints is proposed. In that regard, predefined tolerances are assigned to hardly achievable constraints, mostly at low aggregation levels, or sample estimates with non-negligible variances. In addition, the presented generalized calibration approach allows the use of box-constraints for the calibration weights in order to avoid an inappropriate high variation of the resulting weights. Furthermore, various penalty functions are presented in order to accommodate particular circumstances in applications. The proposed iterative algorithm provably finds the optimal solution and the numerical implementation is able to deal with a huge data base such as the set of all households in Germany. The performance is demonstrated in a short simulation study.

Keywords: Calibration, general regression estimator, coherent estimates, sampling weights, soft constraints, box-constraints, semismooth Newton

*email: {burgardj,muennich,ruppm}@uni-trier.de

1 Introduction

Survey data are often accompanied by a weight vector to provide the necessary base for design-unbiased estimates. One important property, not solely for official statistics, that these weights have to fulfil is coherence. The European Statistics Code of Practice urges coherence in principle 14 (see Eurostat, 2011). One straight forward method to construct coherent weights is the application of calibration methods.

The general idea of calibration, as introduced in Deville and Särndal (1992), is to modify the design weights to a minimal extent in terms of a penalty function such that weighted estimates using the calibration variables satisfy the according known totals (cf. also Särndal, 2007, Kott, 2006, Statistics Canada, 2003, pp. 45-46, or Merkouris, 2004). A recent overview can also be drawn from Haziza and Beaumont (2017). In practice, however, the classical calibration problem may suffer from different uncertainties such as using auxiliary data from different sources, register data with unknown errors or inaccurate estimates. Further, including too many constraints may result in extremely large weights with unacceptable variation. The spread of weights including its importance for statistical modeling was discussed in Gelman (2007) and in the context of survey and small area estimation in Münnich and Burgard (2012).

Since the auxiliary data are treated as known values in general, inaccuracies may hand over to the calibration weights and the respective estimates. Mostly, this cannot be avoided in classical calibration methods. However, there are several methods to consider this fact. Chambers (1996) uses ridge calibration techniques, which is similarly applied in Rao and Singh (1997), Beaumont and Bocci (2008), and Montanari and Ranalli (2009). To further constrain the variation of weights, Théberge (2000) proposed the use of box-constraints. The use of predicts from models within the calibration was proposed by Wu and Sitter (2001) which was denoted by model calibration. Following this, Chen et al. (2002) analyzed box-constraints in a model-calibration environment by computing empirical likelihood estimators and model-calibrated empirical likelihood estimators. Montanari and Ranalli (2005) extend the approach of Wu and Sitter (2001) using non-parametric regression models. However, since model calibration may not satisfy the hard constraints for survey estimates, Lehtonen and Veijanen (2015) proposed hybrid calibration with covers both classical and model calibration.

In this article, we propose a generalized calibration method. The aim of the method is to achieve coherency between estimates gained from different sources and to allow for a high flexibility in the choice of the auxiliary data. The inclusion of subtotals in terms of small area constraints is also possible. In order to avoid empty solution spaces or an inappropriate variation of weights, we include box-constraints as well as the opportunity to relax selected calibration constraints. Then, these only have to be satisfied within specific predefined tolerances. This ensures the feasibility of the calibration problem even for large problem instances and a large number of benchmark totals. The maximal tolerances allowed may also be restricted using additional box-constraints. Finally, a numerical algorithm is proposed that satisfies possible non-differentiabilities introduced by the box-constraints where classical Newton-methods may tend to fail providing the optimum.

The article is structured as follows. In Section 2, the calibration framework is provided. After presenting the classical calibration methods, a general intension is proposed which contains soft and box-constraints. Additionally, a computation algorithms is presented that allows finding the optimum under non-differentiability induced by the box-constraints. A demonstration of the general calibration approach is finally presented in Section 3, which considers national as well as regional estimates. Section 4 contains some brief concluding remarks.

2 The generalized calibration method

2.1 General framework

We consider a finite population $\mathcal{U} = \{1, \ldots, N\}$ and a sample $S \subseteq \mathcal{U}$ of size $n \leq N$. The design weights denoted with d_k are known and strictly positive for each element $k \in U$ of the population. The major aim is to estimate the population total $\tau_y = \sum_{k \in \mathcal{U}} y_k$ of variable of interest y using the calibration estimator $\hat{\tau}_y^{\mathrm{CAL}} = \sum_{k \in S} d_k g_k y_k$ with correction weights g_k $(k \in S)$. The correction weights g_k are the result of the calibration procedure under consideration of the the calibration benchmarks $\tau_{x_i} = \sum_{k \in S} d_k g_k x_{ik}$ $(i = 1, \ldots, q)$ regarding q auxiliary variables with the individual auxiliary values $x_k := (x_{1k}, \ldots, x_{qk})^T \in \mathbb{R}^q$ for all units of the sample. Then, the standard calibration approach is given by

$$\begin{aligned} \min_{g \in \mathbb{R}^n} \; & \sum_{k \in S} d_k D(g_k) \\ \text{s.t.} \; & \sum_{k \in S} d_k g_k x_{ik} = \tau_{x_i} \;\; \forall i = 1, \ldots, q, \end{aligned} \tag{1}$$

where the objective function is characterized by a predefined distance function D. The most common distance function is the GREG-type distance function, since the calibration estimator for the population total based thereon is equivalent to the GREG estimator (cf. Deville and Särndal, 1992). Some traditional choices for D are shown in Table 1 and plotted in Figure 1. For other distance functions we refer to Deville and Särndal (1992), Deville and Särndal (1993), Singh and Mohl (1996), and Stukel et al. (1996). In this article, we focus

Table 1: Common examples of distance functions for the calibration estimator.

	$D(g_k)$
GREG-type	$\frac{1}{2}(g_k - 1)^2$
Raking Ratio	$g_k \log(g_k) - g_k + 1$
Maximum-likelihood Raking	$g_k - 1 - \log(g_k)$

on the GREG-type, the Raking, and the ML-Raking distance function. These functions differ in their treatment of the penalty term, which affects the outcome of the objective function depending on how greatly the calibration weight $w_k := d_k g_k$ differs from the design weight d_k (i.e. the correction weight g_k differs from 1.0). The GREG-type distance function assigns the same penalty to values of g_k with an equal absolute distance to 1.0. The Raking Ratio and the ML-Raking distance functions are based on a nonlinear dependency, where smaller weights are penalized stronger compared to weights greater than 1.0.

In order to limit the spread of the calibration weights Deville and Särndal (1993) and Münnich et al. (2012b) added box-constraints $0 \leq L_{g_k} \leq g_k \leq U_{g_k}$ with $L_{g_k} < U_{g_k}$ for each correction weight g_k. Theses box-constraints are also known as *range-restricted weights* in the literature. For this so-called box-constraint calibration method a very efficient algorithm is published in Münnich et al. (2012b) and Wagner (2013) based on a semismooth Newton method.

As described in the introduction, various sources of the auxiliary data, different stratification levels, and different quantity of auxiliary data may lead to infeasibility issues of the calibration problem. To counteract this, some benchmarks may have to be relaxed, i.e. the restrictions are weaken. As a results of the usage to these so called *soft constraints*, the respective benchmarks only have to be fulfilled with a specific predefined perturbation. The maximal tolerances allowed are added to the problem via additional box-constraints. Since benchmarks possibly on different stratification levels are obtained from known totals and different estimates gained by direct or small-area estimators (cf. Rao, 2003, Münnich et al., 2013, and Tzavidis et al., 2018), the relaxation may also prevent coherence problems between the estimation levels. Thus, an individual adjustment of the tolerance per benchmark can facilitate to incorporate different confidence measures for the different benchmarks. As a consequence, the confidence in a small area estimate is comparably low and the allowed tolerance for this benchmark should be higher than for other benchmarks.

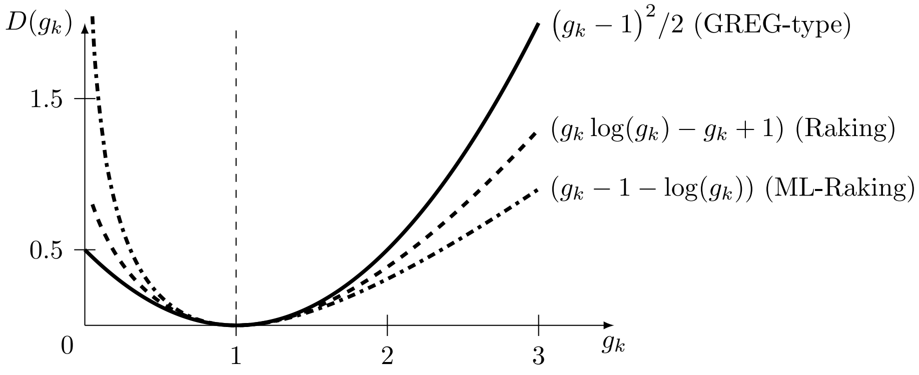

Figure 1: Common examples of distance functions for the calibration estimator.

2.2 Derivation of the method

By extending the calibration approach in (1), the relaxation and box-constraints are treated as real restrictions. This approach amongst others relies on the developments of Wagner (2013, Chapter 7) and is similarly presented in Rupp (2018, Chapter 5). By applying a relaxation and box-constraints, the calibration problem (1) can be extended to

$$\begin{aligned} \min_{(g,\epsilon)\in\mathbb{R}^{n+q_2}} \quad & \sum_{k\in S} d_k D(g_k) + \sum_{j=1}^{q_2} \delta_j D(\epsilon_j) \\ \text{s.t.} \quad & \sum_{k\in S} d_k g_k x_{ik}^{\text{ex}} = \tau_{x_i^{\text{ex}}} \; \forall i = 1,\dots,q_1 \\ & \sum_{k\in S} d_k g_k x_{jk}^{\text{rel}} = \epsilon_j \tau_{x_j^{\text{rel}}} \; \forall j = 1,\dots,q_2 \\ & L_{g_k} \leq g_k \leq U_{g_k} \; \forall k = 1,\dots,n \\ & L_{\epsilon_j} \leq \epsilon_j \leq U_{\epsilon_j} \; \forall j = 1,\dots,q_2 \end{aligned} \tag{2}$$

with a distance function D defined in Table 1 and the total number of $q = q_1 + q_2$ equality constraints. The auxiliary variables are denoted with $x_{1k}^{\text{ex}}, \dots x_{q_1k}^{\text{ex}} \in \mathbb{R}$ and $x_{1k}^{\text{rel}}, \dots x_{q_2k}^{\text{rel}} \in \mathbb{R}$ $(k \in S)$ for the restrictions which have to be fulfilled exactly and with a certain degree of precision, respectively. The degree of precision is defined by the box-constraints of $\epsilon_j \in \mathbb{R}_+$, denoted by L_{ϵ_j} and U_{ϵ_j} $(j = 1,\dots,q_2)$. The vector $\delta \in \mathbb{R}_+^{q_2}$ used as factors in the objective function determines the magnitude of penalization within these bounds. The box-constraints for the correction weights g_k are denoted with $0 \leq L_{g_k} \leq g_k \leq U_{g_k}$ $(k \in S)$. As described before, benchmark totals $\tau_{x_i^{\text{ex}}}$ $(i = 1,\dots,q_1)$ and $\tau_{x_j^{\text{rel}}}$ $(j = 1,\dots,q_2)$ may also be estimated totals instead of known totals. Due to simplifications, the common *hat*-notation is omitted.

In the following, the matrix X^{ex} is defined as design weighted auxiliary matrix for the q_1 auxiliary variables $x_{1k}^{\text{ex}}, \dots, x_{q_1k}^{\text{ex}} \in \mathbb{R}$ for all units $k \in S$, whose benchmark totals need to be satisfied exactly. Analogously, the matrix X^{rel} contains the q_2 auxiliary variables $x_{1k}^{\text{rel}}, \dots, x_{q_2k}^{\text{rel}} \in \mathbb{R}$, whose benchmarks have only to be fulfilled within a predefined tolerance:

$$X^{\text{ex}} = \begin{bmatrix} d_1 x_{11}^{\text{ex}} & \dots & d_n x_{1n}^{\text{ex}} \\ \vdots & & \vdots \\ d_1 x_{q_11}^{\text{ex}} & \dots & d_n x_{q_1n}^{\text{ex}} \end{bmatrix} \in \mathbb{R}^{q_1\times n} \quad \text{and} \quad X^{\text{rel}} = \begin{bmatrix} d_1 x_{11}^{\text{rel}} & \dots & d_n x_{1n}^{\text{rel}} \\ \vdots & & \vdots \\ d_1 x_{q_21}^{\text{rel}} & \dots & d_n x_{q_2n}^{\text{rel}} \end{bmatrix} \in \mathbb{R}^{q_2\times n}.$$

The benchmark totals are denoted by $\tau_{x_1^{\text{ex}}}, \dots, \tau_{x_{q_1}^{\text{ex}}} \in \mathbb{R}$ and $\tau_{x_1^{\text{rel}}}, \dots, \tau_{x_{q_2}^{\text{rel}}} \in \mathbb{R}$, respectively. The approved perturbations of the relaxed variables are given by $\epsilon_1, \dots, \epsilon_{q_2} \in \mathbb{R}_+$. In the notation considered here, the matrices X^{ex} and X^{rel} correspond to population total benchmarks. If regional benchmarks are added for auxiliary variable $i \in \{1, \dots, q_1\}$, the i^{th} row of X^{ex} is extended. The amount of the constraints considered q_1 is then updated. This procedure can be done consecutively for several variables or stratification levels and is analogously possible for relaxed auxiliary variables. In addition, it is also valid to assume an auxiliary variable with totals that have to be fulfilled exactly on highly aggregated stratification levels, but the totals may be relaxed on more disaggregated levels. This procedure is common in applications as, in general, the variance of estimates on low aggregation levels is higher than on higher aggregation levels.

Using the above notation, the restriction matrix of problem (2) can be formulated as

$$A := \left[\begin{array}{c|ccc} X^{\text{ex}} & 0 & \dots & 0 \\ \hline & -\tau_{x_1^{\text{rel}}} & & 0 \\ X^{\text{rel}} & & \ddots & \\ & 0 & & -\tau_{x_{q_2}^{\text{rel}}} \end{array}\right] \in \mathbb{R}^{(q_1+q_2)\times(n+q_2)}, \tag{3}$$

where q_1 is the number of benchmarks to be fulfilled exactly and q_2 is the number of benchmarks to be fulfilled with a tolerance. Whereas the totals for the relaxed benchmarks are included in the low right block of the matrix A, the benchmarks for the exact benchmarks are included in the right-hand side vector given by

$$b := \left(\tau_{x_1^{\text{ex}}}, \dots, \tau_{x_{q_1}^{\text{ex}}}, 0, \dots, 0\right)^T \in \mathbb{R}^{q_1+q_2}. \tag{4}$$

To measure the deviations of the design weights d_k from the calibration weights w_k and the perturbations ϵ_j to 1.0, we choose one of the distance functions $D : \mathbb{R}_+ \to \mathbb{R}_{0_+}$ presented in Table 1, i.e.

1. GREG-type: $D(z_\kappa) = \frac{1}{2}(z_\kappa - 1)^2$,
2. Raking Ratio: $D(z_\kappa) = z_\kappa \log(z_\kappa) - z_\kappa + 1$, or
3. ML-Raking: $D(z_\kappa) = z_\kappa - 1 - \log(z_\kappa)$.

In that regard, $\kappa = 1, \dots, n + q_2$ is the composed index for the respective component of the objective function of problem (2), i.e. indices $\kappa \leq n$ correspond to the n sampled units (index k) and indices $\kappa > n$ correspond to one of the q_2 relaxed benchmarks (index j). Then, problem (2) can then be equivalently rewritten as

$$\begin{aligned} \min_{z\in\mathbb{R}^{n+q_2}} \quad & P(z) := \sum_{\kappa=1}^{n+q_2} \tilde{d}_\kappa D(z_\kappa) \\ s.t. \quad & A\,z - b = 0 \\ & L \leq z \leq U \end{aligned} \tag{5}$$

with objective function $P : \mathbb{R}_+^{n+q_2} \to \mathbb{R}_{0_+}$, where $z := (g, \epsilon)^T \in \mathbb{R}^{n+q_2}$ is the dependent variable of the problem, $\tilde{d} := (d, \delta)^T \in \mathbb{R}^{n+q_2}$ the vector of design weights and degrees of penalization for the relaxed benchmarks, $L := (L_g, L_\epsilon)^T \in \mathbb{R}^{n+q_2}$ the lower bounds, and $U := (U_g, U_\epsilon)^T \in \mathbb{R}^{n+q_2}$ the upper bounds for g and ϵ.

As shown in Rupp (2018, Lemma 5.2.1), the objective function P of problem (5) is twice continuously differentiable, strictly convex, and separable for the three distance functions $D : \mathbb{R}_+ \to \mathbb{R}_{0_+}$ of Table 1. Moreover, D' is strongly monotonically increasing, such that the inverse of function D'^{-1} of D' is well-defined for the three distance functions.

Based on that specific structure of the problem, the first order necessary optimality conditions of problem (5) are also sufficient if the Slater condition is satisfied. This can be proved using Theorem 3.8 in Horst (1979), the affine-linearity of all constraint functions, the strict convexity of the objective function, and the strict convexity of the feasible set. Since the objective function is also separable, the problem (5) can be equivalently reformulated as a non-linear system of equations

$$\Psi(\lambda) = 0 \tag{6}$$

in analogy to Münnich et al. (2012b) with

$$\Psi : \mathbb{R}^{q_1+q_2} \to \mathbb{R}^{q_1+q_2}, \quad \lambda \mapsto A\, z(\lambda)\ - b. \tag{7}$$

In that regard, the function $z(\cdot) : \mathbb{R}^{q_1+q_2} \to \mathbb{R}^{n+q_2}$ is component-wise defined as

$$z_\kappa(\lambda) := \mathrm{Pr}_{[L_\kappa, U_\kappa]}\left(D'^{-1}\left(-\frac{A_\kappa^T \lambda}{\tilde{d}_\kappa} \right) \right) \tag{8}$$

for $\kappa = 1, \ldots, n + q_2$ with the projection function $\mathrm{Pr}_{[L_\kappa, U_\kappa]}(\cdot)$ (cf. Definition A.3). Due to regularity of D, the following theorem proves the equivalence of solving (5) and (6):

Theorem 2.1. *A vector $z^* \in \mathbb{R}^{n+q_2}$ is the unique solution of the optimization problem (5) if and only if there exist Lagrangian multipliers $\lambda^* \in \mathbb{R}^{q_1+q_2}$ such that $\Psi(\lambda^*) = 0$ defined in (6) is satisfied.*

For the proof of Theorem 2.1, we refer to Münnich et al. (2012b, Theorem 3). Thus, only the $(q_1 + q_2)$-dimensional nonlinear system of equations in (6) has to be solved to achieve the optimal solution of the $(n + q_2)$-dimensional optimization problem in (5). Then, the solution $z^* \in \mathbb{R}^{n+q_2}$ of problem (5) is component-wise given by

$$z_\kappa^* = z_\kappa(\lambda^*) \tag{9}$$

for all $\kappa = 1, \ldots, n + q_2$ with $z_\kappa(\cdot)$ computed by (8). Finally, the optimal solution $g^* \in \mathbb{R}^n$ and the optimal penalty parameter $\epsilon^* \in \mathbb{R}^{q_2}$ are determined by

$$g^* = (z_1^*, \ldots, z_n^*)^T \quad \text{and} \quad \epsilon^* = (z_{n+1}^*, \ldots, z_{n+q_2}^*)^T. \tag{10}$$

Since $q_1 \ll n$ and $q_2 \ll n$ in the most common applications, the computational burden to solve (6) is supposed to be significantly lower than the computational effort needed to solve problem (5).

2.3 The semismooth Newton algorithm

By introducing box-constraints to the standard calibration method (1), the function Ψ of the non-linear system of equations (6) is not continuously differentiable which prohibits us to apply the classical Newtons method. However, Ψ defined in (7) is (strongly) semismooth (cf. Definition A.2) which is shown in Rupp (2018, Theorem 5.3.2). Thus, the application of a semismooth Newton method (cf. Qi and Sun, 1993) is possible. Fortunately, in analogy to the classical Newtons method Qi (1993) verify a local q-superlinear convergence rate of the semismooth Newton method for semismooth functions Ψ and a local quadratic convergence rate for strongly semismooth functions respectively.

For a detailed analysis of the theory of semismooth functions we refer to Clarke (1979). Concerning the calibration problem, a more adapted overview about the generalized subdifferential theory and semismooth functions is given in Münnich et al. (2012b), Wagner (2013), and Rupp (2018, Chapter 3.2). These sources also consider the convergence results of the semismooth Newton method.

The algorithm is shown in Algorithm 1 with the generalized Jacobian $\partial\Psi(\cdot)$ (cf. Definition A.1) and an appropriate step-size rule to ensure numerical stability.

Algorithm 1 Semismooth Newton method

```
**Input:** $\Psi : \mathbb{R}^{n+q_2} \to \mathbb{R}^{n+q_2}$ locally Lipschitz, $\lambda^0 \in \mathbb{R}^{n+q_2}$ initial iterate, $\lambda^0 \in B(\lambda^*)$
**while** $||\Psi(\lambda^k)|| \geq$ tol **do**
    choose $H^k \in \partial\Psi(\lambda^k)$
    solve $H^k s^k = -\Psi(\lambda^k)$
    compute step-size $\gamma^k \in (0, 1]$
    $\lambda^{k+1} = \lambda^k + \gamma^k s^k$
    $k \leftarrow k + 1$
**end while**
**return** Solution $\lambda^* \leftarrow \lambda^k$
```

Aside from the semismooth Newton method presented in this article, there are several common calibration algorithms in the literature. One example is the function `calib()` in the `R` package `sampling` (cf. Tillé and Matei, 2016). In addition, Vanderhoeft (2001, pp. 29 f.) proposed a similar algorithm based on a projected Newton algorithm, which is implemented in the `SPSS` module `g-CALIB-S`. Beside these two algorithms, there are other algorithms like the function `calibrate()` (`R` package `survey`; cf. Lumley, 2011) and the `SAS` module `CALMAR` (cf. Sautory, 1993), which are mostly based on Newton techniques. In extreme cases (high number of constraints, strict box-constraints) some of these algorithms may have issues in finding the optimal solution due to non-differentiabilities. As an alternative, Wagner (2013) proposed to solve problem (2) using the highly efficient commercial software *IBM ILOG CPLEX Optimization Studio*[1], which provides the optimal solution. Our approach also yields the optimal solution, but performs generally faster since it is well-tailored to the structure of the optimization problem.

[1] https://www.ibm.com/analytics/data-science/prescriptive-analytics/cplex-optimizer

3 Simulation study

The generalized calibration method is applied on the AMELIA dataset (cf. Burgard et al., 2017) consisting of 3 781 289 households accommodating 10 012 600 individuals. The population is stratified by the stratification levels REG (4 regions), PROV (11 provinces), DIS (40 districts), and CIT (1 592 cities/communities). Thus, the sampling design is based on the 1 592 cities (CIT), also called strata, and the overall sampling fraction is fixed to 5%, i.e. the total sample size is given by $n = 189\,064$ households. The stratified samples are drawn using a univariate box-constraint optimal allocation concerning variable *Personal Income* (INC; cf. Gabler et al., 2012 and Münnich et al., 2012c). Further designs can be found on the AMELIA homepage[2]. To allow a stable calibration procedure and estimation, the lower bounds for the stratum-specific sample sizes are set to 100 households per stratum. For the calibration benchmarks, the totals of considered auxiliary variables can be assumed to be known by registers or other surveys, namely

- ZEN (number of persons living in the household),
- EF117A, EF117B (occupational status),
- ILO1, ILO3, ILO4 (type of employment), and
- ISCEDB, ISCEDC, ISCEDD (highest graduation)

as well as additional classes of cross-classifications of age (4 classes) and gender (2), namely

- AGE4.1_Sex.1, AGE4.2_Sex.1, AGE4.3_Sex.1, AGE4.4_Sex.1, and
  AGE4.1_Sex.2, AGE4.2_Sex.2, AGE4.3_Sex.2, AGE4.4_Sex.2.

Some characteristics of these variables are omitted, such as ILO2 due to its rare appearance. We consider three scenarios described in Table 2, depending on the amount of auxiliary variables and stratification level. The number of benchmarks increases from the first to the last row of the table. The stratum-specific (i.e. city-specific) benchmarks for ZEN are included in each of the three scenarios without relaxation. In the first of the three cases, region-specific benchmarks for the auxiliaries are added (as strict benchmarks). In the second case, stratum-specific benchmarks are added for the auxiliaries as soft constraints with a predefined tolerance of 7.5%. The third case additionally contains benchmarks for Age×Gender classes (exact for the regions and relaxed for cities/strata). The overall number of benchmarks is tabulated in the last column.

As no relaxed benchmarks are used, the calibration estimator for the population total concerning scenario `REG.exact` is equivalent to the GREG estimator for the population total with the REG-specific totals of the auxiliaries and stratum-specific totals for ZEN used as benchmarks. To evaluate the functionality of the calibration method, the HT estimates

[2] http://www.amelia.uni-trier.de

Table 2: Scenarios applied in the simulation study. "✓" means the benchmarks are included. The values in percent refer to the maximal allowed tolerance.

	ZEN	Auxiliaries		Age×Gender		Benchmarks
	CIT	REG	CIT	REG	CIT	
`REG.exact`	✓	✓	–	–	–	1 624
`CIT.rel(Aux)`	✓	✓	±7.5%	–	–	14 360
`CIT.rel(Aux&AxG)`	✓	✓	±7.5%	✓	±7.5%	27 128

are also computed. Aside from the analysis of weights and benchmarks in Sections 3.1 and 3.2, the accuracy of point estimates is shown in Section 3.3. The accuracy is measured by the Monte-Carlo RRMSE and RBIAS computed on the basis of the 1 000 Monte-Carlo replications. For each evaluation, the results for the three distance functions of Table 1 are analyzed. The evaluations of Sections 3.1 and 3.2 are based on the results of *one* sample. A randomly conducted study yields similar results for other samples.

3.1 Distribution of calibration weights

In general, the higher the influence of the calibration process (i.e. the higher the number of benchmarks or the more restrictive the benchmarks), the more the correction weights g_k deviate from 1.0. To illustrate this, the correction weights g_k are plotted using density

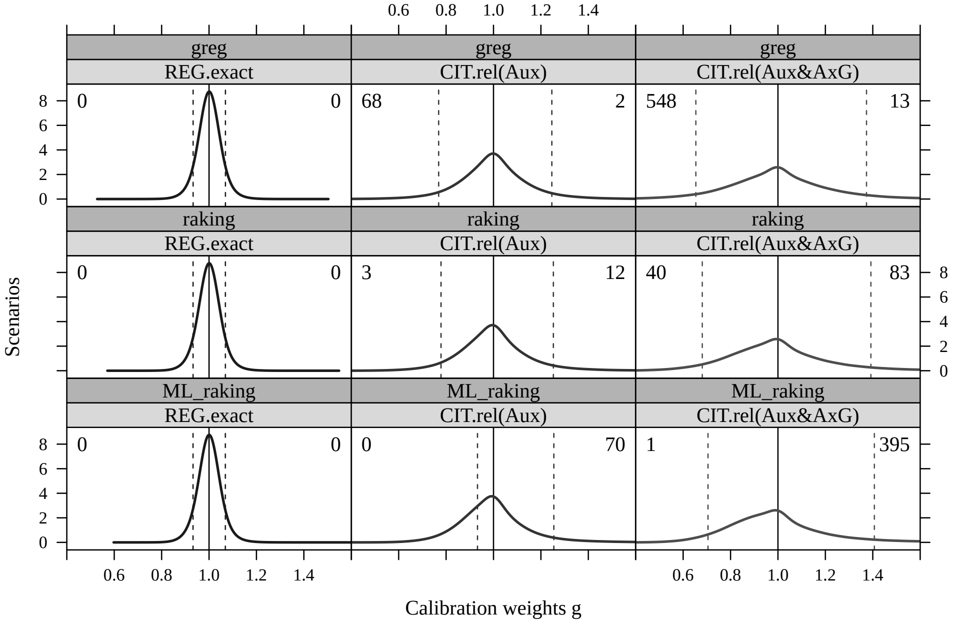

Figure 2: Density plots of correction weights g for three scenarios and distance functions.

plots in Figure 2 for the three scenarios and distance functions. The vertical dashed lines highlight the position of the 5%- and the 95%-quantiles of the weights. Firstly, we observe a significant increase in the variance of the weights for the scenarios two and three with relaxed stratum-specific benchmarks. This is a result of the increased number of benchmarks on disaggregated stratification levels. If we attempt to fulfill the stratum-specific benchmarks exactly without relaxation, the semismooth Newton algorithm breaks down due to the non-feasibility of the problem. When looking at the shapes of the density plots, a similar behavior can be observed for the three distance functions. However, in considering the numbers at the top-right and top-left of each plot (corresponding to the number of weights reaching the respective box-constraints), there are significant differences between the three distance functions. This is consistent with the analysis of the distance function in Figure 1, where the weight g_k being significantly smaller than 1.0 is more penalized for Raking and ML-Raking compared to the GREG-type distance functions.

3.2 Strict versus relaxed benchmarks

As described before, the important feature of the generalized calibration method is to permit specific benchmarks to be relaxed. In Figure 3 the functionality of the relaxation is highlighted. Each boxplot contains the deviation of the totals estimated by the calibration estimator and the benchmark totals for all restrictions which are included in the third scenario (i.e. both stratum- and REG-specific benchmarks). The boxplots are divided into two columns for the *Auxiliaries* and *Age*×*Gender*, respectively. The dashed vertical lines correspond to the maximal allowed tolerance for the relaxed benchmarks. In the scenario without consideration of the stratum-specific benchmarks (`REG.exact`), there are several stratum-specific estimates which substantially differ from their benchmark totals and significantly exceed the maximal tolerance used in the scenarios two and three with stratum-specific benchmarks. The maximal deviations are about +65% and −55%, which are unacceptable if stratum-specific estimates are subject of the survey. This results in regional inconsistencies, coherence problems, and inefficient stratum-specific estimates. The possibility of the relaxation of specific benchmarks (see scenarios two and three) prevents

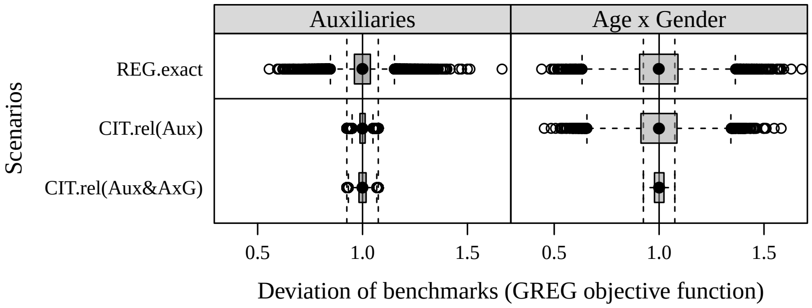

Figure 3: Compliance with benchmarks of SMP-specific estimates for scenarios with the GREG-type objective function.

those issues and enables the inclusion of stratum-specific totals, which may be gained from registers or other surveys. Since benchmarks for highly disaggregated stratification levels may appear to be important, benchmarks gained from small area estimates may also be applied. The circumstance of including estimated benchmarks with different standard errors can be considered by using different tolerance terms.

With regard to Table 2, the high number of benchmarks considered in the scenarios two and three with relaxed benchmarks highlights the fact, that an exact fulfillment of all those benchmarks would simply lead to an empty feasible set of problem (2), and therefore to the non-feasibility of the problem. Thus, using relaxation techniques is the only way to include such a high amount of restrictions.

3.3 Point estimates on regional level

Aside from coherence and consistency, increasing the accuracy of estimates is a key aspect of the generalized calibration method. The RRMSEs of the city-specific point estimates for six selected variables are shown in Figure 4 for all scenarios and for the GREG-type objective functions. The variables comprise three variables which are (partly) included in the calibration (upper panels) and three variables of interest (lower panels), which are not involved in the calibration. Each boxplot contains 1592 points assigned to the 1592 cities of AMELIA.

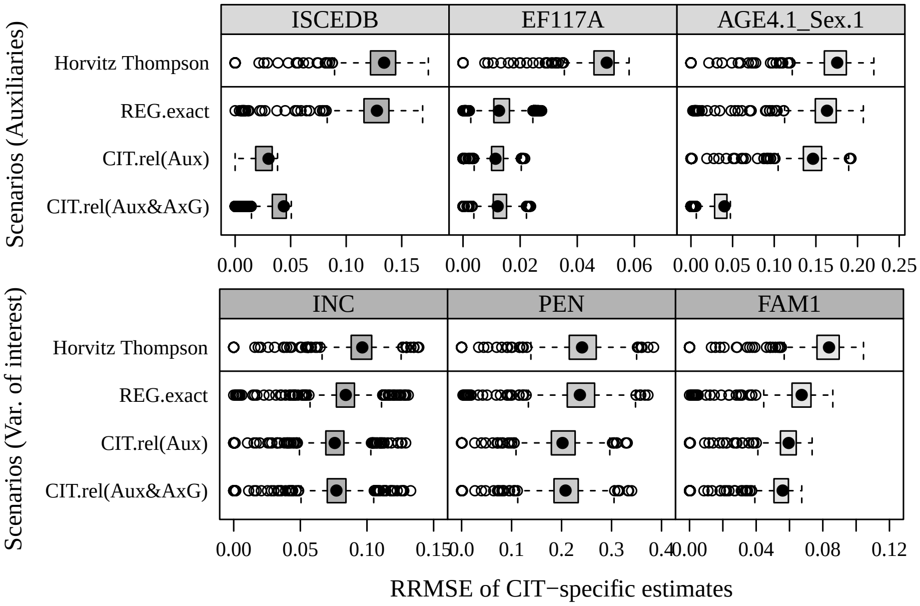

Figure 4: RRMSE of city-specific point estimates for scenarios with a GREG-type objective function.

In general, the point estimates for all variables (auxiliaries and variables of interest) and all scenarios are at least as accurate as the HT estimates. Further, significant differences can be observed in the behavior of the point estimates of the auxiliaries, which is primarily a consequence of the usage of the auxiliary variables as benchmarks. With regard to the scenario `REG.exact`, the accuracy of the stratum-specific (i.e. city-specific or CIT-specific) estimates of all the auxiliaries ISCEDB and AGE4.1_Sex.1 is similar to the HT estimates, since none of the auxiliary variables is included as benchmark on stratum-level. In contrast, the accuracy of the stratum-specific estimates of EF117A increases due to the high correlation between EF117A and ZEN (which is already included as stratum-specific benchmark; see Table 2). In case of the other scenarios, the accuracy of the stratum-specific estimates is significantly improved for the auxiliary variables, since they are applied as relaxed benchmarks with predefined maximal perturbations. Moreover, the inclusion of specific benchmarks does not necessarily lead to accuracy-changes for other not included variables. In scenario `CIT.rel(Aux&AxG)`, the accuracy of the estimates for ISCEDB and EF117A slightly suffer due to the high amount of benchmarks included, as the feasible set is shrunken.

With regard to the variables of interest, different behaviors can also be observed. The accuracy of the stratum-specific estimates increases if additional benchmarks are included due to the correlation structure between the variables of interest and the auxiliaries. However, as the increase of accuracy compared to the HT estimator may only be small it has to be considered that the main task of the generalized calibration is not only to obtain accuracy increased estimates but also to have consistency between benchmarks and auxiliaries.

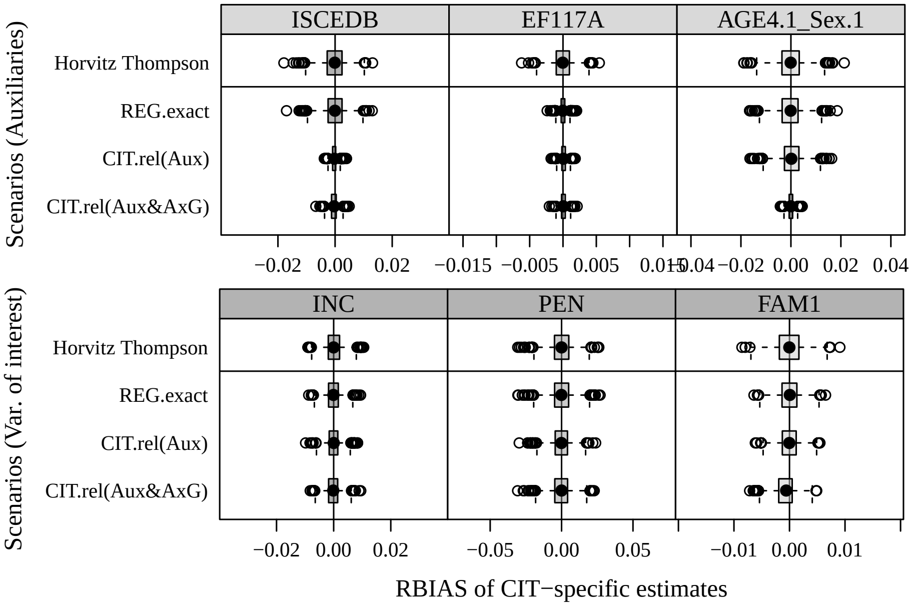

Figure 5: RBIAS of city-specific point estimates for scenarios with a GREG-type objective function.

In Figure 5, the RBIAS is presented for the city-specific point estimates for the same scenarios and variables. Per definition, the HT estimator is unbiased for all variables. Since the GREG estimator is model-assisted and an calibrated estimation can also be associated to this type of estimators, estimates based on the generalized calibration method should at least be asymptotically unbiased. Generally, unbiased estimates for most scenarios are observed, including those with relaxed benchmarks. In addition, the bias is reduced if the respective auxiliary variable is included as relaxed benchmarks in the calibration process. The amount of the reduction highly depends on the predefined tolerances.

4 Concluding remarks

Calibration estimation has shown to provide a sound basis for producing coherent weights. However, practical settings with many constraints may imply unacceptable weight distributions or even non-solvable problems. This problem becomes even more evident when considering estimates on sub-groups or smaller regions. The inclusion of box-constraints and tolerances helps to overcome non-solvability of the calibration problem and to produce more appropriate weights. In Germany, this is essentially important for the register-assisted *German Census* (cf. Münnich et al., 2012a) and the new integrated *German System of Household Surveys* (cf. Riede et al., 2013).

The proposed generalized calibration method allows smoothly to include many different additional constraints in an appropriate manner. This includes the accuracy of the benchmarking information, small area constraints or hierarchical settings. The iterative algorithm, which is based on applying the semismooth Newton algorithm with step-control, ensures the global optimum. This may not be achieved when using hard constraints on the weights in combination with fully Newton-based methods. The solution strategy can also be applied to large-scale calibration problems with acceptable computation time.

Acknowledgements

The research is part of the project *Research Innovation for Official and Survey Statistics (RIFOSS)*. We acknowledge the kind support by the German Federal Statistical Office which made this research possible.

References

Beaumont, J.-F. and Bocci, C. (2008). Another look at ridge calibration. *Metron - International Journal of Statistics*, LXVI(1):5–20.

Burgard, J. P., Kolb, J.-P., Merkle, H., and Münnich, R. (2017). Synthetic data for open and reproducible methodological research in social sciences and official statistics. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 11(3):233–244.

Chambers, R. (1996). Robust case-weighting for multipurpose establishment surveys. *Journal of Official Statistics*, 12(1):3–32.

Chen, J., Sitter, R. R., and Wu, C. (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika*, 89(1):230–237.

Clarke, F. H. (1979). *Nichtlineare Optimierung*. Carl Hanser Verlag.

Deville, J. and Särndal, C. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382.

Deville, J. and Särndal, C. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88:1013–1020.

Eurostat (2011). European Statistics Code Of Practice. Retrieved from http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15. Adopted by the European Statistical System Committee, visited 14 May 2019.

Gabler, S., Ganninger, M., and Mïnich, R. (2012). Optimal allocation of the sample size to strata under box constraints. *Metrika*, 75(2):151–161.

Gelman, A. (2007). Struggles with survey weighting and regression modeling. *Statistical Science*, 22.

Haziza, D. and Beaumont, J.-F. (2017). Construction of weights in surveys: A review. *Statistical Science*, 32(2):206–226.

Horst, R. (1979). *Nichtlineare Optimierung*. Carl Hanser Verlag.

Kott, P. (2006). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors. *Survey Methodology*, 32:133–142.

Lehtonen, R. T. and Veijanen, A. K. K. (2015). Estimation of poverty rate for small areas by model calibration and "hybrid" calibration methods. In *Proceedings of the New Techniques and Technologies for Statistics 2015 Conference. Brussels: European Commission.*

Lumley, T. (2011). R package survey: Analysis of complex survey samples. http://CRAN.R-project.org/package=survey. R package version 3.24.

Merkouris, T. (2004). Combining independent regression estimators from multiple surveys. *Journal of the American Statistical Association*, 99(468):1131–1139.

Montanari, G. E. and Ranalli, M. G. (2005). Nonparametric model calibration estimation in survey sampling. *Journal of the American Statistical Association*, 100:1429–1442.

Montanari, G. E. and Ranalli, M. G. (2009). Multiple and ridge model calibration for sample surveys. *Proceedings of the Workshop in Calibration and Estimation in Surveys, Ottawa, October 2007*.

Münnich, R. and Burgard, J. P. (2012). On the influence of sampling design on small area estimates. *Journal of the Indian Society of Agricultural Statistics*, 66:145–156.

Münnich, R., Burgard, J., P., and Vogt, M. (2013). Small Area-Statistik: Methoden und Anwendungen. *AStA Wirtschafts- und Sozialstatistisches Archiv*, 6(3/4):149–191.

Münnich, R., Gabler, S., Ganninger, M., Burgard, J. P., and Kolb, J.-P. (2012a). *Statistik und Wissenschaft: Stichprobenoptimierung und Schätzung im Zensus 2011*, volume 21. Statistisches Bundesamt, Wiesbaden.

Münnich, R., Sachs, E., and Wagner, M. (2012b). Calibration of estimator-weights via semismooth Newton. *Journal of Global Optimization*, 52(3):471–485.

Münnich, R., Sachs, E. W., and Wagner, M. (2012c). Numerical solution of optimal allocation problems in stratified sampling under box constraints. *AStA Advances in Statistical Analysis*, 96:435–450.

Qi, L. (1993). Convergence analysis of some algorithms for solving nonsmooth equations. *Mathematics of Operations Research*, 18(1):227–244.

Qi, L. and Sun, J. (1993). A nonsmooth version of newton's method. *Mathematical Programming*, 58.

Rao, J. N. K. (2003). *Small Area Estimation*. John Wiley and Sons, Ltd.

Rao, J. N. K. and Singh, A. C. (1997). A ridge-shrinkage method for range-restricted weight calibration in survey sampling. *Proceedings of the Survey Research Methods Section, American Statistical Association*, pages 57–65.

Riede, T., Bechtold, S., and Ott, N. (2013). *Weiterentwicklungen der amtlichen Haushaltsstatistiken, 1. Auflage*. Scivero Verlag, Berlin.

Rupp, M. (2018). *Optimization for Multivariate and Multi-domain Methods in Survey Statistics*. PhD thesis, Trier University.

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

Sautory, O. (1993). La macro CALMAR, redressement d'un échatillon par calage sur marges. https://www.insee.fr/fr/information/2021902. SAS macro.

Singh, A. and Mohl, C. (1996). Understanding Calibration Estimators in Survey Sampling. *Survey Methodology*, 22:107–115.

Statistics Canada (2003). *Quality Guidelines*. Ottawa: Minister of Industry, fourth edition edition. Catalogue no. 12539XIE.

Stukel, D., Hidiroglou, M., and Särndal, C. (1996). Variance Estimation for Calibration Estimators: A Comparison of Jackknifing Versus Taylor Linearization. *Survey Methodology*, 22:117–125.

Théberge, A. (2000). Calibration and restricted weights. *Survey Methodology*, 26:99–107.

Tillé, Y. and Matei, A. (2016). R Package Sampling: Survey sampling. http://CRAN.R-project.org/package=sampling. R package version 2.8.

Tzavidis, N., Zhang, L.-C., Luna, A., Schmid, T., and Rojas-Perilla, N. (2018). From start to finish: a framework for the production of small area official statistics. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(4):927–979.

Vanderhoeft, C. (2001). Statistics belgium working papers. Generalized calibration at statistics belgium. *Tech. Rep. STATBEL*, pages 1–192.

Wagner, C. (2013). *Numerical Optimization in Survey Statistics*. PhD thesis, Trier University.

Wu, C. and Sitter, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96(453):185–193.

A Appendix

The theory of semismooth functions is based on Clarke (1979), where a generalized subdifferential theory is introduced for semismooth functions. Moreover, the semismooth Newton method and its convergence results are presented in Qi and Sun (1993) and Qi (1993). In addition to that, Münnich et al. (2012b), Wagner (2013), and Rupp (2018) give a more application-specific overview about the semismooth theory. Thus, we largely skip this theory and concentrate on the main results relating to the topic of this paper. Firstly, we define the generalized Jacobian of a function F:

Definition A.1. *(Generalized Jacobian) Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz, $x \in \mathbb{R}^n$, and let and let $F'(x) \in \mathbb{R}^{m\times n}$ denote the Jacobian of F in $x \in D_F := \{x \in \mathbb{R}^n : F$ is differentiable in $x\}$. Then*

$$\partial F(x) := conv\{H \in \mathbb{R}^{m\times n} : \exists (x^k)_{k\in\mathbb{N}} \subset D_F : x^k \to x \text{ and } D'(x^k) \to H\}$$

is called the ***generalized Jacobian*** *of F in x.*

We note, that for all $x \in \mathbb{R}^n$ where F is differentiable, the generalized Jacobian $\partial F(x)$ is of cardinality one and equals the Jacobian $F'(x)$. Next, we define semismoothness of a function F which is locally Lipschitz:

Definition A.2. *(Semismoothness) Let $D \subseteq \mathbb{R}^n$ and $F : D \to \mathbb{R}^m$ be a locally Lipschitz function and let all directional derivatives $F'(x; r)$ in direction r exist in $x \in D$. Then F is called*

(i) ***semismooth*** *in $x \in D$, if* $\displaystyle\lim_{r^k\to 0, H^k\in\partial F(x+r^k)} \frac{H^k r^k - F'(x; r^k)}{||r^k||} = 0,$

(ii) ***strongly semismooth*** *in $x \in D$, if* $\displaystyle\limsup_{r^k\to 0, H^k\in\partial F(x+r^k)} \frac{H^k r^k - F'(x; r^k)}{||r^k||^2} < \infty,$

(iii) (strongly) semismooth on D, if F is (strongly) semismooth in all $x \in D$.

We remark that the property of semismoothness is weaker than the differentiability. In contrast to differentiable functions, semismooth functions may have some *kinks*. One example for a semismooth function is the projection:

Definition A.3. *(Projection) Let $[a, b] \in \mathbb{R}$ be an interval and $x \in \mathbb{R}$. The projection of x on $[a, b]$ is defined as*

$$Pr_{[a,b]}(x) := \begin{cases} a\ , & \text{if } x \le a \\ x\ , & \text{if } a < x < b\ . \\ b\ , & \text{if } x \ge b \end{cases}$$