Optimization for Multivariate and Multi-domain Methods in Survey Statistics

Dissertation

zur Erlangung des akademischen Grades eines
Dr. rer. pol.

Dem Fachbereich IV der Universität Trier
vorgelegt von

Martin Johann Andreas Rupp

Gutachter:
Prof. Dr. Ralf Münnich (Universität Trier)
Prof. Dr. Ekkehard Sachs (Universität Trier)

Trier, 2018

Die Dissertation wurde im Rahmen des Forschungsprojektes *Research Innovation for Official and Survey Statistics (RIFOSS)* angefertigt.

Datum der mündlichen Doktorprüfung: 05. Oktober 2018

Gutachter: Prof. Dr. Ralf Münnich (Universität Trier)
Prof. Dr. Ekkehard Sachs (Universität Trier)

Vorwort

Zuallererst möchte ich mich sehr herzlich bei Prof. Dr. Ralf Münnich für die exzellente Betreuung während meiner Promotionszeit bedanken. Ich bin nicht nur sehr dankbar dafür, dass ich die Möglichkeit bekommen habe, an seinem Lehrstuhl als wissenschaftlicher Mitarbeiter tätig zu sein, sondern insbesondere auch für seine inspirierenden Ideen, fachlich kompetente Beratung und stets unterstützende Betreuung. Die ausgezeichnete Zusammenarbeit mit ihm schätze ich sehr. Darüber hinaus möchte ich Prof. Dr. Ekkehard Sachs sehr herzlich für seine Unterstützung sowohl während meines Studiums als auch während meiner Zeit als wissenschaftlicher Mitarbeiter danken. Während der Promotionszeit trugen seine numerische Expertise und zahlreiche anregende Diskussionen entscheidend zu den Entwicklungen der Arbeit bei.

Die Forschungsarbeiten wurden vom Statistischen Bundesamt in Wiesbaden innerhalb des Forschungsprojektes *Research Innovation for Official and Survey Statistics (RIFOSS)* finanziell gefördert. Für die Möglichkeit, die Dissertation im Rahmen des Forschungsprojektes anzufertigen sowie die ausgezeichnete Zusammenarbeit bedanke ich mich sehr. Als assoziiertes Mitglied gilt mein Dank zudem dem Graduiertenkolleg *Algorithmic Optimization (ALOP)*, finanziert von der Deutschen Forschungsgemeinschaft, für die Unterstützung während meiner Promotionszeit.

Des Weiteren möchte ich mich bei allen Kollegen vom Lehrstuhl für Wirtschafts- und Sozialstatistik sowie aus dem Graduiertenkolleg *ALOP* sehr herzlich bedanken. Sie alle haben maßgeblich zu einer sehr kollegialen und freundschaftlichen Arbeitsatmosphäre beigetragen. Ihre Türen standen mir jederzeit sowohl für fachliche als auch persönliche Anliegen offen. Die zahlreichen inhaltlichen Diskussionen haben einen nicht zu unterschätzenden Mehrwert für meine Dissertation geliefert. Neben der exzellenten fachlichen Zusammenarbeit sind in dieser Zeit auch enge private Freundschaften entstanden, die ich keinesfalls mehr missen möchte.

Zu guter Letzt möchte ich mich bei meiner Familie bedanken, die mich während meiner gesamten akademischen Karriere in jeder erdenklichen Situation uneingeschränkt unterstützt und gefördert haben. Sie gaben mir stets die Kraft und Zuversicht, die nötig war, um die Arbeit fertigzustellen. Insbesondere meiner Freundin Julia Manecke bin ich für die Unterstützung und die unermüdliche Geduld in der Schlussphase meiner Promotion unendlich dankbar. Die Zeit mit ihr gab mir den Ausgleich vom Arbeitsumfeld, der mich abschalten und neue Energie tanken ließ.

Contents

German Summary

Diese Dissertation mit dem Titel *Optimization for Multivariate and Multi-domain Methods in Survey Statistics* beschäftigt sich mit Methoden zur Verbesserung der Schätzgüte von designbasierten und modell-assistierten Schätzern für Stichprobenerhebungen auf Basis finiter Populationen. Dabei stehen sowohl die entwickelten statistischen Methoden als auch deren Umsetzung mittels darauf zugeschnittener numerischer Optimierungsstrategien im Vordergrund. Die Motivation zur Entwicklung der statistischen Methoden entstammt sowohl den vielseitigen Anforderungen an die Ergebnisse von Stichprobenerhebungen als auch der gestiegenen Menge verfügbarer Hilfsinformationen. Die Anforderungen resultieren aus Vorgaben zur Schätzung von Statistiken für mehrere potenziell konfliktäre interessierende Variablen auf unterschiedlichen Ebenen der Population anhand einer einzigen Erhebung. Neben den üblichen nationalen Statistiken sind somit auch Statistiken für regional und inhaltlich differenzierte Subpopulationen von Bedeutung.

Die Arbeit lässt sich in zwei Hauptforschungsfragen untergliedern. Zum einen wird eine optimale multivariate Allokationsmethode unter Berücksichtigung mehrerer Schichtungsebenen entwickelt (Kapitel 4), die den Gesamtstichprobenumfang in einem stratifizierten Stichprobendesign unter vielseitigen Restriktionen varianzoptimal auf die einzelnen Schichten aufteilt. Zum anderen wird eine generalisierte Kalibrierungsmethode vorgestellt (Kapitel 5). Diese soll kohärente und effiziente Schätzungen auf unterschiedlichen Schichtungsebenen unter Beachtung von Hilfsinformationen erzielen, die aus einer Vielzahl unterschiedlichster Quellen gewonnen wurden. In den folgenden beiden Abschnitten werden die Methoden kurz erläutert.

In stratifizierten Stichprobendesigns besteht ein zentraler Teil des Auswahlprozesses in der Allokation des Gesamtstichprobenumfangs auf die einzelnen Schichten. Übliche Maßstäbe für diese Aufteilung sind bei gegebenem Gesamtstichprobenumfang die Gleichheit aller schichtspezifischen Stichprobenumfänge, die Proportionalität zur Schichtgröße oder die varianzoptimale Verteilung im Sinne der Minimierung der Varianz des Populationsschätzers für interessierende Variablen. Das entwickelte Modell basiert grundsätzlich auf der varianzminimalen Verteilung, deren Betrachtung durch die gleichzeitige Berücksichtigung mehrerer interessierender Variablen zu einem Mehrzieloptimierungsproblem führt. Als Vorbereitung der numerischen Lösung werden neben Standardisierungsverfahren insbesondere auch mehrere Skalarisierungsmethoden vorgestellt, die die unterschiedlichen Präferenzen möglicher Anwender abbilden. Zudem wird gezeigt, dass sich durch die Lösung des mit einer gewichteten Summe skalarisierten Problems für alle Gewichtekombinationen die gesamte Pareto Front des Ursprungsproblems erzeugen lässt. Durch die Ausnutzung der speziellen Struktur des Problems lässt sich dieses sehr effizient mit Hilfe eines semismooth Newton Verfahrens lösen, sodass sich eine sehr genaue Approximation der Pareto Front berechnen lässt. Um diese Methode auch für andere Skalarisierungsmethoden anwenden zu können, wird unter Verwendung eines projizierten inexakten Quasi-Subgradientenverfahrens eine Verbindung zwischen der gewichteten Summe und den übrigen Skalarisierungsmethoden ausgenutzt. Zur Berücksichtigung von Anforderungen an regionale Schätzqualitäten auf mehreren Schichtungsebenen wird die Verwendung von unteren Schranken für die Schätzqualität in das Modell integriert. Diese Schranken dienen im Speziellen nicht der Optimierung von regionalen Schätzungen, sondern der Vermeidung unerwünscht hoher regionaler Schätzfehler aufgrund von ungewöhnlichen Strukturen innerhalb der Regionen sowie kleinen Stichprobenumfängen. Neben den Restriktionen für regionale Schätzungen ermöglicht die Methode auch die Verwendung von Box-Constraints für die schichtspezifischen Stichprobenumfänge, wodurch minimale und maximale schichtspezifische Auswahlsätze festgelegt werden können.

Die Kalibrierung der aus dem Allokationsprozess resultierenden Designgewichte ist eine verbreitete Methode zur Erzielung effizienter und kohärenter Schätzungen. Die entwickelte generalisierte Kalibrierungsmethode berücksichtigt dabei eine sehr hohe Anzahl von Benchmarks auf unterschiedlichen Schichtungsebenen. Da diese Benchmarks möglicherweise aus unterschiedlichen Quellen wie Registern, Paradaten oder anderen Umfragen mit unterschiedlichen Schätzmethoden gewonnen werden, ist die Qualität dieser Benchmarks sehr heterogen. Um der Qualität und der Vielzahl von Benchmarks gerecht zu werden, wird eine Relaxierung ausgewählter Benchmarks vorgeschlagen, wobei problematischen Benchmarks auf niedrigen Aggregationsebenen vordefinierte Toleranzen zugeordnet werden, um eine exakte Erfüllung zu vermeiden. Neben der GREG-typischen Penalisierungsmethode zur Bestimmung der Höhe der Bestrafung von Abweichungen zwischen Design- und Kalibrierungsgewichten werden weitere Zielfunktionen untersucht. Darüber hinaus ermöglicht die generalisierte Kalibrierungsmethode die Verwendung von Box-Constraints für die Korrekturgewichte, um zu extreme Gewichte zu vermeiden und um eine Beschränkung der Variation der Gewichte zu ermöglichen. Da die Verwendung eines Residualvarianzschätzers für den Kalibrierungsschätzer auf Basis der entwickelten Methode insbesondere für regionale Schätzer nicht ohne Weiteres möglich ist, wird eine Varianzschätzung mittels eines *Rescaling Bootstraps* vorgestellt, welcher eine spezielle Form der bootstrapbasierten Resamplingverfahren darstellt.

Beide entwickelten Methoden werden in umfangreichen Simulationsstudien auf Basis eines realitätsnahen synthetischen Datensatzes aller Haushalte Deutschlands analysiert und mit existierenden Methoden verglichen. Obwohl die beiden Methoden unterschiedlichen Teilen des Survey-statistischen Prozesses zuzuordnen sind (Auswahl- bzw. Schätzprozess), stehen sie in engem Zusammenhang zueinander. Aufgrund der bereits genannten ähnlichen Grundvoraussetzungen und Zielsetzungen beider Verfahren lassen sich diese sukzessive auf eine einzelne Stichprobenerhebung anwenden, um die jeweiligen Effizienzvorteile zu kombinieren. Zudem lassen sich beide Methoden anhand vergleichbarer Optimierungsansätze zeiteffizient lösen. Diese basieren auf Umformungen der Optimalitätsbedingungen. Die Dimension der resultierenden nichtlinearen und nicht differenzierbaren Gleichungssysteme ist letztendlich unabhängig von der Dimension der ursprünglichen Probleme, was insbesondere für die Lösung von sehr großen Problemstellungen extreme Zeitersparnisse mit sich bringt.

List of Figures

List of Tables

List of Algorithms

List of Symbols

Survey framework

$N \in \mathbb{N}$	total size of population
$n_{\text{s}} \in \mathbb{N}$	total sample size
$\mathcal{U} = \{1, \ldots, N\}$	finite population of size N
$\mathcal{U}_h$	finite population in stratum h ($h = 1, \ldots, H$) of size N_h
$\mathcal{U}_{l_r}$	finite population in area l_r ($l_r = 1, \ldots, L_r$) of stratification level $r = 1, \ldots, R$
$S \subseteq \mathcal{U}$	sample (without replacement) of population $\mathcal{U}$ of size $n_{\text{s}} \in \mathbb{N}$
$S_h \subseteq \mathcal{U}$	stratum-specific sample of stratum $\mathcal{U}_h$ of size $n_h \in \mathbb{N}$
$\mathbb{S}$	set of all possible samples (without replacement) of population $\mathcal{U}$
$\mathbb{S}_{n_{\text{s}}}$	set of all possible samples of population $\mathcal{U}$ of size $n_{\text{s}} \in \mathbb{N}$
$n \in \mathbb{R}^H$	vector of stratum-specific sample sizes $n_1, \ldots, n_H$
$p(\cdot) : \mathbb{S} \to [0, 1]$	sampling design
π_k	first-order inclusion probability of unit $k \in \mathcal{U}$
π_{kl}	second-order inclusion probability of units $k, l \in \mathcal{U}$
d_k	design weight of unit $k \in \mathcal{U}$
g_k	correction weight of unit $k \in \mathcal{U}$
$w_k := d_k g_k$	calibration weight of unit $k \in \mathcal{U}$
$D : \mathbb{R}_+ \to \mathbb{R}_{0_+}$	distance function for calibration
$y \in \mathbb{R}^N$	vector of variable of interest
y_k	value of variable of interest of unit $k \in \mathcal{U}$
$x_k \in \mathbb{R}^q$	vector of auxiliary variables $i = 1, ..., q$ for unit $k \in \mathcal{U}$
$X := [x_1, \ldots, x_N]$	matrix of auxiliary variables $i = 1, ..., q$
x_{ik}	value of auxiliary variable $i \in \{1, ..., q\}$ for unit $k \in \mathcal{U}$
S_y^2, S_{hy}^2	variance of variable y, stratum-variance of variable y
s_y^2, s_{hy}^2	estimated variance of variable y, estimated stratum-variance of y
$\bar{y}$, $\bar{y}_h$, $\bar{y}_S$	mean, stratum-specific mean, or sample mean of variable y
ϑ	general statistic
$\hat{\vartheta} : \mathbb{S} \to \mathbb{R}$	estimator for statistic ϑ
$\hat{\vartheta}(S)$	estimate for statistic ϑ concerning sample $S \in \mathbb{S}$
$E(\cdot)$	expected value
$\text{Var}(\cdot)$	variance
$\text{BIAS}(\cdot)$	bias

$\mathrm{MSE}(\cdot)$	mean squared error
$\mathrm{cv}(\cdot)$	coefficient of variation
$\mathrm{cor}(\cdot,\cdot)$	correlation between two variables
$\beta \in \mathbb{R}^q$	regression coefficient of q auxiliary variables X
$\hat{\beta} \in \mathbb{R}^q$	estimated regression coefficient of q auxiliary variables X
τ_y, τ_{hy}, $\tau_{l_r y}$	total value of population, stratum h, or area l_r of variable y
μ_y, μ_{hy}, $\mu_{l_r y}$	mean value of population, stratum h, or area l_r of variable y
$\hat{\tau}_y^{\text{HT}}$, $\hat{\tau}_{yh}^{\text{HT}}$, $\hat{\tau}_{yl_r}^{\text{HT}}$	HT estimate of (population, stratum-specific, area-specific) total value of variable y
$\hat{\tau}_y^{\text{GREG}}$, $\hat{\tau}_{yh}^{\text{GREG}}$, $\hat{\tau}_{yl_r}^{\text{GREG}}$	GREG estimate of (population, stratum-specific, area-specific) total value of variable y
$\hat{\tau}_y^{\text{CAL}}$, $\hat{\tau}_{yh}^{\text{CAL}}$, $\hat{\tau}_{yl_r}^{\text{CAL}}$	calibration estimate of (population, stratum-specific, area-specific) total value of variable y

Indices and notations

R_{MC}	number of Monte-Carlo replicates
R_{Boot}	number of replicates of rescaling bootstrap
H	number of cross-classification strata
R	number of stratification level
L_r	number of areas in stratification level r
r	index of stratification level $r \in R$
h	index of strata $h = 1, \dots, H$
l_r	index of areas with $l_r = 1, \dots, L_r$ and $r = 1, \dots, R$
k	index of unit $k \in \mathcal{U}$
i	index of variables $i = 1, \dots, q$
q	general number of variables of interest or auxiliaries
q_1	`MMDopt`: number of objective functions
q_2	`MMDopt`: number of equality constraints
q_3	`MMDopt`: number of inequality constraints
q_1	`GCAL`: number of strict restrictions
q_2	`GCAL`: number of relaxed restrictions
κ	`GCAL`: composed index of unit $k = 1, \dots, n_{\text{s}}$ and relaxed restrictions $j = 1, \dots, q_2$
$\nu \in \mathbb{N}$	natural number
$\alpha_k > 0$	step-size in iteration k
$\gamma_i \in \mathbb{R}$	standardization factor in multivariate allocation ($i = 1, \dots, q_1$)
$x^k, x^{k+1}, \dots$	iterates of an algorithm
Φ	vector-valued function for the system of equations to solve `MMDopt`
Ψ	vector-valued function for the system of equations to solve `GCAL`
φ	real valued function used to define Φ and Ψ
F, G	vector-valued functions
F_i	real-valued components of vector-valued function F
f, g, h	real-valued functions
ϵ	vector of perturbations for relaxed benchmarks in the calibration

δ	vector of penalties for relaxed benchmarks in the calibration
m, M, L, U	vectors of box-constraints
A	matrix of affine-linear equality constraints
b	right-hand side of affine-linear equality constraints
D	matrix of inequality constraints
c	right-hand side of inequality constraints
w	vector of weights for scalarization
$d_h(w)$	function depending on weights to built Φ for `MMDopt`

Optimization framework

$n, m \in \mathbb{N}$	natural number (only in Chapter 3)
$x, y \in \mathbb{R}^n$	vectors (only in Chapter 3)
$x^* \in \mathbb{R}^n$	denotes the optimal solution of an optimization problem
$\mathbb{R}^n_+$	set of all positive vectors, i.e. $z \in \mathbb{R}^n$ with $z_i > 0$
$\mathbb{R}^n_{+0}$	set of all non-negative vectors, i.e. $z \in \mathbb{R}^n$ with $z_i \geq 0$
$0_{\mathbb{R}^n}$	n-dimensional zero-vector
Z	set in $\mathbb{R}^n$
$\|Z\|$	cardinality of a set
$\|x\|$	absolute value of a vector
$\\|x\\|$	2-norm of a vector
U_x	unspecified neighborhood of $x \in \mathbb{R}^n$
$\mathbb{1}_{(\cdot)}$	indicator function
id$(\cdot)$	identity function
Proj$_{[a,b]}(\cdot)$	projection on the interval $[a, b]$ with $a, b \in \mathbb{R}$
Pr$(\cdot)$	probability
diag$(\cdot)$	diagonal elements of matrix
θ	merit function
$F'(x, r)$	directional derivative of F at x in direction r
$J_F(x)$	Jacobian of F at x
$\partial_B F(x)$	B-subdifferential of F at x
$\partial F(x)$	generalized Jacobian of F at x
H	element of B-subdifferential or generalized Jacobian
$\mathcal{L}$	Lagrangian function
$\lambda, \beta, \mu, \kappa$	Lagrangian multipliers
$\mathcal{O}(\cdot)$	Landau symbol to describe the order of running times / growth rates
$\binom{N}{n}$	binomial coefficient
R	relation
$\preceq$	partial ordering relation
$\phi(\cdot)$	model map for multi-criteria optimization
$C_{\mathbb{R}^n}$	cone on the $\mathbb{R}^n$
S_Z	section of the set $Z \subseteq \mathbb{R}^n$
$I(x)$	set of active constraints in $x \in \mathbb{R}^n$
$C_+(F)$	epigraph of F
$b[-\text{ind}]$	in pseudo-code; vector b without components contained in ind

List of Abbreviations

Algorithms and methods

`GCAL`	generalized calibration method
`genCalib`	`R` package for `GCAL`
`GTM`	projected inexact quasi-subgradient method / algorithm
`MMDopt`	optimal multivariate and multi-domain allocation method
`MultOptAlloc`	`R` package for `MMDopt`
`ProjN`	special case of projected Newton method / algorithm
`SQP`	sequential quadratic programming
`SSN`	semismooth Newton method / algorithm
`TRUNC`	truncated method / algorithm
`.reg`	additional description for `MMDopt` with constraints for regional efficiency

Abbreviations

HT estimator	Horvitz-Thompson estimator
GREG estimator	generalized regression estimator
SRS	simple random sampling
StrRS	stratified random sampling
SMP	sampling point
NUTS	regional structure; nomenclature des unités territoriales statistiques
EQ	equal allocation
PROP	proportional allocation
OPT	optimal allocation
KKT	Karush-Kuhn-Tucker
NCP	nonlinear complementarity problem
(cv)	standardization with coefficient of variation
(opt)	standardization with variance of the optimal univariate allocation
F-differentiable	Fréchet-differentiable
B-differentiable	Bouligand-differentiable
`R`	programming software `R`
`C++`	programming software `C++`
MSE	mean squared error

relMSE	relative mean squared error
RRMSE	relative root mean squared error
BIAS	bias
RBIAS	relative bias
MC	Monte-Carlo
(EF)	characterization of optimality: class of efficient (Pareto optimal) solutions
(norm)	characterization of optimality: class of p-norm optimal solutions
(max)	characterization of optimality: class of min-max optimal solutions
(VP)	multi-criteria optimization problem
(P.OP)	p-norm scalarized problem (VP)
(Ws.OP)	weighted sum scalarized problem (VP)
(VP.P)	multi-criteria optimization problem (VP) to the power of p

Chapter 1

Introduction

1.1 Motivation

The demand for statistically processed information has been rising within the past few decades and this trend is most likely to continue in the next years. Even Särndal et al. (1992, p. 3) phrased that "the need for statistical information seems endless in modern society". As the amount of statistical information grows, the complexity of the underlying models has also increased significantly. In politics and economics, several wide-ranging decisions are made based on statistical analyses. The corresponding decisions may have an impact on the future of individuals, companies, or even entire nations. These aspects illustrate the societal interest in the provision of accurate statistical data in order to prevent mismanagement or wrong decisions. With regard to the high amount of information available and the complexity of the underlying models, this thesis deals with tailor-made multivariate and multi-domain methods in order to provide high quality statistics.

In official statistics, the framework of collecting and processing data as well as publishing statistics is generally mandated by law to specific administrative authorities, such as the Federal Statistical Office of Germany (Destatis) or the European Statistical Office (Eurostat). The legislative basis is governed by the *Bundesstatistikgesetz (BStatG)* for Germany (Destatis, 2016) and by the *European Statistics Code of Practice* for the European Union (Eurostat, 2011). This framework contains major principles, including quality (Principle 4), cost effectiveness (Principle 10), accuracy and reliability (Principle 12) as well as coherence and comparability (Principle 14). For this reasons, the profile of requirements for statistics is versatile. Results have to be provided simultaneously for several variables of interest, which may conflict with one another. Concurrently, the results may have to be determined on various stratification levels of the population, such as for the entire Germany and for all cities separately. In addition to the increased requirements, the amount of available data has dramatically increased within the last decade. Especially the technological progress enables the collection of a huge amount of data within a short period of time and its processing with manageable computational burden.

In particular, the implementation of solution strategies for complex statistical methods may yield high analytical and computational expense. In many cases, these models can be expressed as optimization problems of high dimensions. The solution of these problems often requires certain strategies, whose application may be unsuitable using standard optimization tools due to the computation time needed. To maintain practicability and to improve the computational performance, innovative numerical optimization strategies have to be developed or existing approaches have to be further expanded.

In the context of survey statistics, the statistics are produced using the data collected from a sample of the whole population (cf. Särndal et al., 1992, Chapter 1.2). In this way, the survey process can be divided into two main parts, namely the *selection process* and the *estimation process*. The selection process contains the specification of the rules and operations, according to which the units of the population are determined to be included in the sample (*design stage*) on the one hand, and the drawing procedure of the sample (*sampling stage*) on the other hand. The major task of the estimation process is the computation of the statistics, i.e. the computation of specific point estimates of population or sub-population values, which is also known as *estimation stage* (cf. Kish, 1965). In addition to the point estimation, the evaluation of the quality of the estimates is a further aspect that can be added to the estimation process, which is referred to as *validation stage* in this thesis. Since the quantification of the exact quality is mostly impossible in practice, this also needs to be estimated, which is known as variance estimation (cf. Wolter, 2007, Chapter 1.1).

Each of the two main research questions addressed in this thesis belongs to one of the two parts of the survey statistical process. In addition, both research topics are strongly related to one another for several reasons. Firstly, both topics offer new opportunities for handling complex surveys regarding multivariate objectives, high dimensional auxiliary data, and greatly diverse constraints. Along this line, statistical models are established, and numerical solutions are presented in this thesis in order to generate statistics, which fulfill the multilateral requirements. Secondly, since both topics are constructed on and for similar survey frameworks, they can be applied successively in one survey in order to exploit the potential of the developed approaches. Aside from the similarities in their content, the mathematical structures of the two underlying optimization problems resemble each other, which allows for the application of similar solution strategies.

The thesis is intended to cover relevant aspects and develop strategies to address the two main research questions. The explanation starts with gathering the requirements and circumstances for the desired statistical model. Subsequent to the process of the statistical modeling, the problems need to be formally rewritten to allow for the application of numerical solution strategies. Prior to this, the theory of existence and uniqueness of the solution also needs to be postulated and proved. Tailor-made solvers are then either proposed or extended on the basis of existing algorithms. After the implementation, the models and their solvers are tested, and the results are simulated and analyzed under realistic conditions.

The further proceedings applied in this thesis are briefly sketched for each of the two topics in the following paragraphs with a particular emphasis on the newly developed aspects. The introduction of the definitions and notations as well as a complete overview of the current literature are all found in detail in the respective chapters.

Optimal Multivariate and Multi-domain Allocation

As a part of the selection process in stratified sampling designs, the allocation of the total sample size to the strata of the stratified population is mandatory. In the literature, several methods have been proposed over the past decades, such as the equal and proportional allocations (cf. Lohr, 2009, pp. 104 ff.) as well as the optimal allocations with various strategies based on Tschuprow (1923) and Neyman (1934). With regard to cost efficiency and the accuracy of estimates gained by modern complex and multifarious surveys, the method developed in this thesis needs to consider several aspects. Firstly, the allocation is meant to be optimal for estimates of several variables of interest, which may possibly be unrelated or even mutually contradictory. This results in a multi-criteria optimization problem. Based on the suggestions of Friedrich et al. (2018), the relevance of the variables of interest can be controlled by a few means, such as using various scalarization techniques, certain appropriate decision-making functions, and several standardization techniques. Secondly, the method needs to provide accurate estimates on aggregated stratification levels (such as Germany and its federal states) as well as on disaggregated levels, such as cities or towns. The accuracy of estimates on prioritized levels can be improved by an up-weighting of the corresponding part of the objective, also known as compensatory optimal allocation (cf. Münnich et al., 2012a, pp. 31 ff.). In addition, restrictions for the stratum-specific sample sizes and minimal quality requirements for regional estimates should be included in order to be as flexible as possible. If desired, the stratum-specific sample sizes may be bounded by box-constraints, which has also been suggested by Gabler et al. (2012) and Münnich et al. (2012c) for the univariate case. Finally, the developed model needs to be solved in an appropriate computing time, which requires the application of tailor-made numerical solution strategies. By taking all these aspects into account, the developed method is called *optimal multivariate and multi-domain allocation method*, which is referred to as `MMDopt` (`M`: multivariate, `MD`: multi-domain, `opt`: optimal) hereinafter.

With regard to the requirements mentioned above, the `MMDopt` approach results in the following multi-criteria optimization problem:

$$\begin{aligned}\min_{n\in\mathbb{R}^H_+} \; & \left(\text{Var}(\hat{\tau}^{\text{StrRS}}_{y_1}), \ldots, \text{Var}(\hat{\tau}^{\text{StrRS}}_{y_{q_1}})\right)\\ \text{s.t.} \; & \sum_{h=1}^{H} n_h = n_\text{s}\\ & \text{Var}(\hat{\tau}^{\text{StrRS}}_{y_i l_r}) \leq \text{Vmax}_{(i,r,l_r)}\\ & m \leq n \leq M\end{aligned}$$

In that regard, the variances of the total estimators $\hat{\tau}^{\text{StrRS}}_{y_i}$ of the variables of interest y_i with $i = 1, \ldots, q_1$ are simultaneously minimized with regard to the vector of the stratum-specific sample sizes $n \in \mathbb{R}^H_+$ and the total sample size $n_\text{s} \in \mathbb{N}$. $\text{Vmax}_{(i,r,l_r)}$ respresents predetermined maximal regional estimation errors allowed for the estimation of variable y_i in area or region l_r of stratification level r. While solving this multi-criteria optimization problem, it is mandatory to have a scalarization of the objective functions in order to achieve a real-valued optimization problem. Here, we consider the weighted sum, p-norms (weighted or non-weighted), and the min-max approach (cf. Jahn, 1986, Chapter 5). In the first instance, the weighted sum scalarization is focused on, as it coincides with the theory of Pareto optimization. The dimension of the resulting (weighted) real-valued optimization problem is equal to the number of strata, which tends to be high in several applications. Moreover, the feasible set is built by linear and nonlinear equality and inequality constraints. To avoid the direct solution via a standard solver of nonlinear optimization, we are able to rewrite the problem as a lower dimensional nonlinear system of equations following the derivations for the box-constrained optimal univariate allocation proposed by Gabler et al. (2012) and Münnich et al. (2012c). Nevertheless, it has to be mentioned that the reformulation is only valid for the weighted sum scalarization. Despite the non-differentiability of the reformulated system, it can be solved in an appropriate time by applying a semismooth Newton method (SSN). In that regard, the computational burden is only linearly dependent on the dimension H, which omits an exponential increase of the running time depending on H. As already shown in Friedrich et al. (2018) for a simpler case, the solution of the system for all possible combinations of weights (up to a discretization) provably yields the whole Pareto frontier of the multi-criteria problem. On the one hand, this allows for a comparison of each Pareto optimal solution with one another. On the other hand, the Pareto frontier can be used to solve MMDopt for other scalarizations such as p-norms ($p \neq 1$) and the min-max approach. In general, this is more challenging than for the weighted sum case, given that the objective function is not separable. Instead of the cost-intensive direct solution via standard solvers, this thesis proposes the use of an innovative approach based on a projected inexact quasi-subgradient method (GTM). The iterates of this algorithm are only located in the Pareto frontier of the multi-criteria problem, rather than in the whole feasible set. Due to this significant shrinkage of the feasible set, the computational burden of GTM is clearly lower than using other solvers.

Finally, the proposed method MMDopt and the algorithms SSN and GTM are tested, and the results are computed using a synthetic dataset under realistic circumstances. The thesis also presents further studies of robustness and sensitivity with regard to the quality of the input data. Beyond the statistical properties, the computational expenses of the algorithms SSN and GTM are evaluated and compared with standard solvers.

A Generalized Calibration Method

While the allocation method MMDopt is designed to improve the selection process with regard to versatile requirements such as conflicting objectives, various stratification levels, and diversified restrictions, the second major research objective of the thesis is supposed to improve the estimation process considering similar goals. Regarding this, a generalized calibration method is proposed, which is referred to as GCAL (G: generalized, CAL: calibration) hereinafter. In general, calibration methods deal with the question of how to make use of known auxiliary information to improve the accuracy of the estimates and to achieve coherent results (cf. Kott, 2006 and Särndal, 2007). This is achieved by an adjustment of the a priori given design weights with regard to the auxiliaries. Classical calibration techniques have been proposed by Deville and Särndal (1992). Complex circumstances and the required goals of modern surveys may necessitate further extensions of the existing methods. Firstly, GCAL is designed to deal differently with benchmarks gained by known auxiliary data. For this reason, it needs to be distinguished between a strict compliance with these benchmarks and a weak compliance with a permitted tolerance. This *relaxation* of benchmarks has already been proposed by Guggemos and Tillé

(2010) and Wagner (2013, Chapter 7), and it allows for both coherence and consistency to be achieved between various sources without neglecting disruptive factors, e.g. inaccurate survey data or incorrect register data. Moreover, a high number of benchmarks particularly on different stratification levels can be considered in CGAL, since the relaxation enlarges the feasible set. Secondly, various objective functions are proposed to utilize a penalty structure of the deviation from the design weights, which is well-tailored for the considered application. Finally, box-constraints are added to restrict the deviations of the calibration weights from the design weights and to allow for the control of the variation of the calibration weights.

Generally, the derivations are closely related to the methods proposed by Münnich et al. (2012b) and Burgard et al. (2018). Under the consideration of the mentioned requirements, the resulting optimization problem is given by the following:

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

The problem dimension is equal to the sum of the number of sampled units n_s and the number of relaxed benchmarks q_2. In common surveys, the number of sampled units n_s may in fact exceed the size of one million; an example of this is given by the *German Census 2011* (cf. Münnich et al., 2012a). Usually the number of equality constraints $q_1 + q_2$ (i.e. all benchmarks) is significantly lower than the number of units in the sample. To benefit from the structure of the problem, the problem is rewritten as a lower dimensional nonlinear system of equations in analogy to Münnich et al. (2012b). Similar to the solution strategy of the allocation method MMDopt, GCAL can also be solved with the SSN method. As we will see in the simulation study, the SSN algorithm is able to solve a GCAL problem with a dimension of around $250\,000$ within one second. Moreover, a linear dependence between the sample size and the computation time of SSN can be observed. In contrast to common truncated algorithms (TRUNC), which amongst others can be found in the R package sampling (cf. Tillé and Matei, 2016) and whose structure is based on a similar approach, SSN converges provably to the unique optimal solution.

The variance estimation of GCAL has to handle the relaxation of benchmarks and box-constraints, which is not possible using common variance estimation techniques based on linearization techniques proposed by Deville and Särndal (1992), Estevao and Särndal (2006), and D'Arrigo and Skinner (2010). Instead, a rescaling bootstrap (cf. Preston, 2009) is proposed in analogy to Burgard et al. (2018) for dealing with these extensions. Additionally, this bootstrap enables to provide vectors of replication weights in order to facilitate the variance estimation for GCAL users and to allow further social-scientific studies.

Finally, the proposed method of GCAL is extensively tested on a realistic dataset. In addition to the statistical structures such as point and variance estimates, this study presents computational expenses of the SSN algorithms as well as possibilities for sensitivity analyses. Moreover, the study compares performance of the SSN algorithms to other solvers.

1.2 Outline

This section presents an overview of the chapter contents in this thesis.

Chapter 2: Fundamentals of Survey Statistics
Some basis aspects of the theory of survey statistics are briefly introduced in Chapter 2. In that regard, we focus on the concepts that are relevant for a finite population framework including relevant estimators and common sampling designs. In preparation for the following chapters, classical allocation methods for stratified sampling designs and common calibration techniques are mentioned. Finally, the chapter concludes with the structure of the RIFOSS dataset.

Chapter 3: Fundamentals of Numerical Optimization
In Chapter 3, fundamental frameworks of numerical optimization are presented, which are required in order to develop and solve the statistical models in Chapters 4 and 5. These models contain restricted nonlinear optimization problems for which the optimality theory is introduced. In order to define the `SSN` method, the property of semismooth functions is considered. As a basis for the derivations in Chapter 4, the theory of multi-criteria optimization is explained.

Chapter 4: Optimal Multivariate and Multi-domain Allocation
In Chapter 4, the `MMDopt` method is developed, the discussion of `MMDopt` is embedded in the context of existing methods, and `MMDopt` is extensively analyzed in an application study.

Chapter 5: A Generalized Calibration Method
The `GCAL` method is developed with the major aim to increase the accuracy of the estimates and to ensure coherence between different sources in Chapter 5. Aside from the general method, the study also considers additional aspects such as the variation of the weights and techniques for the variance estimation. Finally, the functionality of `GCAL` is analyzed in a simulation study.

Chapter 6: Conclusion and Outlook
In Chapter 6, the developments of this thesis are summarized and concluded. Advantages and drawbacks are found, both of which are stated and elucidated in the context of survey statistics in modern societies. Moreover, further outstanding and unprocessed research potentials are shortly discussed and further potential scopes of application are explored.

Appendix A, B, and C
An overview of the theory of quality measurements in survey statistics is given in the Appendix A. In Appendix B, additional simulation results are presented, which go beyond the scope of the discussions in Chapters 4 and 5. Finally, the `R`-packages which are still under development are briefly discussed in Appendix C.

To simplify the reading of this thesis, a few conventions are explained here. Different font styles characterize various meanings. Specific expressions and particularly emphasized words are written in *italics*. Abbreviated methods and algorithms are printed in `typewriter font`. Variables of the dataset are written in sans serif letters. In the evaluation of the results of the application and simulation studies, graphics are often indicated by different colors to distinguish between variables of interest (unknown) and auxiliaries (known). The results of the auxiliaries are generally shown in plots with headers shaded in blue and green. By contrast, red and orange shades are chosen for the results of the variables of interest.

Chapter 2

Fundamentals of Survey Statistics

In this chapter, fundamental definitions, notations, and strategies of survey statistics are presented. We mainly focus on aspects which are pertinent to the topics of the thesis while less relevant concepts are only quoted. The exposition begins with some basic remarks on the framework of finite population sampling in Section 2.1. Afterwards, different types of point estimators are explained in Section 2.2, and selected estimators for the estimation of population and sub-population aggregates are introduced. Wide-spread sampling designs are discussed in Section 2.3. In addition to the classical designs, we also focus on balanced sampling designs, where auxiliary information is used within the sampling process to improve the accuracy of the estimates. In Section 2.4, calibration methods are presented whose application is exploited in order to either improve estimates or prevent convergence issues in the context of multiple data sources or multiple surveys. Concerning stratified sampling designs, common allocation techniques are presented in Section 2.5, whose analysis is a principal part of this thesis. The chapter concludes with presenting the RIFOSS dataset of Germany, which is the basis for the application and simulation studies.

2.1 Framework of finite population sampling

The following definitions and notations are based on Cassel et al. (1977), Särndal et al. (1992), and Lohr (2009) unless otherwise stated, and their usage is consistent in the entire thesis.

We assume a fixed and finite population approach, which is also known as a design-based framework following the definitions given by Cassel et al. (1977, Chapter 1). In that regard, the statistical inference depends on the distribution generated by the selection process while the population is treated as a fixed and finite set, i.e. the randomness is induced by the probability whether a unit of the population is selected or not (cf. Lehtonen and Veijanen, 2009, p. 219). The finite population of size $N \in \mathbb{N}$ is assigned with

$$\mathcal{U} := \{1, \ldots, N\}, \tag{2.1}$$

where each unit is uniquely labeled with a fixed and known index $k = 1, \ldots, N$. A set S consisting of units of the population $\mathcal{U}$ is called a *sample*. The size of S is denoted with $n_\mathrm{s} \in \mathbb{N}$. Generally, the techniques for drawing a sample can be divided into techniques *with* and *without* replacement. According to Cassel et al. (1977, Section 1.4), a *without replacement* sampling techniques prohibits a unit $k \in \mathcal{U}$ from being contained more than once in the sample S. In the context of this thesis, without replacement sampling techniques are considered exclusively. Thus, we assume that each unit $k \in \mathcal{U}$ can only occur once in the sample S, i.e. the sample S is a subset of the population $\mathcal{U}$. Hence, S is an element of the set of all possible samples $\mathbb{S} := \{S : S \subseteq \mathcal{U}\}$ with a cardinality of 2^N. Moreover, the set of all possible samples of size n_s is denoted by $\mathbb{S}_{n_\mathrm{s}} := \{S : S \subseteq \mathcal{U} \text{ with } |S| = n_\mathrm{s}\}$. Its cardinality is given by $\binom{N}{n_\mathrm{s}}$ (cf. Lohr, 2009, pp. 30 ff.). If $S = \mathcal{U}$ and therefore $n_\mathrm{s} = N$, the survey is referred to as a *full census*.

As the first step of the selection process, the *sampling design* needs to be specified, defined by a function

$$p(\cdot) : \mathbb{S} \to [0, 1], \tag{2.2}$$

which matches each possible sample $S \in \mathbb{S}$ to a specific probability of being chosen with $\sum_{S \in \mathbb{S}} p(S) = 1$. In a fixed and finite population framework, the sampling design $p(\cdot)$ is the only stochastic element which an inference can be based upon (cf. Cassel et al., 1977, p. 32 and Lehtonen and Veijanen, 2009, p. 219). In analogy to the samples itself, sampling designs can be divided into designs *with* and *without* replacement. Since the without replacement designs are considered in this thesis exclusively, the with replacement designs are not further mentioned. Thus, the formal distinction between *with* and *without* is omitted, meaning that all mentions of sampling designs refer to without replacement designs.

In accordance with the sampling design, each unit k of the population $\mathcal{U}$ is associated a known and strictly positive probability of being selected in the sample (cf. Särndal et al., 1992, p. 32). This probability is known as the (first-order) *inclusion probability* given by

$$\pi_k := \Pr(k \in S) = \sum_{S \in \mathbb{S}} \mathbb{1}_{(k \in S)} p(S) \in \mathbb{R}_+, \tag{2.3}$$

where $\Pr(\cdot)$ denotes the probability and $\mathbb{1}_{(\cdot)}$ is the indicator function. The first-order inclusion probability is the sum over the probabilities of all samples in which unit k is included. Moreover, we refer to the reciprocal of the inclusion probability as the *design weight*, given by

$$d_k := \pi_k^{-1} \in \mathbb{R}_+. \tag{2.4}$$

The design weight is of great importance when using design-based estimators and calibration techniques (see Section 2.4). In analogy to the definition of π_k, the second-order (or joint) inclusion probability

$$\pi_{kl} := \Pr(k \in S \wedge l \in S) = \sum_{S \in \mathbb{S}} \mathbb{1}_{(k \in S)} \mathbb{1}_{(l \in S)} p(S) \in \mathbb{R}_+ \tag{2.5}$$

represents the probability that units k and l are both elements of the sample. We also note that $\pi_{kk} = \pi_k$ and $\pi_{kl} \leq \min\{\pi_k, \pi_l\}$ for all $k, l \in \mathcal{U}$. Regarding Equations (2.3) and (2.5), the definition of the inclusion probabilities depends on the choice of the sampling design $p(\cdot)$. As we will see in Section 2.3, these formulas become significantly simplified if particular sampling designs are considered.

In general, the main purpose of surveys is to derive a statistic ϑ from the population $\mathcal{U}$ (or sub-populations of $\mathcal{U}$) with regard to a *variable of interest*, i.e. a vector

$$y := (y_1, \ldots, y_N)^T \in \mathbb{R}^N \tag{2.6}$$

of parameters of the finite population $\mathcal{U}$. Frequently, this statistic is evaluated with the aim of q *auxiliary variables* or simply *auxiliaries*, denoted by the corresponding q parameter vectors $(x_{i1}, \ldots, x_{iN})^T \in \mathbb{R}^N$ $(i = 1, \ldots, q)$ and assembled to the auxiliary matrix

$$X := [x_1, \ldots, x_N] = \begin{bmatrix} x_{11} & \ldots & x_{1N} \\ \vdots & & \vdots \\ x_{q1} & \ldots & x_{qN} \end{bmatrix} \in \mathbb{R}^{q \times N}. \tag{2.7}$$

In some instances, they are used as individual auxiliary vectors $x_k := (x_{1k}, \ldots, x_{qk})^T \in \mathbb{R}^q$ for all units $k = 1, \ldots, N$. The statistics of the auxiliary variables are generally assumed to be available prior to the sampling process, i.e. non-response or similar issues are neglected in the context of this thesis. By contrast, the statistic of the variable of interest y is not given in advance. Regarding this point, the main task in the estimation process is to evaluate an *estimate* $\hat{\vartheta}(S)$ of a specific statistic ϑ of the variable of interest y, which only requires sample information y_k, $k \in S$. For this reason, the measurable function $\hat{\vartheta} : \mathbb{S} \to \mathbb{R}$, which maps the space of all samples to the real values, is called an *estimator* of ϑ. For a more detailed analysis of the definition of estimators, the work of Witting (1978, Chapter 1.2 and 1.3) and Durrett (2010, Chapter 1.3) provide the necessary references and groundwork.

Following the principles of Särndal et al. (1992, Section 2.7), the expectation of the estimator $\hat{\vartheta}$ is given by

$$\mathrm{E}(\hat{\vartheta}) := \sum_{S \in \mathbb{S}} p(S) \hat{\vartheta}(S), \tag{2.8}$$

where the estimate $\hat{\vartheta}(S)$ is the specific value of the estimator concerning sample S. Thus, the expected value of $\hat{\vartheta}$ is computed as an average of all possible values of $\hat{\vartheta}(S)$, weighted with the probabilities $p(S)$. The *variance* of an estimator is given by

$$\mathrm{Var}(\hat{\vartheta}) := \sum_{S \in \mathbb{S}} p(S) \Big(\hat{\vartheta}(S) - \mathrm{E}(\hat{\vartheta}) \Big)^2. \tag{2.9}$$

Informally, the variance is the expectation of the squared deviation of a random variable from its mean (cf. Durrett, 2010, p. 32). The square root of the variance is called *standard deviation*, formally given by

$$\mathrm{sd}(\hat{\vartheta}) := \sqrt{\mathrm{Var}(\hat{\vartheta})}. \tag{2.10}$$

The variance and the standard deviation are measurements for the *accuracy* and *preciseness* of an estimator. Beyond this, we define the *bias* of an estimator $\hat{\vartheta}$ by

$$\mathrm{Bias}(\hat{\vartheta}) := \vartheta - \mathrm{E}(\hat{\vartheta}). \tag{2.11}$$

If $\mathrm{Bias}(\hat{\vartheta}) = 0$, the estimator is called *unbiased*. An estimator is *asymptotically unbiased* in a finite population framework, if $\mathrm{Bias}(\hat{\vartheta}) = 0$ holds at least for assuming that the population size and sample size tend to infinity. As the expression *bias* suggests, the bias is a parameter to measure the *distortion* of the estimated value from the real value. As a combination of accuracy and distortion, the *efficiency* of an estimator is commonly measured by the *mean squared error (MSE)*, which additively combines the variance and the squared bias

$$\mathrm{MSE}(\hat{\vartheta}) := \mathrm{Var}(\hat{\vartheta}) + \mathrm{Bias}(\hat{\vartheta})^2 = \sum_{S \in \$} p(S)\Big(\hat{\vartheta}(S) - \vartheta\Big)^2. \tag{2.12}$$

If $\hat{\vartheta}$ is an unbiased estimator for ϑ, it follows from Equation (2.12) that $\mathrm{MSE}(\hat{\vartheta}) = \mathrm{Var}(\hat{\vartheta})$. We remark that the estimator $\hat{\vartheta}$ may be an estimator for the whole population, but it may also be for a sub-population. To make variances of various variables of interests with highly different scales comparable, the *coefficient of variation*

$$\mathrm{cv}(\hat{\vartheta}) := \frac{\sqrt{\mathrm{Var}(\hat{\vartheta})}}{\vartheta} \tag{2.13}$$

is commonly used for standardizing the standard deviation of an estimator $\hat{\vartheta}$ (if $\vartheta > 0$). Usually, the estimator $\hat{\vartheta}$ is an estimator for the total τ_y or the mean μ_y of the variable of interest y, but other types of estimators are possible. Since the estimator for the mean can easily be derived from the total estimator, we mainly focus on total estimators within this thesis. Nevertheless, all methods proposed are similarly applicable for the estimator of the mean. For notational simplicity, the estimate $\hat{\tau}_y(S)$ of the total τ_y computed regarding sample $S \in \$$ is denoted as $\hat{\tau}_y$ hereafter, so that $\hat{\tau}_y$ refers to the estimator as well as the specific estimate with regard to sample $S \in \$$.

2.2 Selected estimators

In this section, two common estimators are presented which are vital for this thesis. For a broader overview, we refer to Särndal et al. (1992). In general, the focus is on the estimation of the population total

$$\tau_y := \sum_{k \in \mathcal{U}} y_k \tag{2.14}$$

of the variable of interest y for population $\mathcal{U}$. The estimator of τ_y is denoted by $\hat{\tau}_y$. The estimator of a sub-population total

$$\tau_{yd} := \sum_{k \in \mathcal{U}_d} y_k \tag{2.15}$$

for an arbitrary subset $\mathcal{U}_d$ of the population $\mathcal{U}$ is analogously defined by $\hat{\tau}_{yd}$, whereby an index is added for the considered sub-population. The estimators $\hat{\mu}_y$ and $\hat{\mu}_{yd}$ of the corresponding means μ_y and μ_{yd} can be derived by dividing the total estimates by the size of the (sub-)population. The sample S is drawn under consideration of a specific sampling design $p(\cdot)$ without replacement (see Equation (2.2)), which determines the inclusion probabilities π_k for all $k \in S$ (see Equation (2.3)). For a definition of the respective specific sampling designs, we refer to Section 2.3.

In general, we distinguish between *direct* and *indirect* estimators. Direct estimators only incorporate the information available from inside the population of interest itself. On the other hand, an indirect estimator utilizes additional information from outside the population of interest as well to improve the estimation (cf. You and Rao, 2002, p. 431). This particular utilization referred to as *borrowing strength* may be especially reasonable if the size of the population of interest or the respective sample size is comparatively small (cf. Rao and Molina, 2015, p. 1). Such indirect estimation techniques are not further considered in this thesis, however, small area estimates may be incorporated as calibration benchmarks in the generalized calibration method in Chapter 5. In addition to distinction between direct and indirect estimators, they can be grouped into three types: design-based estimators, model-assisted estimators, and model-based estimators. In the following, two common estimators are presented for various sampling designs $p(\cdot)$ without replacement, which belong to the class of design-based and model-assisted estimators respectively. After this, the characteristics of model-based estimators are briefly sketched without details, since they are not the primary focus of this thesis.

Horvitz-Thompson estimator

Design-based estimators do not explicitly use auxiliary information in the estimation process. Nevertheless, some auxiliary information may be incorporated in the design $p(\cdot)$. The most popular design-based estimator for the population total (2.14) is the *Horvitz-Thompson estimator* (HT estimator), defined by

$$\begin{aligned}\hat{\tau}_y^{\mathrm{HT}} :&= \sum_{k\in\mathcal{U}} \mathbb{1}_{(k\in S)}\frac{y_k}{\pi_k} = \sum_{k\in\mathcal{U}} \mathbb{1}_{(k\in S)} d_k y_k \\ &= \sum_{k\in S}\frac{y_k}{\pi_k} = \sum_{k\in S} d_k y_k.\end{aligned} \tag{2.16}$$

It is built as the sum of the values of the variable of interest over the sampled units weighted by their respective design weights, which have been determined by means of the sampling design $p(\cdot)$. The HT estimator is design-unbiased for τ_y, which can be proved applying the definitions of the inclusion probability (2.3) and the expected value (2.8) (cf. Horvitz and Thompson, 1952 and Särndal et al., 1992, Section 2.8). The HT estimator is a direct estimator, as it does not require any information from outside the sample, e.g. from other sources such as administrative data or other surveys. In the presence of strongly correlated auxiliary data, the HT estimator may have a higher variance in comparison to estimators that utilize auxiliary information by means of an assisting model, since the HT estimator does not make use of them at the estimation stage (cf. Särndal et al., 1992, pp. 219 ff.). Nevertheless, the application of the HT estimator is, not least due to its simplicity, widespread in the field of survey statistics and may work well in many situations. For comparably high sampling fractions in particular, the HT estimator yields generally accurate point and variance estimates (cf. Münnich et al., 2012a, p. 40). The variance of (2.16) can be derived using the fact that the estimator is a linear function of the random variables $\mathbb{1}_{(k\in S)}$ and this is then given by

$$\mathrm{Var}(\hat{\tau}_y^{\mathrm{HT}}) = \sum_{k\in\mathcal{U}}\sum_{l\in\mathcal{U}}\left(\frac{\pi_{kl}}{\pi_k\pi_l} - 1\right) y_k y_l. \tag{2.17}$$

Moreover, based on Särndal et al. (1992, p. 43), the variance can be unbiasedly estimated by the residual variance estimator

$$\widehat{\text{Var}}(\hat{\tau}_y^{\text{HT}}) = \sum_{k \in S} \sum_{l \in S} \left(\frac{\pi_{kl}}{\pi_k \pi_l} - 1 \right) \frac{y_k y_l}{\pi_{kl}}, \tag{2.18}$$

which prevents the need for resampling strategies for the variance and MSE estimation. Nevertheless, the computation of both (2.17) and (2.18) may suffer for general sampling designs, since the second-order inclusion probabilities π_{kl} are required. However, Section 2.3 shows that the formulas can be simplified under specific sampling designs like stratified random sampling.

Generalized regression estimator

As mentioned, in the case of known auxiliary data, design-based estimators may be inefficient due to the neglecting of auxiliary information at the estimation stage. Procedures that make use of potentially correlated auxiliary information within a model to reduce the variance compared to design-based estimators are called *model-assisted*, if their design-based properties are not dependent upon the validity of the model (cf. Särndal et al., 1992, Remark 6.4.1). The most prominent example of such estimators is the *generalized regression estimator* (GREG estimator), which was primarily published by Cassel et al. (1976). It is given by

$$\hat{\tau}_y^{\text{GREG}} := \hat{\tau}_y^{\text{HT}} + \hat{\beta}^T (\tau_X - \hat{\tau}_X^{\text{HT}}), \tag{2.19}$$

whereby $\hat{\beta} \in \mathbb{R}^q$ is the vector of the estimated regression coefficients between the q auxiliary variables X and the variable of interest y. Since generally the variable of interest and the auxiliary information are only available for the sampled units $k \in S$, the regression coefficient has to be estimated by

$$\hat{\beta} = \left(\sum_{k \in S} d_k x_k x_k^T \right)^{-1} \sum_{k \in S} d_k x_k y_k. \tag{2.20}$$

Its population-based value $\beta \in \mathbb{R}^q$ is analogously defined by the sum over $k \in U$ instead of $k \in S$. With regard to Equation (2.20), the GREG estimator consists of the HT estimator corrected by a linear assisting model depending on the correlation of the variable of interest y and the auxiliary variables X (cf. Särndal et al., 1992, Chapter 6 and 7). We also note that $\tau_X \in \mathbb{R}^q$ and $\hat{\tau}_X^{\text{HT}} \in \mathbb{R}^q$ are the vector-valued equivalents to the totals and the estimated totals of the q auxiliary variables respectively. Compared to the HT estimator (2.16), the efficiency of the GREG estimator (2.19) decisively depends on the goodness of the underlying fit between the auxiliaries and the variable of interest generated by the assisting model. Since this is a linear regression model, the goodness is generally strong, if the correlations between the auxiliaries and the variable of interest are high (cf. Särndal et al., 1992, pp. 227 ff.). While the goodness of the model does affect the accuracy of the estimate, basic properties such as the asymptotic design-unbiasedness (which holds under weak assumptions; cf. Cassel et al., 1976) or the validity of the variance formulas are not dependent on the strength of the model. Thus, the GREG estimator is given the term model-assisted. Moreover, the GREG estimator can either be direct or indirect depending on the level which the assisting regression model is fitted at, i.e.

the level at which the vector $\hat{\beta}$ is computed. If the model is fitted only within the population of interest, the estimator (2.19) is a direct estimator. Alternatively, if additional information from beyond the bounds of the population of interest is used to fit the model, (2.19) is called an indirect estimator. This borrowing of strength might be especially useful, if the population of interest or its sample size is relatively small. In this case, the computation of the regression coefficient $\hat{\beta}$ may be unstable, as the number of sampled units is not large enough to allow an accurate fit to the considered covariables. The variance of the GREG estimator cannot be explicitly formulated due to the complex nature of the estimator. Hence, the variance needs to be approximated by applying Taylor approximations as shown in Särndal et al. (1992, Result 6.6.1). The approximated variance is given by

$$\text{Var}(\hat{\tau}_y^{\text{GREG}}) = \sum_{k \in \mathcal{U}} \sum_{l \in \mathcal{U}} \left(\frac{\pi_{kl}}{\pi_k \pi_l} - 1 \right) (y_k - x_k^T \beta)(y_l - x_l^T \beta), \tag{2.21}$$

whereby $\beta \in \mathbb{R}^q$ is computed over all units of the population $k \in \mathcal{U}$ in analogy to the sample-based computation of $\hat{\beta}$ defined in (2.20). In that regard, the terms $(y_k - x_k^T \beta)$ represent the residuals. The corresponding residual variance estimator is defined as

$$\widehat{\text{Var}}(\hat{\tau}_y^{\text{GREG}}) = \sum_{k \in S} \sum_{l \in S} \left(\frac{\pi_{kl}}{\pi_k \pi_l} - 1 \right) \frac{(y_k - x_k^T \hat{\beta})(y_l - x_l^T \hat{\beta})}{\pi_{kl}}, \tag{2.22}$$

where only sample information is used to compute the variance and the regression coefficient $\hat{\beta}$. As specified in Särndal et al. (1992, Chapter 6.5), the GREG estimator has several equivalent expressions apart from (2.19). We will focus on one alternative expression in Section 2.4, which will build a bridge between the GREG estimator and the calibration techniques.

Model-based estimators

In model-based estimation, implicit or explicit models are commonly applied to borrow strength by using auxiliary information from outside the population of interest. In contrast to model-assisted estimators, the properties of model-based estimators are strictly dependent on the validity of the model (cf. Särndal et al., 1992, Remark 6.4.1). Hence, the validity of the underlying model has to be assumed and checked by means of the data. In addition, the property of design-unbiasedness is not given in general, meaning that the occurrence of a systematic bias is possible. Model-based techniques are popular in the context of small area estimation, where linear mixed models are commonly applied to borrow strength from outside the population of interest (cf. Rao and Molina, 2015, p. 2). We will not focus on these model-based techniques in this thesis, however, small area estimates may be incorporated as calibration benchmarks in the generalized calibration method presented in Chapter 5. Thus, a short overview of the basic literature is provided here. Model-based estimators in the context of small area estimation have been published by Rao and Molina (2015). Moreover, Fay and Herriot (1979), Battese et al. (1988), and You and Rao (2002) have presented examples of empirical best linear unbiased predictors. According to these studies, the use of point estimates is often straightforward, while the evaluation of the quality of these estimates may be very sophisticated. Thus, methods for variance and MSE estimations are most likely based on approximation techniques or resampling methods. Nevertheless, small area estimation has been a widely investigated topic in the research field over the past few years, and it is steadily becoming more applied in official statistics, as its potential for improvement is not negligible.

2.3 Common sampling designs

As defined in Section 2.1, the procedure of drawing a sample is based on the sampling design $p(\cdot) : \mathbb{S} \to [0, 1]$, which matches each possible sample $S \in \mathbb{S}$ to a specific probability of being chosen. In general, we make a distinction between *equal* probability sampling and *unequal* probability sampling according to equal or unequal inclusion probabilities for all units of the population (cf. Lohr, 2009, p. 23 and p. 179). In this section, we briefly introduce the sampling designs which are relevant for this thesis, in particular simple random sampling and stratified random sampling. Since these two designs are the only designs considered in this thesis, we refer to Cochran (1977) and Lohr (2009, Chapters 2 to 6) for a further overview of classical sampling designs.

Simple random sampling

In *simple random sampling* without replacement (SRS), a sample of size n_s is drawn in a way where every possible subset of n_s distinct units of the population has the same probability of being selected as the sample, i.e.

$$p(S) = \frac{1}{\binom{N}{n_\mathrm{s}}} \text{ for all } S \in \mathbb{S}_{n_\mathrm{s}}. \tag{2.23}$$

Consequently, SRS in an equal probability design, since the inclusion probabilities are constant for all units of the population, i.e.

$$\pi_k = \frac{n_\mathrm{s}}{N} \; \forall k = 1, \ldots, N. \tag{2.24}$$

In general, a distinction is made between simple random sampling *with* replacement and *without* replacement (cf. Lohr, 2009, Chapter 2.3). Following our definition of a sample $S \in \mathbb{S}_{n_\mathrm{s}}$ in Section 2.1, SRS is exclusively referred to the without replacement design, since it is more common in household and business surveys. Since π_k is constant in SRS, both (2.17) and (2.18) as well as (2.21) and (2.22) of the variances of the presented estimators in Section 2.2 are simplified under the SRS design. The variance of the HT estimator of the population total of variable y under SRS is then given by

$$\mathrm{Var}(\hat{\tau}_y^{\mathrm{SRS,HT}}) = \frac{S^2 N^2}{n_\mathrm{s}} \left(1 - \frac{n_\mathrm{s}}{N}\right), \tag{2.25}$$

where

$$S^2 = \frac{1}{N-1} \sum_{k \in \mathcal{U}} \left(y_k - \bar{y}\right)^2 \tag{2.26}$$

is the population variance of variable of interest y with the population mean $\bar{y} = \frac{1}{N}\sum_{k\in\mathcal{U}} y_k$ (cf. Cochran, 1977, p. 90). In general, S^2 is not known in advance and needs to be estimated using

$$s^2 = \frac{1}{n_\mathrm{s}-1}\sum_{k\in S}\left(y_k - \bar{y}_S\right)^2 \tag{2.27}$$

with the sample mean $\bar{y}_S = \frac{1}{n_\mathrm{s}}\sum_{k\in S} y_k$. Hence, the variance can be estimated by

$$\widehat{\mathrm{Var}}(\hat{\tau}_y^{\mathrm{SRS,HT}}) = \frac{s^2 N^2}{n_\mathrm{s}}\left(1 - \frac{n_\mathrm{s}}{N}\right). \tag{2.28}$$

The approximated variance of the GREG estimator $\mathrm{Var}(\hat{\tau}_y^{\mathrm{SRS,GREG}})$ under SRS equals Equation (2.25), whereas S^2 is exchanged by

$$S_\mathrm{e}^2 = \frac{1}{N-1}\sum_{k\in\mathcal{U}}\left(y_k - x_k^T\beta\right)^2 \tag{2.29}$$

with the regression coefficient β (cf. Lohr, 2009, pp. 372 ff.). If S_e^2 needs to be estimated, an estimator is given by

$$s_\mathrm{e}^2 = \frac{1}{n_\mathrm{s}-1}\sum_{k\in S}\left(y_k - x_k^T\hat{\beta}\right)^2 \tag{2.30}$$

with the estimated regression coefficient $\hat{\beta}$ defined in Equation (2.20). Thus, the estimated variance $\widehat{\mathrm{Var}}(\hat{\tau}_y^{\mathrm{SRS,GREG}})$ is given by Equation (2.28) with (2.30) instead of (2.27).

Despite its simplicity, SRS is rarely used in modern surveys in official statistics with design-based estimation strategies for several reasons. Firstly, with auxiliary information neglected at the design stage, the opportunity is missed to generate efficiency gains of the estimates through a *smart* structuring of the population, such as by incorporating geographic regions in the design (cf. Lehtonen and Veijanen, 2009, pp. 222 f.). Secondly, the sampled units are randomly distributed over the whole population, such that the survey costs might be unacceptable high, given that is is necessary for interviewers to cover huge regions. A more detailed discussion about this issue is given in Lohr (2009, Chapter 5) in the context of cluster sampling. Moreover, since regional estimates are desired, SRS implies the presence of unplanned regions, as the region-specific sample sizes are not known at the estimation stage, which is an assumption for the validity of the variance formulas of the HT and the GREG estimator. Nevertheless, SRS plays an important role in multi-stage designs such as the two-stage cluster design, where SRS can be applied to sample clusters at the first stage and to sample units within each cluster at the second stage (cf. Lohr, 2009, Chapter 5.3). Since SRS may be the most intuitive design, the accuracy of an estimate with another design is often measured in relation to the accuracy of an estimate under SRS.

Stratified random sampling

In Section 2.2, common estimators have been introduced for the total of population $\mathcal{U}$ and the total of an unspecific sub-population $\mathcal{U}_d$. Henceforth, the definition of such sub-populations will be specified by a stratification of the population substantiated by the *stratified random sampling* *design* (StrRS; cf. Lohr, 2009, Chapter 3). In that regard, the population is divided into H pairwise disjoint and exhaustive strata $\mathcal{U}_h$ and stratum sizes $N_h \in \mathbb{N}$ with

$$\mathcal{U} = \bigcup_{h=1}^{H} \mathcal{U}_h. \tag{2.31}$$

Hence, each unit of the population belongs to exactly one stratum. Since the focus is on a design-based context, we assume that the stratum sizes are known at the design stage, which implies that the H strata are determined in the sampling design. In light of the fact that we apply the developed statistical methods on the whole population, on each specific stratum as well as on other subsets of the whole population, any unions of some strata h are referred to as *areas* from this point onward. To be more precise, areas can be defined on different stratification levels $r = 1, \dots, R$, where L_r areas exists for each level r. The areas on stratification level $r \in \{1, \dots, R\}$ are denoted with $\mathcal{U}_{l_r}^{(r)}$ ($l_r \in \{1, \dots, L_r\}$). Moreover, the areas of each level exhaust the whole population, i.e.

$$\mathcal{U} = \bigcup_{l_r=1}^{L_r} \mathcal{U}_{l_r}^{(r)} \text{ for all } r = 1, \dots, R. \tag{2.32}$$

It should be noted that these areas are also disjoint among one another. Since all the areas are unions of the strata, they are also considered in the sampling design. This is important for the validity of the variance formulas of area-specific HT and GREG estimates (unplanned strata or areas are omitted in this way). The stratification structure is exemplarily illustrated in Figure 2.1 for $R = 3$, $L_1 = 2$, $L_2 = 3$, $L_3 = 3$, and $H = 10$. Each color (red, blue, and green) corresponds to one stratification level. The areas of one stratification level are consecutively numbered and the numbers are plotted with the respective color. The numbers of the 10 strata are plotted in black.

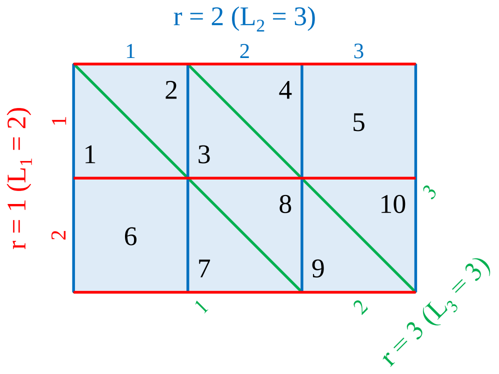


Figure 2.1: Example for a multi-domain stratification ($R = 3$, $L_1 = 2$, $L_2 = 3$, $L_3 = 3$ and $H = 10$).

In analogy to the population, a sample $S \in \mathbb{S}_{n_s}$ of the size $n_s \leq N$ and can be divided into disjoint stratum-specific samples S_h with

$$S = \bigcup_{h=1}^{H} S_h \tag{2.33}$$

and $S_h = S \cap \mathcal{U}_h$ (cf. Lohr, 2009, Chapter 4). The stratum-specific sample sizes n_h are stated within the vector

$$n := (n_1, \ldots, n_H)^T, \tag{2.34}$$

where the sum over all components is equal to the total sample size, i.e. $\sum_{h=1}^{H} n_h = n_{\text{s}}$. Similar to (2.32), area-specific samples are denoted with $S_{l_r}^{(r)}$. Moreover, $S_{l_r}^{(r)} = S \cap \mathcal{U}_{l_r}^{(r)}$ holds for all $l_r \in \{1, \ldots, L_r\}$ and $r \in \{1, \ldots, R\}$.

Using StrRS, a SRS design is conducted separately within each stratum h, where n_h units are drawn from the h^{th} stratum. In this way, the inclusion probabilities are given by

$$\pi_k = \frac{n_h}{N_h} \; \forall k \in \mathcal{U}_h, \tag{2.35}$$

i.e. the inclusion probabilities are constant within each stratum. In contrast to SRS, StrRS ensures that units from all strata are present in the sample (subject to the condition that $n_h > 0$). Moreover, StrRS enables the computation of stratum- or area-specific estimates and their variances since, in contrast to SRS, the stratum- and area-specific sample sizes are known at the estimation stage. The HT estimator (2.16) and the GREG estimator (2.19) are applicable with StrRS for the estimation of population as well as of stratum- and area-specific totals (see inter alia Särndal et al., 1992, Result 3.7.2). Since we assume StrRS, the variance of the HT estimator for the population total $\tau_y = \sum_{k \in \mathcal{U}} y_k$ is defined as the sum of the stratum-specific variances of the HT estimator, given by

$$\text{Var}(\hat{\tau}_y^{\text{StrRS,HT}}) = \sum_{h=1}^{H} \frac{S_h^2 N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right). \tag{2.36}$$

It can be estimated by

$$\widehat{\text{Var}}(\hat{\tau}_y^{\text{StrRS,HT}}) = \sum_{h=1}^{H} \frac{s_h^2 N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right), \tag{2.37}$$

where S_h^2 and s_h^2 are the stratum-specific variance of variable y and its estimate in stratum h in analogy to (2.26) and (2.27) respectively (Lohr, 2009, p. 100). An expression for the variance of the GREG estimator under StrRS can be formulated by (2.36) and (2.37), with a replacement of S_h^2 and s_h^2 in analogy to (2.29) and (2.30).

The definition of the strata is crucial for the efficiency of the estimates under a StrRS design. As illustrated by Särndal et al. (1992, Example 3.7.1), the variance reduction of the estimate for τ_y compared to SRS increases with an increasing homogeneity of the strata with respect to the variable of interest y. If homogeneity increases in stratum h, the stratum-specific variance S_h^2 decreases, which leads to a smaller variance of the estimate. However, it is not possible to construct strata based on the variable of interest y, since its information is collected by means of the survey. Thus, external auxiliary information or proxies of y, such as those obtained from previous surveys or registers, are used in practice to build the strata. Often, regional properties of the population and individual properties of the units are simultaneously assessed to construct the strata. In the *German Census 2011* (where each address in Germany is a unit of the population), the strata are built as cross-classifications between sampling points (SMP; regional areas) and classes depending on the registered inhabitants within an address (areas by content) (cf. Münnich et al., 2012a, p. 31). In business surveys, the company size and the industry sector of the company are frequently used in addition to some regional areas in order to build the cross-classification strata (cf. Hidiroglou and Lavallée, 2009).

One main task of the application of a stratified sampling design is the allocation of the total sample size to the strata, which is discussed in more detail in Section 2.5.

Other sampling designs

In addition to SRS and StrRS, *cluster sampling* is also popular in modern survey sampling. One major type of cluster sampling can be characterized as a multi-stage SRS, where groups of units (clusters) are sampled in the first stage(s). The units of the population are only sampled at the last stage. Cluster sampling is not further considered in this thesis, we refer to Lohr (2009, Chapter 5) for more information on this topic. Another common design is the *balanced sampling* design, which makes use of auxiliary data at the sampling stage (cf. Tillé, 2006, Chapter 8). It has the property that the HT estimators of the totals for a set of auxiliary variables are equal to the totals that have to be estimated. For more information about balanced sampling and its implementation, we refer to Deville and Tillé (2004) and Tillé (2006, Chapter 8).

2.4 Calibration of survey weights

The term *calibration* is widely used, and in general it has different meanings in various sciences. In survey statistics, it is considered in the context of calibration of estimator weights. Around a decade ago, Särndal (2007) praised calibration as "an important methodological instrument in large-scale production of statistics". The calibration approach given by Särndal is defined as follows.

Definition 2.4.1 (calibration approach)**.** The calibration approach to estimation for finite populations consists of

(a) a computation of weights that incorporate specified auxiliary information and are restrained by calibration equation(s),

(b) the use of these weights to compute linearly weighted estimates of totals and other finite population parameters: weight times variable value, summed over a set of observed units,

(c) an objective to obtain nearly design unbiased estimates as long as nonresponse and other non-sampling errors are absent.

Other definitions can also be found in the literature. Ardilly (2006) defined calibration similarly to Definition 2.4.1 (a), namely as a re-weighting method to adjust the design weights with regard to several auxiliary variables. Kott (2006) described calibration weights as a set of weights that satisfy known population totals. Moreover, Kott characterized the calibration estimator as design consistent, i.e. the design bias is, under weak conditions, asymptotically insignificant for the MSE of the estimator. Finally, a proper summary of all these definitions is given in Statistics

Canada (2003, pp. 45-46): “Calibration is a procedure that can be used to incorporate auxiliary data. This procedure adjusts the sampling weights by multipliers known as calibration factors that make the estimates agree with known totals. The resulting weights are called calibration weights or final estimation weights. These calibration weights will generally result in estimates that are design consistent, and that have a smaller variance than the Horvitz-Thompson estimator”. In addition to these characterizations, Merkouris (2004) delineated calibration as a powerful tool to achieve coherent estimates, for example between several survey, for known administrative data, and on different levels (e.g. unit and household level). This property is highly relevant for official statistics (cf. Riede et al., 2013). Concerning these definitions, we are able to define a standard calibration estimator.

Definition 2.4.2 (calibration estimator)**.** Let the units $k \in S$ be numbered consecutively with $k = 1, \dots, n_\text{s}$ without loss of generality. Then, the estimator for the total τ_y of variable y

$$\hat{\tau}_y^{\text{CAL}} := \sum_{k \in S} d_k g_k y_k \tag{2.38}$$

with g_k $(k = 1, \dots, n_\text{s})$ is called *calibration estimator* with regard to a convex and continuously differentiable distance function $D : \mathbb{R} \to \mathbb{R}_{0+}$, if the vector $g = (g_1, \dots, g_{n_\text{s}})^T \in \mathbb{R}^{n_\text{s}}$ is the unique optimal solution of the optimization problem

$$\begin{aligned} \min_{g \in \mathbb{R}^{n_\text{s}}} \; & \sum_{k \in S} d_k D(g_k) \\ \text{s.t.} \; & \sum_{k \in S} d_k g_k x_k = \sum_{k \in \mathcal{U}} x_k . \end{aligned} \tag{2.39}$$

The components g_k of the solution g are then called *correction weights* (also known as *g-weights*). Furthermore, the products $w_k := d_k g_k$ are called *calibration weights*.

The distance function D in Definition 2.4.2 quantifies the amount of penalization when the calibration weights $w_k = d_k g_k$ differ from the design weights d_k. Some traditional choices for D are shown in Table 2.1. These and other distance functions are listed in Deville and Särndal (1992), Singh and Mohl (1996), and Stukel et al. (1996). A detailed comparison of the distance functions is given in the simulation study in Subsection 5.6.7.

Table 2.1: 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	$g_k - 1 - \log(g_k)$

In comparing the calibration estimator (2.38) with the GREG estimator (2.19), similarities can be observed. Indeed, we can prove equivalence between (2.38) and (2.19) for the estimation of the population total if the GREG-type distance function is chosen (cf. Deville and Särndal, 1992). Prior to that, Lemma 2.4.3 is proved.

Lemma 2.4.3. The GREG estimator (2.19) is equivalent to the calibration estimator (2.38), if

$$g_k = 1 + \left(\tau_X - \hat{\tau}_X^{\mathrm{HT}}\right)^T \left(\sum_{l \in S} d_l x_l x_l^T\right)^{-1} x_k \tag{2.40}$$

for each $k \in S$.

Proof. Using (2.19) and (2.20), we can compute

$$\begin{aligned}
\hat{\tau}_y^{\mathrm{GREG}} &= \hat{\tau}_y^{\mathrm{HT}} + (\tau_X - \hat{\tau}_X^{\mathrm{HT}})^T \hat{\beta} \\
&= \sum_{k \in S} d_k y_k + \left((\tau_X - \hat{\tau}_X^{\mathrm{HT}})^T \left(\sum_{l \in S} d_l x_l x_l^T\right)^{-1} \sum_{k \in S} d_k x_k y_k \right) \\
&= \sum_{k \in S} \underbrace{\left(1 + (\tau_X - \hat{\tau}_X^{\mathrm{HT}})^T \left(\sum_{l \in S} d_l x_l x_l^T\right)^{-1} x_k \right)}_{= g_k} d_k y_k \\
&= \sum_{k \in S} d_k g_k y_k = \hat{\tau}_y^{\mathrm{CAL}}
\end{aligned}$$

with g_k defined by (2.40). □

Using the characterization of the correction weights g_k given in Lemma 2.4.3, the equivalence between (2.38) and (2.19) for the estimation of population total can be proved.

Theorem 2.4.4. The GREG estimator (2.19) is equivalent to the calibration estimator (2.38), if $D(g_k) = \frac{1}{2}(g_k - 1)^2$ for all $k = 1, \ldots, n_{\mathrm{s}}$.

Proof. Since the objective function of (2.39) is strictly convex, solving the Karush-Kuhn-Tucker (KKT) conditions

$$d_k(g_k - 1) + d_k \lambda^T x_k = 0 \;\; \forall k = 1, \ldots, n_{\mathrm{s}} \tag{2.41}$$

$$\sum_{k \in S} d_k g_k x_k - \sum_{k \in \mathcal{U}} x_k = 0 \tag{2.42}$$

is necessary and sufficient for the evaluation of the unique optimal solution of (2.39) (Theorem 3.1.7 or Geiger and Kanzow, 2002, Definition 2.35) with Lagrangian multipliers $\lambda \in \mathbb{R}^q$. We transform (2.41) to

$$g_k = 1 - \lambda^T x_k \tag{2.43}$$

and place it into (2.42). After some transformations, the following is obtained:

$$\lambda^T = -\left(\tau_X - \hat{\tau}_X^{\mathrm{HT}}\right)^T \left(\sum_{k \in S} d_k x_k x_k^T\right)^{-1}. \tag{2.44}$$

The insertion of (2.44) into (2.43) shows the equivalence of (2.40) and (2.43), which completes the proof. □

Theorem 2.4.4 shows, that the GREG estimator for the population total can be interpreted as calibration estimator with the auxiliary variables X as calibration benchmarks. Moreover, the calibration estimator defined in Definition 2.4.2 is a general case of the GREG estimator, since Definition 2.4.2 is valid for various distance functions. Alternative formulations of the GREG estimator and its connection to the field of calibration techniques are extensively discussed in the literature. Since this is not considered to be a main research question within this thesis, we refer to the references Deville and Särndal (1992, Chapter 1), Cassel et al. (1976), and Zieschang (1990, Chapter 3) for a detailed overview of this topic.

In comparing balanced sampling (cf. Tillé, 2006, Chapter 8) with the GREG estimator (2.19) and calibration techniques, the methods differ in their handling of the auxiliary data. A balanced sampling approach can be interpreted as an *a priori calibration strategy*, since auxiliary data is already used at the design and sampling stage. On the other hand, the application of the model-assisted GREG estimator (2.19) and the calibration estimator (2.38) can be expressed as an *a posteriori balancing strategy*, since the auxiliary data is not used before the estimation stage. A comparative glance at both strategies yields no general statement whether the usage of auxiliary information is already sensible at the design stage, since the effect strongly depends on the specific application. Basically, calibration techniques are preferred in practice, since the auxiliary information needs to be available not before the estimation stage. By contrast, balanced sampling requires the auxiliaries to be available already at the sampling stage. For a more detailed discussion concerning calibration and balanced sampling, we refer to Tillé (2006, Chapters 8 and 9).

One major aim of this thesis is the development of a more generalized form of the calibration estimator (see Definition 2.4.2) in order to gain more flexibility and usability in real applications, especially in official statistics. As explored in Chapter 5, where we discuss the additional consideration of box-constraints, regional or area-specific benchmarks, and the relaxation of specific benchmarks. Moreover, the control of the variation of the calibration weights is discussed, which can be measured by the ratio of the largest to the smallest calibration weight (cf. Gelman, 2007). This ratio is referred to as *Gelman bound* (or Gelman factor) from this point onward in accordance with Münnich and Burgard (2012).

2.5 Allocation methods

In Section 2.3, the StrRS design was introduced as one of the most common sampling designs in modern survey sampling. In this section, we address the problem of how to allocate an a priori fixed sample size $n_s \leq N$ to the H strata. Generally, the costs of a survey will increase if the total sample size n_s increases. Thus, the total sample size n_s is strongly correlated with the costs of the survey and is therefore limited in several applications. The other major trade-off is the one between efficient population estimates on the one hand and efficient stratum- or area-specific estimates on the other hand. The development of an innovative allocation method, which considers both aspects at once, is one of the major tasks of this thesis and is extensively covered in Chapter 4. In general, (stratum-specific) sample sizes need to be integer numbers, since it is not possible to sample fractions of units (e.g. companies or households) of the population. Thus, either integer solvers need to be applied or the continuous solution needs to be rounded such that the following conditions hold:

1. Each stratum-specific sample size needs to be an integer number: $n_h \in \mathbb{N} \, \forall h = 1, \ldots, H$.
2. The sum over the (rounded) stratum-specific sample sizes needs to be equal to the total sample size: $\sum_{h=1}^{H} n_h = n_{\mathrm{s}}$.

Assumption 1 requires certain rounding procedures. The two assumptions are assumed to be fulfilled in the following description (without being mentioned in each situation). We start by introducing some standard techniques to allocate an a priori fixed sample size n_{s} to the disjoint and exhaustive strata $h = 1, \ldots, H$ in StrRS.

A commonly used and very simple allocation technique is the *equal* allocation, where the total sample size is equally distributed amongst the strata $\mathcal{U}_h$, i.e.

$$n_h^{\mathrm{EQ}} = \frac{n_{\mathrm{s}}}{H} \, \forall h = 1, \ldots, H. \tag{2.45}$$

The equal allocation (2.45) does not prevent an overallocation meaning that the desired stratum-specific sample size may exceed the stratum size in at least one stratum ($n_h > N_h$). Choudhry et al. (2012) presented a modification which may be considered instead of (2.45), where n_h is upper-bounded by N_h, and the remaining sample size is equally distributed to the other strata. An advantage of the equal allocation is that design-based area- and stratum-specific estimates may be efficient (i.e. their variance may be comparatively small), since, very small sample sizes are generally avoided. However, the accuracy of population estimates may not be at a very high level.

In contrast to (2.45), the *proportional* allocation given by

$$n_h^{\mathrm{PROP}} = \frac{n_{\mathrm{s}}}{N} N_h \, \forall h = 1, \ldots, H \tag{2.46}$$

yields the same sampling fraction $f = \frac{n_{\mathrm{s}}}{N}$ in all strata, up to rounding effects (cf. Särndal et al., 1992, p. 107). Thus, all units within the population share (almost) the same probability of being included in the sample. Moreover, Equation (2.46) prevents overallocation, since the ratio $\frac{n_{\mathrm{s}}}{N}$ does not exceed a value of 1.0. Comparing the accuracy of population estimates for τ_y using SRS and StrRS with a proportional allocation, Särndal et al. (1992, pp. 108 f.) proved that StrRS with proportional allocation yields smaller variances compared to SRS, except in the rather theoretical case where all stratum means $\bar{y}_h$ are almost equal. The more unequal the stratum means $\bar{y}_h$ are, the more precision will be gained by using StrRS with proportional allocation instead of SRS (cf. Lohr, 2009, p. 105 f.). Hence, a sensible construction of the strata should yield heterogeneity between the strata and homogeneity within the strata. If the population is not completely homogeneous, this results in a kind of homogeneity within each stratum, which is illustrated by Särndal et al. (1992, Example 3.7.1). Thus, if the stratum means $\bar{y}_h$ significantly differ among each other, we can basically assume

$$\mathrm{Var}(\hat{\tau}_y^{\mathrm{StrRS.PROP,HT}}) \leq \mathrm{Var}(\hat{\tau}_y^{\mathrm{SRS,HT}}). \tag{2.47}$$

As the stratification of the population is sensible, in general, the inequality (2.47) holds for almost all practical applications. A drawback of the proportional allocation is that it tends to disregard design-based area- and stratum-specific estimates, as it may lead to small stratum-specific sample sizes for smaller strata. Even stratum-specific sample sizes of zero may occur due to the rounding procedure.

In general, surveys focus on the computation of population estimates. To meet this condition, the total sample size may be allocated to the strata in a way that minimizes the variance of the HT estimator (2.16) of the variable of interest y for the population, as given by (2.17). This procedure is referred to as *optimal* allocation and was introduced by Tschuprow (1923) and Neyman (1934) for the case of one variable of interest y, also known as the *univariate* case. In Tschuprow (1923) and Neyman (1934), the variance $\text{Var}(\hat{\tau}_y^{\text{StrRS,HT}})$ is minimized under the assumption of a fixed total sample size n_{s}. It can be easily shown that some derivations of the first-order optimality conditions lead to the closed form

$$n_h^{\text{OPT}} = \frac{N_h S_h}{\sum_{\iota=1}^{H} N_\iota S_\iota} \cdot n_{\text{s}} \tag{2.48}$$

for all $h = 1, \ldots, H$. This approach was extended in Gabler et al. (2012), Münnich et al. (2012c), Wagner (2013), Friedrich et al. (2015), and Friedrich (2016). In these studies, upper bounds $M_h \leq N_h$ and lower bounds $m_h \geq 2$ for each stratum-specific sample size n_h, namely *box-constraints*, are added to the optimization problem to prevent overallocation as well as extremely small stratum-specific sample sizes. A lower bound of $m_h \geq 2$ ensures that the variance formula (2.36) is well-defined, as the denominator in S_h^2 defined in (2.26) is greater than zero. As a consequence, the resulting optimization problem

$$\begin{aligned} \min_{n \in \mathbb{R}^H} \; & \text{Var}(\hat{\tau}_y^{\text{StrRS,HT}}) \\ \text{s.t.} \; & \sum_{h=1}^{H} n_h = n_{\text{s}} \\ & 2 \leq m_h \leq n_h \leq M_h \; \forall h = 1, \ldots, H \end{aligned} \tag{2.49}$$

needs to be solved numerically, since the Lagrangian approach does not lead to a closed form solution. In Gabler et al. (2012), the theory of the solution strategy is introduced and verified. In Münnich et al. (2012c), the optimization problem is reformulated as a nonlinear system of equations which results from the Lagrangian approach. The authors made use of the special structure of the objective function (2.36), since it can be easily rewritten as

$$\text{Var}(\hat{\tau}_y^{\text{StrRS,HT}}) = \sum_{h=1}^{H} \frac{S_h^2 N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) = \sum_{h=1}^{H} \left(\frac{d_h}{n_h} - S_h^2 N_h\right) \tag{2.50}$$

with some constants $d_h \in \mathbb{R}_+$ for $h = 1, \ldots, H$. Since we ignore empty strata or strata with $S_h^2 = 0$, the constants d_h are supposed to be strictly positive. Then, the problem (2.49) can equivalently be reformulated with the inequality constraint $\sum_{h=1}^{H} n_h \leq n_{\text{s}}$, but equality holds at every optimal solution: assuming $n^* \in \mathbb{R}^H$ is optimal for (2.49) and $\sum_{h=1}^{H} n_h = n_{\text{s}} - \epsilon$ with $\epsilon > 0$. Then

$$\text{Var}^{\text{OPT}}(\hat{\tau}_y^{\text{StrRS,HT}}) = \left(\sum_{h=1}^{H} \frac{d_h}{n_h^*} - S_h^2 N_h\right) > \frac{d_1}{n_1^* + \epsilon} + \sum_{h=2}^{H} \frac{d_h}{n_h^*} - \sum_{h=1}^{H} S_h^2 N_h.$$

Since $\frac{d_1}{n_1^* + \epsilon} > 0$, the resulting variance for $\sum_{h=1}^{H} n_h = n_\mathrm{s} - \epsilon$ is bigger than for $\sum_{h=1}^{H} n_h = n_\mathrm{s}$, such that equality holds in (2.49) for each optimal solution. The objective function (2.50) is continuously differentiable on the feasible set and fulfills the property of separability (cf. Boyd and Vandenberghe, 2004, pp. 248 f.), which is a necessary condition for the application of some efficient numerical solvers.

Remark 2.5.1. A function $F : \mathbb{R}_+^H \to \mathbb{R}_{0+}$ is *separable* if it can be rewritten as an independent sum over its components F_h depending on the individual variables n_h, i.e.

$$F(n) = \sum_{h=1}^{H} F_h(n_h).$$

Moreover, a numerical algorithm using a fixed point iteration is presented to solve (2.49) continuously in a short processing time in Münnich et al. (2012c). This approach is discussed and extended in Chapter 4 in order to generalize the box-constraint optimal allocation with regard to multivariate allocations and including additional restrictions. Because the stratum-specific sample sizes n_h all need to be integer numbers, a rounding problem might occur. Since the rounded optimal (continuous) solution may not be optimal at all, Friedrich et al. (2015) provided a further extension that ensures integrality of the solution of the optimization problem (2.49). This was done based on Greedy algorithms, which avoids rounding.

Since, in practice, the total of the variable of interest is estimated by means of the survey, the variable of interest (or its variance) is not known at the design stage, when the sample allocation is computed (cf. Särndal et al., 1992, p. 106). Hence, the stratum-specific variances S_h^2 of variable of interest y, which are necessary for the computation of the optimal allocation, are not available. Alternatively, either proxies or highly correlated auxiliary data need to be used. These proxies can be values from previous years, for instance. Generally, the quality of the resulting optimal sample sizes depends on the accuracy of the proxy or auxiliary information used. If (2.48) or (2.49) is calculated using past values of y and meanwhile the variable of interest has been subject to a large shock (e.g. as a consequence of demographic changes or economic crises), the resulting allocation may be far from optimal. In any case, this statistical relationship between the possibly unsteady population changes and the accuracy of the estimates is also investigated in Chapter 4. An optimal *multivariate* allocation approach is proposed to weaken this effect since several variables of interest are considered simultaneously. The consideration of several variables within the allocation is vital in modern surveys, as the user pursues various possibly conflicting goals within one single survey. To solve optimal multivariate allocation problems, strategies of Pareto optimization are proposed in this thesis, and this will be the starting point of the developments in Chapter 4. We have discussed the strategy of applying Pareto optimization for a simpler framework in Friedrich et al. (2018). Moreover, optimal allocation only concentrates on the minimization of the variance of population total estimates. Variances for area- or stratum-specific estimates are neglected, which may lead to inefficient estimates on disaggregated stratification levels. This issue is also addressed in Chapter 4.

2.6 The RIFOSS dataset

To test the developed statistical methods and to accentuate their functionality, the applications and simulations considered in this thesis are performed on a synthetic dataset representing all inhabitants of Germany. The dataset[1] has been generated at Trier University within the RIFOSS[2] project funded by the Federal Statistical Office of Germany. It is based on the simulation dataset used for the evaluation of the methodology of the German Census 2011. This dataset has also been generated at Trier University, and it is based on an extract from the official population register of Germany (cf. Münnich et al., 2012a, Chapter 3.2). Additional variables have been added with the aid of a scientific use file of the German Microcensus 2008[3]. Generally, the variables have been included applying multinomial logistic regression model. The generation process of the data has been done separately for three times. Subsequently, approximately a third of all households has been selected based on a simulated annealing strategy (cf. Laarhoven and Aarts, 1987, Chapter 3) in a way where population and sub-population totals are fitted to the results of the German Census 2011. The generation process has been inspired by Alfons et al. (2011) and Rahman and Harding (2016, Section 4.4.1). For a more detailed analysis of the generation of synthetic datasets, we also refer to Burgard et al. (2017).

In representing real sizes of the German population, the dataset consists of approximately 85 million individuals inhabiting approximately 39 million households. The 85 million contain approximately 82 million individuals living in their main residence. Additionally, approximately 3 million of them are also listed with a secondary residence. Although the methods developed in the thesis are even able to handle these dimensions in an appropriate time, the dataset is reduced to four federal states of Germany due to the simpler manageability (with regard to the working memory and hard disk memory sizes). Thus, the dataset is reduced to the federal states of Hesse, North Rhine-Westphalia, Rhineland-Palatinate, and Saarland. Moreover, households are defined as sampling units. The reduction results in a population size of 11 121 631 households accommodating 30 077 329 individuals.

The population is stratified by the following stratification levels by region or by content:

1. FS: 4 federal states (by region),
2. NUTS2: 12 NUTS2 regions (by region),
3. NUTS3: 121 NUTS3 regions (by region),
4. SMP: 784 sampling points (by region; same structure as in the German Census 2011),
5. HHS: 8 classes of household sizes (by content; same number of persons in each class).

Corresponding to the five stratification levels, the sampling design is defined on 6 272 strata built as cross-classifications of the five stratification levels that is, the 784 sampling points and the 8 classes of household sizes. In addition to estimates of the population total and of stratum-

[1] Version: RIFOSS_GG_v0.1.1_vanilla_inc_cream, accessed 22 June 2018

[2] RIFOSS: Research Innovation for Official and Survey Statistics

[3] http://www.forschungsdatenzentrum.de/bestand/mikrozensus/suf/2008/fdz_mz_suf_2008_schluesselverzeichnis.pdf

specific totals, namely *area-specific* estimates will also be evaluated for NUTS2 and NUTS3 regions as well as for sampling points and classes of household sizes. The four regional stratification levels are plotted in the Figures 2.2 and 2.3 for Germany and for the four federal states respectively. The maps exemplarily show the mean number of individuals under the age of 20 within a household per region. The values of the respective regions illustrate the heterogeneous structure of the dataset. When an aggregated region may be shaded in dark blue, the color structure of a corresponding disaggregated levels may be highly heterogeneous. However, it has to be remarked that the values are *not* real values since the dataset is synthetically generated.

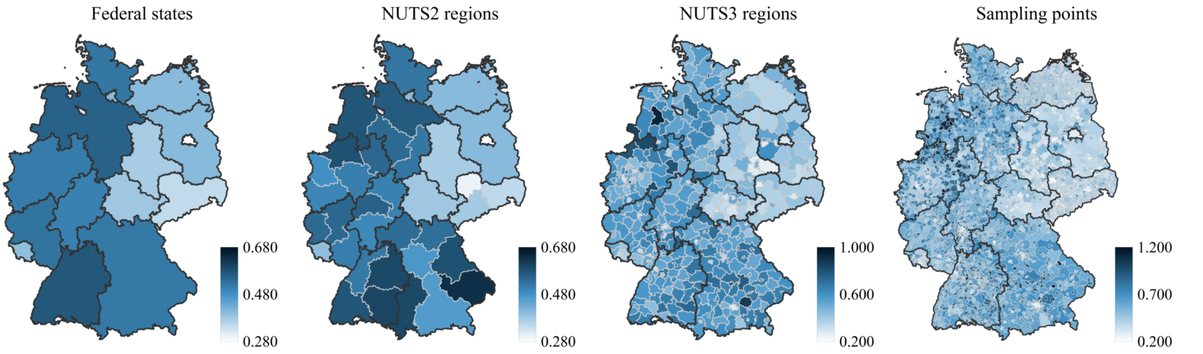


Figure 2.2: Regional stratification levels of Germany (exemplary content: mean number of persons under the age of 20 within a household).

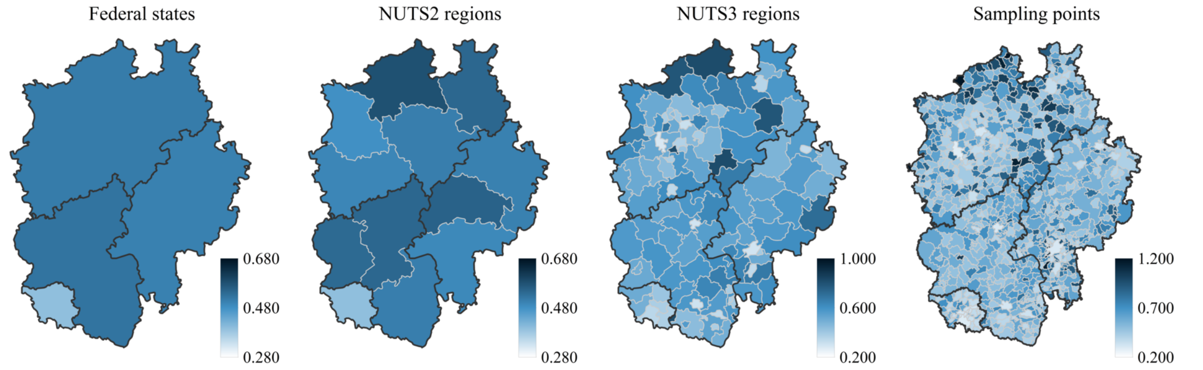


Figure 2.3: Regional stratification levels of four federal states (exemplary content: mean number of persons under the age of 20 within a household).

The variables considered in the applications and simulations in this thesis are tabulated in the original form on individual level in Table B.1 in Appendix B.1. Before the application, the variables are suitably transformed to the household structure. The resulting variables are tabulated in Table B.2. The expression *suitably transformed* is exemplary meant to refer to the transformation of the variable EF44 (age of person) to the variables AGE4.1, AGE4.2, AGE4.3, and AGE4.4 corresponding to the number of persons who are (a) under the age of 20, (b) from 20 to 39, (c) from 40 to 59, and (d) 60 or older living in a household respectively. Since the population is fully available on household level, the evaluation of the quality of the estimates computed by the application and simulation studies is then possible based on the true values.

Chapter 3

Fundamentals of Numerical Optimization

In this chapter, fundamental frameworks and definitions of numerical optimization are presented. The scope of the content is limited to aspects that are pertinent to the thesis. In Section 3.1, the basic theory of nonlinear optimization is introduced. As we have already seen in Chapter 2, the optimal allocation as well as the calibration methods are based on nonlinear optimization problems. Thus, the theory of nonlinear optimization is the theoretical foundation to solve these problems. Afterwards, some aspects of non-smooth optimization are addressed in Section 3.2 with a focus on a property of functions called *semismoothness*. This property is a weaker assumption than continuous differentiability. Under this assumption, a semismooth Newton method (`SSN`) is applicable to solve special cases of non-smooth equations. This is essential for the developments in Chapters 4 and 5, as the reduction of the dimension of the original optimization problems to be solved results in non-differentiable, but semismooth nonlinear systems of equations. In Section 3.3, the theory of multi-criteria (or multi-objective) optimization is introduced as it is the basis of the derivation of the optimal multivariate and multi-domain allocation method in Chapter 4. As the optimal multivariate allocation considers more than one variable of interest, a multi-criteria optimization problem has to be solved. Due to reasons of comprehensibility, the notation used in this chapter differs from the conventional notation of this thesis. For instance, $n, m \in \mathbb{N}$ are natural numbers, and x, y, and z are dependent variables of functions F and G.

3.1 Nonlinear optimization

We have seen in Chapter 1, that almost all solution strategies for survey statistical models desire the application of optimization techniques. Regarding this point, the development of the optimal multivariate and multi-domain allocation methods (`MMDopt`) based on (2.49) and the extension of general calibration methods (`GCAL`) such as (2.39) require the solution of a restricted nonlinear optimization problem. For this reason, we provide a short overview of the theoretical background of common numerical solution strategies. A more detailed theoretical discussion can be found in Horst (1979), Geiger and Kanzow (2002), Ruszczynski (2006), and Jahn (2007).

Firstly, we will revise the elementary concept of continuity of a vector-valued function F given by Walter (2002, pp. 41 ff.).

Definition 3.1.1 (Lipschitz-continuous)**.**

1. A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called *Lipschitz-continuous* if there exists a constant $L \geq 0$ such that
$$\|F(x_1) - F(x_2)\|_2 \leq L\|x_1 - x_2\|_2 \; \forall x_1, x_2 \in \mathbb{R}^n.$$
2. A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called *locally Lipschitz-continuous* if for every $x \in \mathbb{R}^n$ there exists a neighborhood U_x of x such that F is Lipschitz-continuous in U_x.

Definition 3.1.2 (Positive homogeneity, linearity)**.** A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called

1. *positively homogeneous*, if $F(\alpha x) = \alpha F(x)$ for all $x \in \mathbb{R}^n$ and $\alpha \geq 0$.
2. *linear*, if $F(x_1 + x_2) = F(x_1) + F(x_2)$ and $F(\alpha x_1) = \alpha F(x_1) \; \forall x_1, x_2 \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.

Definition 3.1.2 implies that a linear function is also positively homogeneous. The other implication does not hold. For example, the absolute value function $f : \mathbb{R} \to \mathbb{R}, x \mapsto |x|$ is positively homogeneous but is not linear, indeed. In the further course of the chapter, we define two different types of differentiability of a function F, each is characterized by one of the two properties in Definition 3.1.2 (cf. Yamamuro, 1974, Section 1.2 and Werner, 2007, Section 3.5). The first type of differentiability will be the basis of convergence results of the classical Newton method, while the second one enables the definition of a non-smooth Newton method (see Section 3.2).

The most common type of differentiability is the *Fréchet-differentiability*, which is referred to generally as differentiability. For simplicity, we characterize this property in Definition 3.1.3 based on Werner (2007, Lemma 3.5.2). For a general definition, we refer to Werner (2007, Definition 3.5.1).

Definition 3.1.3 (Frechét-differentiability)**.** Let $Z \subseteq \mathbb{R}^n$ an open set. A function $F : Z \to \mathbb{R}^m$ is called *Fréchet-differentiable* at $x^0 \in Z$, if there exists a continuous linear function $A : \mathbb{R}^n \to \mathbb{R}^m$ with
$$\lim_{\|h\| \to 0} \frac{F(x^0 + h) - F(x^0) - A(h)}{\|h\|} = 0.$$
F is called *Fréchet-differentiable* if F is Fréchet-differentiable at each $x^0 \in Z$.

Fréchet-differentiability of a function F is a most common assumption for numerical solvers. If it does not hold in some applications, we then try to verify a weaker assumption of function F based on another type of differentiability (see Section 3.2) that also allows for the application of common numerical solvers. From this point onward, a Fréchet-differentiable function F is simply called *differentiable*.

In the following definition, different types of convexity are defined for sets and real-valued functions $f : \mathbb{R}^n \to \mathbb{R}$ in accordance to Boyd and Vandenberghe (2004, Sections 2.1.4, 3.1.1, 3.4.1) and Jahn (2007, Definition 4.15).

Definition 3.1.4.

1. A set $Z \subseteq \mathbb{R}^n$ is a *convex* set, if for any $x_1, x_2 \in Z$ and any $\alpha \in [0, 1]$

$$\alpha x_1 + (1 - \alpha)x_2 \in Z.$$

2. Let $Z \subseteq \mathbb{R}^n$ be a convex set. Then, the function $f : Z \to \mathbb{R}$ is a *convex* function, if for any $x_1, x_2 \in Z$ and any $\alpha \in [0, 1]$

$$f\Big(\alpha x_1 + (1 - \alpha)x_2\Big) \leq \alpha f(x_1) + (1 - \alpha)f(x_2).$$

If the inequality even holds with $<$, the function f is *strictly convex*.

3. Let $Z \subseteq \mathbb{R}^n$ be a convex set. Then, the function $f : Z \to \mathbb{R}$ is called *quasi-convex*, if

$$Z_\alpha := \{x \in Z : f(x) \leq \alpha\}$$

is a convex set for all $\alpha \in \mathbb{R}$. Alternatively, f is also quasi-convex, if for any $x_1, x_2 \in Z$ and any $\alpha \in [0, 1]$

$$f\Big(\alpha x_1 + (1 - \alpha)x_2\Big) \leq \max\Big(f(x_1), f(x_2)\Big).$$

4. Let $Z \subseteq \mathbb{R}^n$ and let the directional derivatives of function $f : Z \to \mathbb{R}$ exist in $x^* \in S$ in every direction. Then, the function f is called *pseudo-convex* in x^*, if the implication

$$f'\Big(x^*;(x - x^*)\Big) \geq 0 \quad \Rightarrow \quad f(x) - f(x^*) \geq 0$$

holds for all $x \in Z$, where $f'\Big(x^*;(x - x^*)\Big)$ is the directional derivative of f in x^* in direction $(x - x^*)$.

Jahn (2007, p. 94) shows that the properties *quasi-convex* and *pseudo-convex* are generally weaker assumption than the classical *convexity* for differentiable functions.

Applying the previously defined properties, it is now possible to derive optimality conditions for nonlinear optimization problems. Let the nonlinear optimization problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \; & f(x) \\ \text{s.t. } & h_i(x) = 0 \; (i = 1, \ldots, q_2) \\ & g_j(x) \leq 0 \; (j = 1, \ldots, q_3) \end{aligned} \tag{3.1}$$

be given with (component-wise) twice continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^{q_2}$, and $g : \mathbb{R}^n \to \mathbb{R}^{q_3}$.

Remark 3.1.5. Problem (3.1) is referred to as a *convex optimization problem*, if f and g_j are convex functions for all $j = 1, \ldots, q_3$, and h_i are affine-linear functions for all $i = 1, \ldots, q_2$, i.e. $h_i(x) := a_i^T x + b_i$ for some $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$. Due to Geiger and Kanzow (2002, Lemma 2.14), the feasible set

$$\mathcal{X} := \Big\{x \in \mathbb{R}^n : h_i(x) = 0 \; (i = 1, \ldots, q_2) \text{ and } g_j(x) \leq 0 \; (j = 1, \ldots, q_3)\Big\} \subseteq \mathbb{R}^n. \tag{3.2}$$

of problem (3.1) is convex in this case.

The Lagrangian function of problem (3.1) is defined as

$$\mathcal{L}(x,\lambda,\beta) := f(x) + \sum_{i=1}^{q_2} \lambda_i h_i(x) + \sum_{j=1}^{q_3} \beta_j g_j(x) \tag{3.3}$$

with Lagrangian multipliers $\lambda \in \mathbb{R}^{q_2}$ and $\beta \in \mathbb{R}^{q_3}$ (cf. Geiger and Kanzow, 2002, p. 46 and pp. 241 ff.). The corresponding Karush-Kuhn-Tucker (KKT-) conditions are given by the following nonlinear system of equations

$$\nabla_x \mathcal{L}(x,\lambda,\beta) = 0, \tag{3.4}$$

$$h_i(x) = 0 \; \forall i = 1,\dots,q_2, \tag{3.5}$$

$$g_j(x) \le 0, \; \beta_j \ge 0, \; \beta_j g_j(x) = 0 \; \forall j = 1,\dots,q_3. \tag{3.6}$$

Depending on the properties of the objective function f and the constraint functions, the KKT-conditions may be applied to formulate necessary and sufficient optimality conditions. In that regard, the inequality conditions $g_j(x) \le 0$ and $\beta_j \ge 0$ impede a direct solution of Equations (3.4) to (3.6) via standard techniques for nonlinear systems of equations. Alternatively, Equation (3.6) can be equivalently rewritten by the nonlinear complementarity problem formulation

$$\varphi\Big(-g_j(x), \beta_j\Big) = 0 \; \forall j = 1,\dots,q_3, \tag{3.7}$$

where $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is a NCP-function (i.e. $\varphi(a,b) = 0 \Leftrightarrow a \ge 0, b \ge 0, ab = 0$; cf. Geiger and Kanzow, 2002, p. 242). An example for a NCP-function is the minimum function $\varphi(a,b) = \min\{a,b\}$, which is not differentiable like most of the NCP-functions. However, the reformulation yields a nonlinear system equations (KKT-system)

$$\begin{aligned} \nabla_x \mathcal{L}(x,\lambda,\beta) &= 0, \\ h_i(x) &= 0 \; \forall i = 1,\dots,q_2, \\ \varphi\Big(-g_j(x), \beta_j\Big) &= 0 \; \forall j = 1,\dots,q_3 \end{aligned} \tag{3.8}$$

consisting of Equations (3.4), (3.5), and (3.7). This system depending on $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^{q_2}$, and $\beta \in \mathbb{R}^{q_3}$ forms optimality conditions and can be solved by methods for solving non-smooth nonlinear systems of equations. Beside the fulfillment of the KKT-system, the regularity of a feasible point has to be verified by a constraint qualification condition. Common conditions are the linear independence constraint qualification condition (LICQ, cf. Geiger and Kanzow, 2002, Definition 2.40), the Mangasarian-Fromlovitz constraint qualification condition (MFCQ, cf. Ruszczynski, 2006, Lemma 3.17), and for a convex problem the Slater-condition (cf. Horst, 1979, Collorary 3). The fulfillment of the LICQ condition implies the fulfillment of the MFCQ condition. For a convex problem (cf. Remark 3.1.5), the Slater-condition is often preferred.

Definition 3.1.6. Let $x \in \mathcal{X}$ be a feasible point of problem (3.1) and let

$$I(x) := \Big\{ j \in \{1,\dots,q_3\} : g_j(x) = 0 \Big\}$$

denotes the *active set* of problem (3.1) at $x \in \mathbb{R}^n$. Then, the following constraint qualification conditions are defined:

1. LICQ holds at $x \in \mathcal{X}$, if the gradients

$$\nabla g_j(x) \text{ for all } j \in I(x) \text{ and } \nabla h_i(x) \text{ for all } i = 1, \dots, q_2$$

are linearly independent.

2. MFCQ holds at $x \in \mathcal{X}$, if the gradients

$$\nabla h_i(x) \text{ for all } i = 1, \dots, q_2$$

are linearly independent and there is a vector $s \in \mathbb{R}^n$ such that

$$\nabla g_j(x)^T s < 0 \text{ for all } j \in I(x) \text{ and } \nabla h_i(x)^T s = 0 \text{ for all } i = 1, \dots, q_2.$$

3. Let the problem (3.1) be a convex optimization problem in the sense of Remark 3.1.5. The Slater-condition holds for problem (3.1), if there is a vector $x \in \mathbb{R}^n$ with

$$g_j(x) < 0 \text{ for all } j = 1, \dots, q_3 \text{ and } h_i(x) = 0 \text{ for all } i = 1, \dots, q_2.$$

Following this, we are able to formulate some necessary optimality conditions for problem (3.1). Since the considered problems are exclusively convex problems, we limit ourselves to this case.

Theorem 3.1.7 (Necessary and sufficient optimality conditions)**.** Let the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint function $g : \mathbb{R}^n \to \mathbb{R}^{q_3}$ be twice continuously differentiable and convex. Let $h : \mathbb{R}^n \to \mathbb{R}^{q_2}$ be twice continuously differentiable and affine-linear. Moreover, let the Slater-condition be satisfied. Then, $x^* \in \mathbb{R}^n$ is a global solution of problem (3.1) if and only if there are Lagrangian multipliers $\lambda^* \in \mathbb{R}^{q_2}$ and $\beta^* \in \mathbb{R}^{q_3}$ which fulfill the KKT-system (3.8) for $x^* \in \mathbb{R}^n$.

For the proof we refer to Geiger and Kanzow (2002, Theorems 2.45 and 2.46, and p. 245).

Jahn (2007, Lemma 2.14) proved that under the assumptions of Theorem 3.1.7, the set of optimal solutions of (3.1) is a convex set. In addition, if the objective function f is strictly convex, problem (3.1) has a unique optimal solution.

Algorithms to solve nonlinear optimization problems with constraints are widespread in the literature. An overview can be found in Geiger and Kanzow (2002), Bonnans et al. (2006), and Lange (2013). Most of them attempt to somehow solve the KKT-system (3.8). One evident method is the Lagrange-Newton method, where the KKT-system (3.8) is solved via a Newton method (cf. Geiger and Kanzow, 2002, pp. 239 ff.). In the presence of inequality constraints, the solver needs to be a non-smooth version of the Newton method due to the non-differentiability of φ in (3.7). This issue is discussed in detail in Section 3.2. Moreover, sequential quadratic programming (`SQP`) methods are common solvers for nonlinear optimization problems. Instead of directly solving the KKT-system, quadratic subproblems are solved sequentially. After each iteration, the solution and the Lagrangian multipliers is updated (cf. Geiger and Kanzow, 2002, p. 244). Another class of solvers is based on penalization methods, where the constraint functions are added additively as a penalization parameter to the objective function. One example is the augmented Lagrange method, in which unconstrained optimization problems (with the constraints added as penalty term) are solved iteratively. After each iteration, an approximation of the Lagrangian multipliers is updated (cf. Ruszczynski, 2006, pp. 196 ff.). Since the presented methods are mostly based on solving the KKT-system, e.g. with the Newton method, the Hessian of the objective function plays a significant role. To reduce the computational burden, the use of approximations instead of the Hessian is a popular choice. These algorithms are generally known as *quasi* Newton methods. A common strategy is based on the BFGS-update formulas (cf. Nocedal and Wright, 2006, Section 8.1). Moreover, trust region methods are also applicable for nonlinear optimization problems; the iterative strategy of these algorithms is based on small regions in which quadratic subproblems are solved to compute a correction vector. This vector is utilized to update the solution in each iteration (see Qi and Sun, 1994).

Each of the mentioned algorithms has benefits and drawbacks, but all of them contain iterative strategies whose dimension is at least of the dimension of the underlying optimization problem (3.1). For example, the Lagrange-Newton method comprises the solution of a $(n + q_2 + q_3)$-dimensional nonlinear system of equations. Moreover, the computational effort exponentially increases in the dimension n of the problem. Due to the continuously increasing amount of data, the numeric is faced with ever-growing problem dimensions. Aside from this, there are also other reasons to develop innovative strategies to reduce the computational effort, such as regarding variance or MSE estimation strategies. Since the structure of sampling designs and point estimators are often quite complex, a variance estimation technique applying linearizations via Taylor approximations is not reasonable. Thus, resampling methods have to be applied (cf. Särndal et al., 1992, p. 419) in which sub-samples are sequentially drawn from the sample (see Section 5.4). As a consequence, a problem needs to be solved up to $10\,000$ times (see Section 5.4), such that a reasonable computation time of one hour increases to 420 days (without parallel computing).

Münnich et al. (2012b) and Münnich et al. (2012c) proposed an alternative approach to solve a special case of optimization problems of the form (3.1) concerning specific objective functions and specific structures of constraints (only affine-linear equality constraints and box-constraints, which are treated different to the inequality constraints) respectively. The special cases correspond to the optimal univariate allocation problem with box-constraints (2.49) and the generalized calibration methods of the form (2.39). The main advantage is the reduction of the dimension of the underlying KKT-system from $(n + q_2 + q_3)$ to $(q_2 + q_3)$. Generally, since in our applications $q_2 + q_3 \ll n$ (with n being the number of strata, population, or sample size), the reduction of the computational effort may be significant. Since the dimension of the alternative system is independent of n, the exponential increase of the running time depending on the dimension of the optimization problem (3.1) is omitted. However, the non-differentiability of the KKT-system remains.

Since the statistical applications are becoming more complicated within the past few years, the main goals of this thesis are to extend the approach of Münnich et al. (2012b) and Münnich et al. (2012c) to multivariate and more general problems and to develop a robust and efficient numerical solver for these applications. The main difficulty of the presented approach is the non-differentiability of the KKT-system. In analogy to the mentioned publications, we face this with a semismooth version of the Newton method, as particularly discussed in Section 3.2.

3.2 Semismooth Newton method

As discussed in Section 3.1, the reduction of the dimension of the KKT-system (3.8) may yield to non-differentiability. Nevertheless, we wish to continue solving the reduced KKT-system with a Newton type method, since it ensures (under some assumptions) locally quadratic and superlinear convergence. The classical Newton method is an iterative solver for a nonlinear system of equations $F(x) = 0$ with a continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^n$. In each iteration k , a linear system

$$J_F(x^k)d^k = -F(x^k)$$

with the Jacobian $J_F(x^k)$ of F in $x^k \in \mathbb{R}^n$ is solved. Then, the next iterate is computed by

$$x^{k+1} = x^k + d^k.$$

As shown in Geiger and Kanzow (2002, Theorem 5.26), this method is superlinearly convergent if both the initial value $x^0 \in \mathbb{R}^n$ is close enough to the solution x^* and $J_F(x^k)$ is regular. If the Jacobian is also locally Lipschitz-continuous, the method converges with a locally quadratic convergence rate to the optimal solution x^*. Due to the non-differentiability of the KKT-system, we have to use a non-smooth version of the Newton method, which is referred to as the *semismooth* Newton method suggested by Qi and Sun (1993) and Qi (1993). It has similar convergence rates compared to the classical Newton method.

While the classical Newton method requires continuously (Fréchet-)differentiable functions based on Definition 3.1.3, the semismooth Newton method suffices with a weaker assumption based on the *Bouligand*-differentiability (B-differentiability) originally published in Robinson (1987, p. 60 and Definition A.1). A more intuitive definition is given by Pang (1990).

Definition 3.2.1 (Bouligand-differentiability)**.** Let $Z \subseteq \mathbb{R}^n$ be an open set. A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called *B-differentiable* at $x^0 \in Z$, if there exists a positively homogeneous function $A : \mathbb{R}^n \to \mathbb{R}^m$ with

$$\lim_{\|h\| \to 0} \frac{F(x^0 + h) - F(x^0) - A(h)}{\|h\|} = 0.$$

F is called *B-differentiable*, if F is B-differentiable at each $x^0 \in Z$.

In comparing the Fréchet-differentiability (Definition 3.1.3) and the B-differentiability for functions $F : Z \to \mathbb{R}^m$ in an open set $Z \subseteq \mathbb{R}^n$, the only difference is that the B-differentiability requires positively homogeneous function A, whereas A has to be continuous and linear in the case of Fréchet-differentiability. Thus, the following implication holds for $F : Z \to \mathbb{R}^m$:

$$F \text{ is Fréchet-differentiable at } x^0 \in Z \;\Rightarrow\; F \text{ is B-differentiable at } x^0 \in Z. \tag{3.9}$$

Provided that F is locally Lipschitz-continuous at x^0, Shapiro (1990) has shown that F is B-differentiable at x^0 if and only if F is directionally differentiable at x^0. According to Rademacher (1919), a locally Lipschitz-continuous function is Fréchet-differentiable almost everywhere, i.e. the set of points in which a B-differentiable function is not Fréchet-differentiable has a Lebesgue-measure of zero. Thus, many properties of Fréchet-differentiable functions can be similarly extended to B-differentiable functions.

Since the classical Jacobian $J_F(x)$ is only defined if F is Fréchet-differentiable at $x \in Z$, we have to define an equivalent to the Jacobian for B-differentiable functions. Qi (1993) proposed to use the generalized Jacobian introduced by Clarke (1983, p. 70) and the one similarly denoted in Ito and Kunisch (2009, Equations (1.8) and (1.9)).

Definition 3.2.2 (B-subdifferential, generalized Jacobian)**.** Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz-continuous and Z_F be the set of points at which F is Fréchet-differentiable. Then, the set

$$\partial_B F(x^0) := \left\{ H \in \mathbb{R}^{m\times n} : \exists \{x^k\}_{k\in\mathbb{N}} \subset Z_F \text{ with } x^k \to x^0 \text{ and } J_F(x^k) \to H \right\}$$

is called *B-subdifferential* of F at x^0. Moreover, its convex hull

$$\partial F(x^0) := \text{conv}\, \partial_B F(x^0)$$

is called the *generalized Jacobian* of F at x^0.

Both the B-subdifferential $\partial_B F(x)$ and the generalized Jacobian $\partial F(x)$ have a cardinality of one and are equivalent to the Jacobian $J_F(x)$ for all $x \in \mathbb{R}^n$ where F is Fréchet-differentiable. The existence of a sequence $\{x^k\}_{k\in\mathbb{N}} \subset Z_F$ is given by Rademacher (1919), since a local Lipschitz-continuous function is Fréchet-differentiable almost everywhere.

Example 3.2.3. In this example, the difference between Fréchet- and Bouligand-differentiability is illustrated based on four functions. The corresponding functions are plotted in Figure 3.1.

1. If a function $F : \mathbb{R}^n \to \mathbb{R}^m$ is continuously Fréchet-differentiable, it is also B-differentiable and
$$\partial_B F(x) = \partial F(x) = \{J_F(x)\}\ \forall x \in \mathbb{R}^n.$$
An example is $F_1 : \mathbb{R} \to \mathbb{R}, x \mapsto (x-1)^2 + 1$.
2. Let $F_2 : \mathbb{R} \to \mathbb{R}, x \mapsto |x|$.
F_2 is not Fréchet-differentiable in $x = 0$, but B-differentiable.
$$\begin{aligned} \partial_B F_2(x) &= \partial F_2(x) = \{J_{F_2}(x)\}\ \forall x \in \mathbb{R}^n \setminus \{0\} \\ \partial_B F_2(0) &= \{-1, 1\},\ \ \partial F_2(0) = [-1, 1]. \end{aligned}$$
3. Let $F_3 : \mathbb{R} \to \mathbb{R}, x \mapsto \begin{cases} (x-1)^2 - 1 & , \text{ if } x \le \frac{3}{2} \\ 2x - \frac{7}{4} & , \text{ if } x > \frac{3}{2} \end{cases}$.
F_3 is not Fréchet-differentiable in $x = \frac{3}{2}$, but B-differentiable.
$$\begin{aligned} \partial_B F_3(x) &= \partial F_3(x) = \{J_{F_3}(x)\}\ \forall x \in \mathbb{R}^n \setminus \{3/2\} \\ \partial_B F_3(3/2) &= \{1, 2\},\ \ \partial F_3(3/2) = [1, 2]. \end{aligned}$$

4. Let $F_4 : \mathbb{R} \to \mathbb{R}, x \mapsto \sqrt{|x|}$.
   F_4 is not Fréchet- and B-differentiable in $x = 0$, since no directional derivative exists.

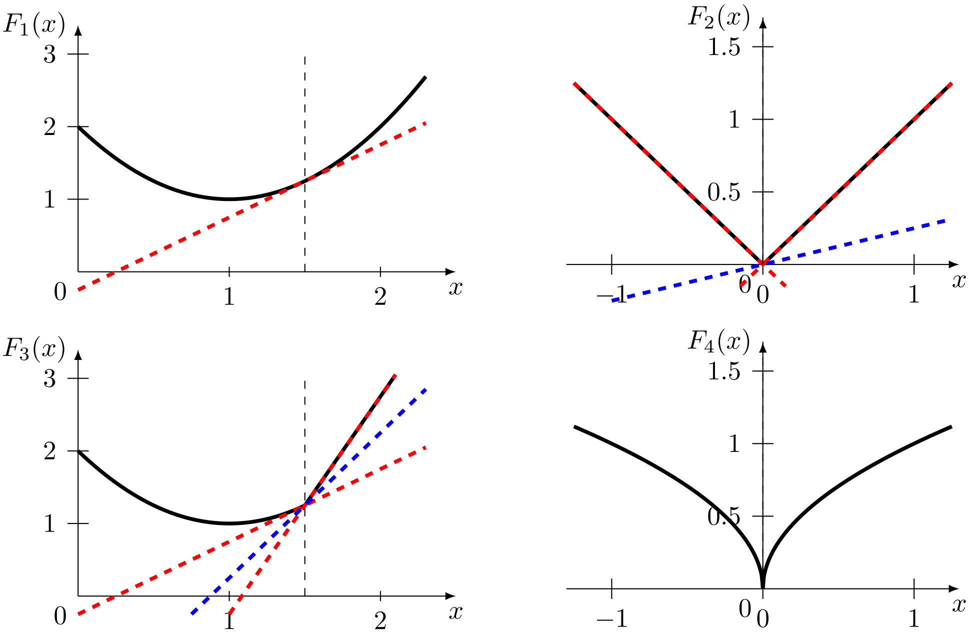


Figure 3.1: Examples for different types of differentiability. Functions $F_1, \dots, F_4$ are defined in Example 3.2.3. The function is plotted in *black*, the generalized Jacobian in red and exemplary elements of the B-subdifferential in blue.

In non-smooth Newton methods, the Jacobian $J_F(x^k)$ is replaced by an element of the generalized Jacobian $H^k \in \partial F(x^k)$ or an element of the B-subdifferential $H^k \in \partial_B F(x^k)$, depending on the considered version. Nevertheless, convergence of such a non-smooth Newton method is generally not given for B-differentiable functions. Therefore, we additionally have to assume that the function F is semismooth, a property which was originally introduced by Mifflin (1977) and was extended in Qi and Sun (1993).

Definition 3.2.4 (Semismoothness)**.** Let $Z \subseteq \mathbb{R}^n$ and $F : Z \to \mathbb{R}^m$ be a locally Lipschitz-continuous function and B-differentiable in $x^0 \in Z$. Then F is called

1. *semismooth* in $x^0 \in Z$, if $\lim\limits_{r^k \to 0, H^k \in \partial F(x^0+r^k)} \dfrac{H^k r^k - F'(x^0;r^k)}{\|r^k\|} = 0,$
2. *strongly semismooth* in $x^0 \in Z$, if $\limsup\limits_{r^k \to 0, H^k \in \partial F(x^0+r^k)} \dfrac{H^k r^k - F'(x^0;r^k)}{\|r^k\|^2} < \infty,$
3. (strongly) semismooth on Z, if F is (strongly) semismooth in each $x^0 \in Z$.

Since the definition of semismoothness is rather complicated than intuitive, Mifflin (1977), Qi and Sun (1993), and Fischer (1997) have proved the following lemmata.

Lemma 3.2.5 (Characterization of semismooth functions)**.**

1. Let $Z \subseteq \mathbb{R}^n$ be an open set, $x \in Z$ and $F : Z \to \mathbb{R}^m$ Lipschitz-continuous.
    - If F is continuously (Fréchet-)differentiable in x, then F is semismooth in x.
    - If F is (Fréchet-)differentiable and J_F locally Lipschitz-continuous in x, then F is strongly semismooth in x.
2. Let $Z \subseteq \mathbb{R}^n$ be an open and convex set and let $F : Z \to \mathbb{R}^m$ be a convex function. Then F is semismooth on Z.
3. Le $F : \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz-continuous. Then F is semismooth in x, if each component F_i of F is semismooth in x.
4. Scalar products as well as sums of semismooth functions are semismooth.
5. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be semismooth in $x \in \mathbb{R}^n$ and $G : \mathbb{R}^m \to \mathbb{R}^n$ be semismooth in $F(x) \in \mathbb{R}^m$. Then, the composition $G \circ F$ is semismooth in x.

For the proof of Lemma 3.2.5, we refer to Qi and Sun (1993) for Items 1 to 3, Mifflin (1977) for Item 4, and Fischer (1997) for Item 5.

If semismoothness can be verified for a non-smooth function F, e.g. with Lemma 3.2.5, we are now able to apply the semismooth Newton algorithm (SSN) published by Qi and Sun (1993) to solve a nonlinear system of equations $F(x) = 0$ with $F : \mathbb{R}^n \to \mathbb{R}^n$. The algorithm (with a certain step-size rule) is presented in Algorithm 1. Here, we choose H^k as an element of the B-subdifferential $\partial_B F(x^k)$ as it is presented in Qi (1993). In the original version by Qi and Sun (1993), H^k is forced to be an element of the generalized Jacobian $\partial F(x^k)$, which is a significantly stronger assumption (cf. Example 3.2.3). Following Qi and Sun (1993), the solvability increases if H^k is chosen from $\partial_B F(x^k)$. Thus, we consider this version hereafter. Moreover, if F is Fréchet-differentiable, Algorithm 1 is equivalent to the classical Newton method, since $\partial_B F(x^k) = \{J_F(x^k)\}$. In Algorithm 1, a step-size strategy with step-size $\alpha_k \leq 1$ can optionally be included. A detailed discussion about its benefits and its calculations is given at the end of this subsection.

In general, the computation of the B-subdifferential $\partial_B F(x^k)$ and the choice of a sensible element $H^k \in \partial_B F(x^k)$ is highly challenging. Thus, several alternative expressions of the semismooth Newton methods have been proposed. Han et al. (1992) discussed an approach based on B-differentiable functions which are additionally assumed to be directionally differentiable. Hence, the *generalized* Newton equation

$$H^k d^k = -F(x^k) \;\text{ with } H^k \in \partial_B F(x^k) \tag{3.10}$$

may be replaced by

$$F'(x^k;d^k) = -F(x^k), \tag{3.11}$$

where $F'(x^k;d^k)$ is the directional derivative of F in x^k in direction d^k. Han et al. (1992) proved both local and global convergence under specific assumptions. However, it will be shown in the following chapters, that the computation of a sensible element $H^k \in \partial_B F(x^k)$ is not a decisive factor in the considered cases. In that regard, an application of the standard semismooth Newton algorithm is sensible in the context of this thesis.

Algorithm 1 Semismooth Newton method with step-size rule (SSN)

```
Input: $F : \mathbb{R}^n \to \mathbb{R}^n$ locally Lipschitz-continuous, $x^0 \in \mathbb{R}^n$ initial value, $x^0 \in U(x^*)$
  while $\|F(x^k)\| \geq$ tol
    choose $H^k \in \partial_B F(x^k)$
    solve $H^k d^k = -F(x^k)$
    [optional] compute step-size $\alpha_k \in (0, 1]$, else $\alpha_k = 1$
    $x^{k+1} = x^k + \alpha_k d^k$
    $k \leftarrow k + 1$
  end while
Return: Solution $x^* \leftarrow x^k$
```

A convergence analysis of the semismooth Newton method is given in Pang (1990) and Qi and Sun (1993), where a locally superlinear convergence rate is proved (for the case without the step-size rule). In addition, a locally quadratic convergence rate is shown if the function F is strongly semismooth. Thus, the convergence results resemble these of the classical Newton method, which is a very strong result. However, assuming that x^* is the solution of the semismooth Newton method, the assumption that all elements $H \in \partial_B F(x^*)$ have to be regular is essential throughout the convergence analysis of the semismooth Newton method. This assumption may be challenging in real applications. Prior the convergence results, we need to verify the regularity of all $H \in \partial_B F(x)$ for all x in a neighborhood of x^*.

Lemma 3.2.6. Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be locally Lipschitz-continuous and B-differentiable in x^* and let all $H \in \partial_B F(x^*)$ be regular. Then there exist $\rho > 0$ and a neighborhood U_{x^*} of x^*, such that the following conditions hold for any $x \in U_{x^*}$ and $H \in \partial_B F(x)$:

$$H \text{ is regular and } \|H^{-1}\| \leq \rho.$$

For the proof of Lemma 3.2.6, we refer to Qi (1993).

In applying Lemma 3.2.6, the following two theorems give convergence results for the semismooth Newton method.

Theorem 3.2.7 (Superlinear convergence)**.** Let x^* be the solution of $F(x) = 0$ and let $F : \mathbb{R}^n \to \mathbb{R}^n$ be semismooth. Moreover let $H \in \partial_B F(x^*)$ be regular. Then, the semismooth Newton method is well-defined and converges in a neighborhood $U(x^*)$ of x^* superlinearly to x^*, i.e. there exists a sequence $\{c_k\}_{k \in \mathbb{N}}$ with $c_k \to 0$ and

$$\|x^{k+1} - x^*\| \leq c_k \|x^k - x^*\| \text{ for all } k = 0, 1, 2, \ldots.$$

For the proof of Theorem 3.2.7, we refer to Qi (1993, Theorem 3.1).

Theorem 3.2.8 (Quadratic convergence)**.** Let x^* be the solution of $F(x) = 0$ and let $F : \mathbb{R}^n \to \mathbb{R}^n$ be strongly semismooth. Moreover let $H \in \partial_B F(x^*)$ be regular. Then, the semismooth Newton method is well-defined and converges in a neighborhood $U(x^*)$ of x^* quadratically to x^*, i.e. there exists a constant $c > 0$ with

$$\|x^{k+1} - x^*\| \leq c\|x^k - x^*\|^2 \text{ for all } k = 0, 1, 2, \ldots.$$

Proof. Since F is locally Lipschitz-continuous and B-differentiable, Shapiro (1990) proved the directional differentiability of F. Moreover, since F is strongly semismooth, the following equation holds for $x^0 \in \mathbb{R}^n$ due to Definition 3.2.4:

$$Hr - F'(x^0;r) = \mathcal{O}(\|r\|^2)\ \forall H \in \partial_B F(x^0 + r).$$

In this way, the proof is completed using Qi (1993, Lemma 2.3 and Theorem 3.1). □

Aside from the local convergence results proved in Theorems 3.2.7 and 3.2.8, global convergence results have also been published, and these basically rely on the usage of a step-size strategy. Regarding this, a brief discussion is given in the following paragraph.

Step-size strategy

As mentioned before, the results in Theorems 3.2.7 and 3.2.8 holds without step-size rule. In order to achieve global convergence, Pang (1990) proposes to include a step-size rule to ensure stability and robustness. In this way, the assumption that the generalized Newton equation in (3.10) has at least one solution at every iteration is essential (cf. Pang, 1990, Chapter 5). A common choice of a step-size strategy is to choose a specific version of the Armijo step-size rule primarily published by Armijo (1966). The step-size rule reduces the impact of the solution d^k of the generalized Newton equation in (3.10) on the update step, since a scalar factor $\alpha_k \in (0, 1]$ is included, e.g.

$$x^{k+1} = x^k + \alpha_k d^k. \tag{3.12}$$

In general, the step-size strategy increases the stability of the algorithm, but it also increases the number of iterations and therefore the computational burden. One possible global convergence result for semismooth Newton methods with step-size rule is presented in Qi and Sun (1993) as an extension of the Newton-Kantorovich theorem (cf. Ortega, 1968). Another result was proposed by Qi (1993), which is based on Pang (1990). We present the method proposed by Ito and Kunisch (2009), where the step-size rule is separated into two parts based on calculations of the squared norm of F, the merit function $\theta : \mathbb{R}^n \to \mathbb{R}_{0_+}$, $\theta(x) = \|F(x)\|^2$ (see Algorithm 2). Firstly, if the normed step-size is low enough and the norm decrease within one iteration k is high enough, the step-size is set to the maximum value of $\alpha_k = 1$. If this is not the case, the step-size is reduced step by step, until Equation (3.13) is satisfied.

In comparison with Theorems 3.2.7 and 3.2.8, the global convergence results require several additional assumptions.

Theorem 3.2.9 (Global convergence)**.** Let x^* be the solution of $F(x) = 0$ and let $F : \mathbb{R}^n \to \mathbb{R}^n$ be semismooth. Moreover, let $H \in \partial_B F(x^*)$ be regular and $x^* \in \mathbb{R}^n$. Assume that the following assumptions hold:

1. The set $\mathcal{X}_{\text{sol}} := \left\{x \in \mathbb{R}^n : \|F(x)\| \le \|F(x^0)\|\right\}$ is bounded.
2. There exist $\bar{\sigma}$ and $b > 0$ such that for each $x \in \mathcal{X}_{\text{sol}}$ there exists $d \in \mathbb{R}^n$ satisfying
$$\theta'(x; d) \le \bar{\sigma}\theta(x) \text{ and } \|d\| \le b\|F(x)\|.$$
3. The following implication holds:
$$x^k \to \tilde{x} \text{ and } d^k \to \tilde{d} \text{ with } x^k \in \mathcal{X}_{\text{sol}} \;\Rightarrow\; \theta'(\tilde{x}; \tilde{d}) \le -\bar{\sigma}\theta(\tilde{x}).$$

Then the sequence $\{x^k\}$ generated by the SSN Algorithm 1 (inclusive step-size rule) is bounded. It satisfies
$$\|F(x^{k+1})\| < \|F(x^k)\| \; \forall k \ge 0,$$
and each accumulation point x^* satisfies $F(x^*) = 0$. Moreover, the sequence $\{x^k\}$ converges to x^* superlinearly.

For the proof of Theorem 3.2.9, we refer to Ito and Kunisch (2009, Theorem 2.1) and (Pang, 1990, Theorem 4).

Algorithm 2 Armijo step-size rule (armijo)

```
Input: $x^k, d^k \in \mathbb{R}^n$;  $\theta : \mathbb{R}^n \to \mathbb{R}_{0_+}, x^k \mapsto \|F(x^k)\|^2$;  $\beta, \delta, \bar{\sigma} \in (0,1)$;  $\sigma \in (0,\bar{\sigma})$.
Ensure: $\theta'(x^k; d^k) \le -\bar{\sigma}\theta(x^k)$
  if $\|d^k\| \le b\|F(x^k)\|$ and $\|F(x^k + d^k)\| \le \delta\|F(x^k)\|$
    set $\alpha_k = 1$
  else
    choose smallest number $m_k \in \mathbb{N}_0$ with
        $\theta(x^k + \beta^{m_k} d^k) - \theta(x^k) \le -\sigma\beta^{m_k}\theta(x^k)$        (3.13)
    set $\alpha_k = \beta^{m_k}$
  end if
Return: Solution $\alpha_k$
```

Finally, we can note that if we apply the step-size rule presented in Algorithm 2, the SSN method in Algorithm 1 is globally convergent under some assumptions specified in Theorem 3.2.9. Nevertheless, the assumptions are solely of a theoretical nature and their verification in real application is generally impossible. Despite this, it is desirable to choose an initial value x^0 as near as possible to the solution x^* to ensure better robustness and convergence rates. The effect of the step-size strategy in the context of the optimal multivariate and multi-domain allocation will also be discussed in the application study in Subsection 4.6.6.

3.3 Multi-criteria optimization

A major part of this thesis is the development of a method to solve optimal multivariate and multi-domain allocation problems. In contrast to optimal univariate allocation techniques as given by Equation (2.49) in Section 2.5, these multivariate allocation techniques consider more than one variable of interest. Since in optimal allocation the variance of the HT estimator for the population total is minimized, the underlying optimization problem in (2.49) needs to contain the minimization of more than one objective function. Thus, problem (2.49) passes over to a multi-criteria (or vector) optimization problem with one equality constraint and box-constraints. Due to the further extensions expained in Chapter 4, we consider various equality and inequality constraints. While the mathematical theory of multi-objective optimization is addressed in this section, both the development of the multivariate allocation method and the survey statistical application are presented and analyzed in Chapter 4. Unless otherwise stated, the following theory is based on Jahn (1986) and Ehrgott (2005), which are two main references in the context of multi-criteria optimization.

The optimal multivariate allocation problem results in the multi-criteria optimization problem

$$\min_{x \in \mathcal{X}} F(x) \tag{3.14}$$

with the feasible set

$$\mathcal{X} := \Big\{ x \in \mathbb{R}^n : h_i(x) = 0 \ (i = 1, \ldots, q_2) \text{ and } g_j(x) \leq 0 \ (j = 1, \ldots, q_3) \Big\} \subseteq \mathbb{R}^n. \tag{3.15}$$

and component-wise continuously differentiable functions

$$F : \mathbb{R}^n \to \mathbb{R}^{q_1}, \ h : \mathbb{R}^n \to \mathbb{R}^{q_2} \text{ and } g : \mathbb{R}^n \to \mathbb{R}^{q_3}. \tag{3.16}$$

The components of F are denoted with $F_i : \mathbb{R}^n \to \mathbb{R}$, i.e. $F(x) = \big(F_1(x), \ldots, F_{q_1}(x)\big)$. Thus, problem (3.14) consists of q_1 objective functions, q_2 equality constraints, and q_3 inequality constraints. The box-constraints of (2.49) are contained in function g. Moreover, the image of $\mathcal{X}$ under F is denoted by

$$\mathcal{Y} := F(\mathcal{X}) := \Big\{ y \in \mathbb{R}^{q_1} : y = F(x) \text{ for } x \in \mathcal{X} \Big\} \subseteq \mathbb{R}^{q_1}. \tag{3.17}$$

Since different interpretations for the operator $\min_{x \in \mathcal{X}}$ can be associated in multi-criteria optimization, we have to characterize the definition of the type of optimality that we want to apply for solving problem (3.14). In that regard, the type of optimality used is closely related to the application-specific preferences. The required theoretical framework is presented hereafter.

3.3.1 Characterization of optimality

Since for $q_1 > 1$ there is no canonical order on $\mathbb{R}^{q_1}$ as there is on $\mathbb{R}^1$, we have to define how to compare the q_1 objective functions $F_1(x), \ldots, F_{q_1}(x)$ for all $x \in \mathcal{X}$ in order to classify the meaning of $\min_{x \in \mathcal{X}}$ in multi-criteria optimization. To do this, we use the classification of Ehrgott (2005, p. 17), according to which a multi-criteria optimization problem is fully characterized by the following properties:

(P1) the feasible set $\mathcal{X} \subseteq \mathbb{R}^n$

(P2) the objective function $F(x) = \big(F_1(x), \dots, F_{q_1}(x)\big)$

(P3) the objective space $\mathbb{R}^{q_1}$

(P4) a model map $\phi : \mathbb{R}^{q_1} \to \mathbb{R}^\nu$ with $\nu \in \mathbb{N}$

(P5) an ordered space $(\mathbb{R}^\nu, \preceq)$.

The properties (P1) to (P3) are clearly given in advance in the considered optimal multivariate allocation framework. The feasible set (P1) is defined with the aid of the restrictions considered in the problem, e.g. the equality constraint and box-constraints of problem (2.49). The objective functions (P2) are given by the q_1 variance functions $\text{Var}(\hat{\tau}_{y_i}^{\text{StrRS,HT}})$ $(i = 1, \dots, q_1)$ defined in (2.50). Their image space yields the objective space (P3). Thus, we have to pay heed to (P4) and (P5). The determination of both the model map ϕ and the ordered space $(\mathbb{R}^\nu, \preceq)$ fully characterizes the concept of optimality applied for the solution of multi-criteria optimization problems. Thus, the choice of the scalarization technique for the optimal multivariate allocation problem (see Subsection 4.2.3) is accompanied by the determination of a suitable model map and an ordered space $(\mathbb{R}^\nu, \preceq)$.

Initially, we will have a closer look at the model map ϕ. The function maps the value of the objective function vector $F(x) \in \mathbb{R}^{q_1}$ to the space $\mathbb{R}^\nu$ in which the comparisons for the determination of the minimum are made. In this thesis, we consider three cases of ϕ given by

- the identity map $\phi_{\text{id}} : \mathbb{R}^{q_1} \to \mathbb{R}^{q_1},\ F(x) \mapsto F(x)$,
- the p-norm $\phi_{\text{norm}} : \mathbb{R}^{q_1} \to \mathbb{R}_{0_+},\ F(x) \mapsto \left(\sum_{i=1}^{q_1} |F_i(x)|^p\right)^{\frac{1}{p}}$, and
- the maximum $\phi_{\text{max}} : \mathbb{R}^{q_1} \to \mathbb{R}_{0_+},\ F(x) \mapsto \max\limits_{i=1,\dots,q_1} F_i(x)$.

As we will see in Subsection 4.2.3, these three cases are strongly related to the choice of a scalarization technique for the optimal multivariate allocation problem.

Subsequent to the choice of ϕ, the ordered space $(\mathbb{R}^\nu, \preceq)$ has to be determined, i.e. a *partial ordering relation* $\preceq$ has to be chosen. Their formal definition requires the knowledge or some theory of geometric structures for the Euclidean space $\mathbb{R}^\nu$ (cf. Jahn, 1986, Definition 1.1), which are presented in the following paragraphs.

Definition 3.3.1.

1. Each subset $R \neq \emptyset$ of the product space $\mathbb{R}^\nu \times \mathbb{R}^\nu$ is called *binary relation R* on $\mathbb{R}^\nu$.
2. Let $C \neq \emptyset$ be a subset of the $\mathbb{R}^\nu$. The set C is called a *cone*, if $x \in C,\ \lambda \geq 0 \ \Rightarrow\ \lambda x \in C$. A cone C is *pointed*, if $C \cap (-C) = \{0_{\mathbb{R}^\nu}\}$, where $0_{\mathbb{R}^\nu} := (0, \dots, 0)^T \in \mathbb{R}^\nu$.

The most trivial cone on the $\mathbb{R}^\nu$ is the component-wise positive cone

$$C_{\mathbb{R}^\nu} := \Big\{x \in \mathbb{R}^\nu : x_i \geq 0\ \forall i = 1, \dots, \nu\Big\} = \mathbb{R}^\nu_{0_+}. \tag{3.18}$$

We can easily show that the cone $C_{\mathbb{R}^\nu}$ defined in (3.18) is pointed, convex, and has the property of $0_{\mathbb{R}^\nu} \in C$. Thus, we are able to define a partial ordering relation on $\mathbb{R}^\nu$ (cf. Jahn, 1986, Definition 1.16).

Definition 3.3.2. Each relation $\preceq$ on $\mathbb{R}^\nu$ is called a *partial ordering relation* on $\mathbb{R}^\nu$, if the following implications hold for $z_1, z_2, z_3, z_4 \in \mathbb{R}^\nu$:

1. $z_1 \preceq z_1$ (reflexive),
2. $z_1 \preceq z_2,\ z_2 \preceq z_3 \ \Rightarrow\ z_1 \preceq z_3$ (transitive),
3. $z_1 \preceq z_2,\ z_3 \preceq z_4 \ \Rightarrow\ z_1 + z_3 \preceq z_2 + z_4$ (additive), and
4. $z_1 \preceq z_2,\ \alpha > 0 \ \Rightarrow\ \alpha z_1 \preceq \alpha z_2$ (scalar multiplicative).

To illustrate Definition 3.3.2, some examples are given in Example 3.3.3.

Example 3.3.3. This example shows four relations on $\mathbb{R}^\nu$, one of which is a partial ordering relation and three of which are no partial ordering relations.

1. Component-wise order $z_1 \leq_c z_2 \quad :\Leftrightarrow \quad z_{1i} \leq z_{2i}$ for all $i = 1, \ldots, \nu$

   $\leq_c$ is a partial ordering relation on $\mathbb{R}^\nu$. For $\nu = 1$, it is equal to the standard $\leq$ (lower or equal) relation.

2. Strict component-wise order $z_1 <_c z_2 \quad :\Leftrightarrow \quad z_{1i} < z_{2i}$ for $i = 1, \ldots, \nu$

   $<_c$ is no partial ordering relation on $\mathbb{R}^\nu$, as it is not reflexive, e.g. $2 \not<_c 2$.

3. p-norm relation $z_1 \leq_p z_2 \quad :\Leftrightarrow \quad \|z_1\|_p \leq \|z_2\|_p$

   $\leq_p$ is no partial ordering relation, since it is not additive. Let $\nu = 2$:

   $$(1.1, 1.1) \leq_p (2, 0) \ \text{ and } \ (1.1, 1.1) \leq_p (0, 2), \ \text{ but } \ (2.1, 2.1) \not\leq_p (2, 2).$$

4. max-relation $z_1 \leq_{\max} z_2 \quad :\Leftrightarrow \quad \max_{i=1,\ldots,\nu} z_{1i} \leq \max_{i=1,\ldots,\nu} z_{2i}$

   $\leq_{\max}$ is no partial ordering relation, since it is not additive. Let $\nu = 2$:

   $$(4, 4) \leq_{\max} (1, 5) \ \text{ and } \ (3, 3) \leq_{\max} (5, 1), \ \text{ but } \ (7, 7) \not\leq_{\max} (6, 6).$$

Nevertheless, the p-norm relation and the max-relation are partial ordering relations on $\mathbb{R}^1$.

The fact that the p-norm relation and the max-relation are not partial ordering relations on the $\mathbb{R}^\nu$ affects the interpretation of the scalarization technique in Chapter 4. In particular, the characterization of optimality for problem (3.14) is not possible with the model map $\phi \equiv \text{id}$ in combination with the p-norm relation or the max-relation. Thus, we will have to choose another model map ϕ for these cases to avoid this inconsistency.

The following theorem states some important properties of partial ordering relations on $\mathbb{R}^\nu$.

Theorem 3.3.4.

1. If $\preceq$ is a *partial ordering relation* on $\mathbb{R}^\nu$, then the set $C_{\mathbb{R}^\nu} = \mathbb{R}^\nu_{0+}$ defined in (3.18) is a convex cone. If, in addition, $\preceq$ is antisymmetric, i.e. $(z_1 \preceq z_2,\ z_2 \preceq z_1 \ \Rightarrow\ z_1 = z_2)$ holds for $z_1, z_2 \in \mathbb{R}^\nu$, the convex cone $C_{\mathbb{R}^\nu}$ is pointed.
2. If C is a convex cone in $\mathbb{R}^\nu$ and $z_1, z_2 \in \mathbb{R}^\nu$, then the relation $z_1 \preceq z_2 \ :\Leftrightarrow\ z_2 - z_1 \in C$ is a partial ordering relation on $\mathbb{R}^\nu$. If, in addition, C is pointed, then $\preceq$ is antisymmetric.

A convex cone characterizing a partial ordering in $\mathbb{R}^\nu$ is called an *ordering cone*. For the proof of Theorem 3.3.4, we refer to Jahn (1986, Theorem 1.18).

Finally, we are able to characterize the ordered space $(\mathbb{R}^\nu, \preceq)$.

Definition 3.3.5. The Euclidean space $\mathbb{R}^\nu$ equipped with a partial ordering relation $\preceq$ and an ordering cone C is called a *partially ordered linear space* or simply *ordered space* $(\mathbb{R}^\nu, \preceq)$.

Since we have linked the Euclidean space $\mathbb{R}^\nu$ with a partial ordering relation $\preceq$, we are able to define minimal (or maximal) elements of a subset of this space concerning the underlying partial ordering relation in analogy to Jahn (1986, Definitions 4.1, 4.8, and 4.12).

Definition 3.3.6. Let Z be a non-empty subset of $\mathbb{R}^\nu$ with an ordering cone C. Then:

1. An element $z^* \in Z$ is called a *minimal element* of Z, if
$$\Big(\{z^*\} - C\Big) \cap Z \subseteq \{z^*\} + C.$$
If C is pointed, the equation can be replaced by $\Big(\{z^*\} - C\Big) \cap Z = \{z^*\}$.
2. An element $z^* \in Z$ is called a *strongly minimal element* of Z, if
$$Z \subseteq \{z^*\} + C.$$
3. An element $z^* \in Z$ is called a *weakly minimal element* of Z, if
$$\Big(\{z^*\} - \text{cor}(C)\Big) \cap Z = \emptyset,$$
where $\text{cor}(C)$ is the algebraic interior of C.

Based on the defined structures, the multi-criteria optimization problem (3.14) can be fully characterized with (P1) to (P5). Since the first three items, i.e. the feasible set $\mathcal{X}$, the objective function F, and the objective space $\mathbb{R}^{q_1}$ arise out of the problem formulation, the decision-maker has to choose a suitable model map $\phi : \mathbb{R}^{q_1} \to \mathbb{R}^\nu$ and an ordered space $(\mathbb{R}^\nu, \preceq)$ with a partial ordering relation $\preceq$. We denote this classification of the multi-criteria optimization problem (3.14) by
$$(\mathcal{X}, F, \mathbb{R}^{q_1})/\phi/(\mathbb{R}^\nu, \preceq). \tag{3.19}$$
Using the characterization defined in (3.19), we are now able to interpret the minimal solution $x^* \in \mathcal{X}$ of problem (3.14) as the inverse image of the minimal element $F(x^*)$ of the (from ϕ dependent) image space $\phi(\mathcal{Y}) = \phi(F(\mathcal{X}))$, as defined in Definition 3.3.6.

Definition 3.3.7. Let the multi-criteria optimization problem (3.14) be given of class $(\mathcal{X}, F, \mathbb{R}^{q_1})/\phi/(\mathbb{R}^{\nu}, \preceq)$ with the feasible set $\mathcal{X}$ (see (3.15)) and functions (3.16). Let the model map $\phi : \mathbb{R}^{q_1} \to \mathbb{R}^{\nu}$ be given with the ordered space $(\mathbb{R}^{\nu}, \preceq)$ and an ordering cone $C_{\mathbb{R}^{\nu}}$. Then:

1. An element $x^* \in \mathcal{X}$ is called a *minimal solution* of problem (3.14), if $\phi(F(x^*))$ is a minimal element of the image set $\phi(\mathcal{Y}) = \phi(F(\mathcal{X}))$.

2. An element $x^* \in \mathcal{X}$ is called a *weakly minimal solution* of problem (3.14), if $\phi(F(x^*))$ is a weakly minimal element of the image set $\phi(\mathcal{Y}) = \phi(f(\mathcal{X}))$.

As a consequence of Definition 3.3.7, the set of minimal solutions of the multi-criteria optimization problem (3.14) depends on the choice of the model map ϕ and the partial ordering relation $\preceq$, which represent the choice of the scalarization technique for the optimal multivariate allocation problem in Chapter 4. The choice is limited, as $\preceq$ has to be a partial ordering relation. Hence, we are not allowed to choose a p-norm or max-relation in combination with $\phi \equiv \text{id}$ (see Example 3.3.3). However, as the p-norm and the max-relation are common scalarization techniques for multivariate allocation problems (see Subsection 4.2.3), it is highly necessary to consider these in the developments in Chapter 4. Thus, we choose different model maps ϕ in order to incorporate these scalarization techniques. To conclude, the following characterizations of optimality are considered in Chapter 4:

(EF)	class of efficient solution	$(\mathcal{X}, f, \mathbb{R}^{q_1})/\text{id}/(\mathbb{R}^{q_1}, \leq_c)$,
(norm)	class of p-norm optimization	$(\mathcal{X}, f, \mathbb{R}^{q_1})/\phi_{\text{norm}}/(\mathbb{R}^1, \leq)$, and
(max)	class of min-max optimization	$(\mathcal{X}, f, \mathbb{R}^{q_1})/\phi_{\text{max}}/(\mathbb{R}^1, \leq)$.

As we will see in Subsection 3.3.2, the class of efficient solutions (EF) coincides with the theory of efficient solutions or Pareto optimal solutions. Since the considered ordered space is of dimension q_1 for this class, the resulting problem is still a multi-criteria problem. The existence of a solution is shown in Subsection 3.3.3 and optimality conditions are given in Subsection 3.3.4. Thereafter, we prove in Subsection 3.3.5 that it is sufficient to solve a weighted sum scalarized problem for all combinations of weights to achieve the whole set of optimal solutions. The problems of the classes of p-norm optimization (norm) and min-max optimization (max) are real-valued problems, since the image space is a subset of $\mathbb{R}^1$. Thus, the theory of nonlinear optimization presented in Section 3.1 can be applied. We note that the objective function of the problem class (max) is not differentiable, so that certain theory of non-smooth optimization has to be applied (cf. Geiger and Kanzow, 2002, Chapter 6). To omit this in Chapter 4, we only compute approximate solutions of class (max) by choosing a high p for the corresponding problem of class (norm). Several tests have shown that the approximations are precise enough for almost all survey statistical applications.

In Section 4.4, we utilize a connection between the class (EF) and the class (norm) to solve problems of class (norm) in an appropriate time. The strategy takes advantage of the fact that in the setting of Chapter 4 the optimal solution of a problem of class (norm) is an element of the set of optimal solutions of the problem of class (EF). This is exploited to restrict the feasible set of the problem of class (norm) to the set of optimal solutions of the corresponding problem of class (EF).

3.3.2 Efficient solutions and Pareto optimality

In this subsection, we consider the special class (EF) of problem (3.14) given by

$$(\mathcal{X}, f, \mathbb{R}^{q_1})/\mathrm{id}/(\mathbb{R}^{q_1}, \leq_c) \tag{3.20}$$

as defined in Subsection 3.3.1, i.e. we choose the component-wise partial ordering relation presented in Example 3.3.3, which bridges the gap between general optimal solutions of (3.14) and the theory of efficient solutions or *Pareto optimality*, well-known from the microeconomic theory (cf. Varian, 2010, pp. 15 ff.). Since $\phi = \mathrm{id}$ and therefore $\nu = q_1$, the consideration of the model map can be ignored hereafter. Accordingly, the definition of Pareto optimality is given in the following definition.

Definition 3.3.8. Let the multi-criteria optimization problem (3.14) be given of class (3.20) with the feasible set (3.15) and functions (3.16). Then $x^* \in \mathcal{X}$ is Pareto optimal (or Pareto-efficient), if and only if there is no $x \in \mathcal{X}$ with

$$F_i(x) \leq F_i(x^*) \;\forall i = 1, \ldots, q_1 \;\text{ and }\; F_j(x) < F_j(x^*) \text{ for at least one } j \in \{1, \ldots, q_1\} \,.$$

The image set $\mathcal{Y}_{\text{OPT}} := \{F(x^*) : x^* \text{ is Pareto optimal for (3.14)}\} \subseteq \mathbb{R}^{q_1}$ of all Pareto optimal solutions is called *Pareto frontier*.

By means of the geometric structures presented in Subsection 3.3.1, we are now able to prove the equivalence of general minimal solutions of multi-criteria optimization in Definition 3.3.7 and of the Pareto optimal solutions in Definition 3.3.8 for class (3.20).

Theorem 3.3.9. Let the multi-criteria optimization problem (3.14) be given of class (3.20). Then, $x^* \in \mathcal{X}$ is an optimal solution of (3.14), if and only if x^* is Pareto optimal for (3.14).

Proof. Using Definition 3.3.7, x^* is an optimal solution of problem (3.14), if and only if $F(x^*)$ is a minimal element of the image space $F(\mathcal{X})$. Moreover, the core $C_{\mathbb{R}^{q_1}} = \mathbb{R}^{q_1}_{0+}$ is pointed. Thus, it is sufficient with Definition 3.3.6 to verify the following equivalence:

$$\Big(\{F(x^*)\} - C_{\mathbb{R}^{q_1}}\Big) \cap F(\mathcal{X}) = \Big\{F(x^*)\Big\} \quad \Leftrightarrow \quad x^* \text{ is a Pareto optimal solution.}$$

Let $x^* \in \mathcal{X}$ be a minimal solution of problem (3.14), then

$$\Big(\{F(x^*)\} - C_{\mathbb{R}^{q_1}}\Big) \cap F(\mathcal{X}) = \Big\{F(x^*)\Big\},$$

which is equivalent to

$$\Big(\{F(x^*)\} - C_{\mathbb{R}^{q_1}}\Big) \cap \Big\{F(x) : x \in \mathcal{X}\Big\} = \Big\{F(x^*)\Big\}.$$

Then, a shift of $-F(x^*)$ yields

$$-C_{\mathbb{R}^{q_1}} \cap \Big\{F(x) - F(x^*) : x \in \mathcal{X}\Big\} = \Big\{0_{\mathbb{R}^{q_1}}\Big\}.$$

This equality holds if and only if x^* is Pareto optimal as defined in Definition 3.3.8, which completes the proof. □

In Theorem 3.3.9, we have shown that the set of all optimal solutions of a multi-criteria optimization problem (3.14) of class (3.20) is equal to the set of all Pareto optimal solutions.

3.3.3 Existence of solutions

In this subsection, assumptions which guarantee the existence of at least one Pareto optimal solution for problem (3.14) of class (3.20) are presented. In this way, the property *compactness* of subsets of the Euclidean space plays a significant role (for the general definition, see Jahn, 1986, Definition 1.29). In general, a closed and bounded subset of the Euclidean space is referred to as a compact set.

Theorem 3.3.10 (Existence of minimal element)**.** Let Z be a non-empty subset of $\mathbb{R}^{q_1}$ with $C_{\mathbb{R}^{q_1}} = \mathbb{R}^{q_1}_{0+}$. Then, there exists at least one minimal element of the set Z, if Z has a compact section, i.e there is a $z \in Z$ for which $S_Z := (\{z\} - C) \cap Z$ is non-empty and compact.

Proof. Special case of Jahn (1986, Theorem 6.3) with the partially ordered topological linear space $\mathbb{R}^{q_1}$. $C_{\mathbb{R}^{q_1}}$ is a closed set as a closed subset of the closed set $\mathbb{R}^{q_1}$. □

In order to simplify the verification of Theorem 3.3.10 for the multi-criteria optimization problem, we deploy the following lemma.

Lemma 3.3.11 (Existence of Pareto optimal solution)**.** Let the multi-criteria optimization problem (3.14) be given of class (3.20). Moreover, let $\mathcal{X} \neq \emptyset$ be closed and bounded. Then there exists at least one Pareto optimal solution of (3.20).

Proof. Since $\mathcal{X} \neq \emptyset$ and F is a continuous function, it follows that $F(\mathcal{X}) \neq \emptyset$. Moreover, let $F(x) \in F(\mathcal{X})$ for $x \in \mathcal{X}$. Then, the section

$$S_{F(\mathcal{X})} := (\{F(x)\} - C_{\mathbb{R}^{q_1}}) \cap F(\mathcal{X})$$

is a non-empty set, as $0_{\mathbb{R}^{q_1}} \in C_{\mathbb{R}^{q_1}}$. Moreover, the ordering cone $C_{\mathbb{R}^{q_1}} = \mathbb{R}^{q_1}_{0+}$ is a closed set as a closed subset of the closed set $\mathbb{R}^{q_1}$. Therefore,

$$\{F(x)\} - \mathbb{R}^{q_1}_{0+}$$

is also a closed set. Since $\mathcal{X}$ is a closed set by assumption, its image $F(\mathcal{X})$ under a continuous function F is a closed set as well. Furthermore, $F(\mathcal{X})$ is bounded because F is continuous and $\mathcal{X}$ is closed and bounded. Consequently $\mathcal{X}$ is compact, and the image spaces of continuous function with a compact feasible set are bounded. This implies that the non-empty section

$$S_{F(\mathcal{X})} = (\{F(x)\} - \mathbb{R}^{q_1}_{0+}) \cap F(\mathcal{X})$$

is a closed and bounded set, and therefore it is a compact set. The existence of at least one Pareto optimal solution of (3.20) follows by Theorem 3.3.10. □

In applying Lemma 3.3.11, we are now able to prove the existence of a Pareto optimal solution of problem (3.14) of class (3.20), if the objective function is continuous and the feasible set is closed and bounded.

3.3.4 Optimality conditions

Since we can verify the existence of a Pareto optimal solution for problem (3.14) of class (3.20) by Lemma 3.3.11, we now present optimality conditions applying a Lagrangian approach in analogy to the KKT-conditions for the classical (single-objective) optimization presented in Section 3.1. The following theorem represents necessary optimality conditions for the multi-criteria optimization problem in (3.14).

Theorem 3.3.12 (Necessary optimality condition)**.** Let the optimization problem (3.14) be given of class (3.20) with the feasible set $\mathcal{X}$ (see (3.15)) and functions defined in (3.16). Moreover, let $x^* \in \mathcal{X}$ be a weakly minimal solution of problem $\min_{x\in\mathcal{X}} F(x)$. Let F and g have partial derivatives at x^*, and let h be continuously partially differentiable at x^*. Let the MFCQ condition (cf. Definition 3.1.6) hold in $x^* \in \mathcal{X}$. Then, there exists multipliers $\lambda_i \geq 0$ ($i = 1, \dots, q_1$; with $\lambda_i > 0$ for at least one i), $\mu_j \geq 0$ ($j \in I(x^*)$) and $\eta \in \mathbb{R}^{q_2}$ with

$$\sum_{i=1}^{q_1} \lambda_i \nabla F_i(x^*) + \sum_{j\in I(x^*)} \mu_j \nabla g_j(x^*) + \sum_{l=1}^{q_2} \eta_l \nabla h_l(x^*) = 0_{\mathbb{R}^n}. \tag{3.21}$$

For the proof of Theorem 3.3.12, we refer to Jahn (1986, Theorem 7.8).

As it is the case in single-criteria optimization, the convexity assumptions are necessary for the formulation of sufficient optimality conditions (see Theorem 3.1.7) in multi-objective optimization. The convexity (quasi- or pseudo-convexity) of real-valued functions is defined in Definition 3.1.4. Regarding this, we can formulate sufficient optimality conditions for (3.14).

Theorem 3.3.13 (Sufficient optimality condition)**.** Let the optimization problem (3.14) be given of class (3.20) with the feasible set $\mathcal{X}$ (see (3.15)) and functions defined in (3.16). Let x^* be given, and assume that F, g, and h have partial derivatives at x^*. Let the functions $F_1, \dots, F_{q_1}$ be pseudo-convex at x^*, and let the functions h_l, $-h_l$ ($l = 1, \dots, q_2$) and g_j ($j \in I(x^*)$) be quasi-convex at x^*. If there exists Lagrangian multipliers $\lambda_i \geq 0$ ($i = 1, \dots, q_1$; with $\lambda_i > 0$ for at least one i), $\mu_j \geq 0$ ($j \in I(x^*)$) and $\eta \in \mathbb{R}^{q_2}$ with

$$\sum_{i=1}^{q_1} \lambda_i \nabla F_i(x^*) + \sum_{j\in I(x^*)} \mu_j \nabla g_j(x^*) + \sum_{l=1}^{q_2} \eta_l \nabla h_l(x^*) = 0_{\mathbb{R}^n}, \tag{3.22}$$

then x^* is a weakly minimal solution of problem $\min_{x\in\mathcal{X}} F(x)$ of class (3.20).

For the proof of Theorem 3.3.13, we refer to Jahn (1986, Corollary 7.24).

3.3.5 Pareto optimization and weighted sum scalarization

In order to numerically solve multi-criteria optimization problems of the form (3.14) of class (3.20), we have to scalarize the q_1 objective functions to obtain a real-valued objective function. After this, the numerical solver can be applied to the resulting single objective optimization problem. With regard to the class (3.20), we focus on the *weighted sum* scalarization which contains the property that we are able to derive results from the relationships between Pareto optimal solutions of the multi-objective problem (3.14) and the scalarized problem.

Definition 3.3.14. Let

$$\min_{x\in\mathcal{X}} F(x) \tag{3.23}$$

be the multi-criteria optimization problem with component-wise continuously differentiable objective function $F : \mathbb{R}^n \to \mathbb{R}^{q_1}$ and non-empty feasible set $\mathcal{X} \subseteq \mathbb{R}^n$. Then, the single-objective optimization problem

$$\min_{x\in\mathcal{X}} \sum_{i=1}^{q_1} w_i F_i(x) \tag{3.24}$$

with weights $w = (w_1, \dots, w_{q_1})^T \in \mathbb{R}^{q_1}$ is called *weighted sum optimization problem.*

Although problem (3.24) can be solved numerically (which is not possible with problem (3.23)), it is in general not possible to find all Pareto optimal points for a multi-criteria optimization problems by solving the weighted sum problem (3.24). While the sufficient condition of Theorem 3.3.15 holds in a very general setting, see for example Folks and Antle (1965), this is not true for the necessary condition of Theorem 3.3.17.

Theorem 3.3.15 (Sufficient Condition)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ and let $F_i : \mathcal{X} \to \mathbb{R}$, $i = 1, \dots, q_1$. For every optimal solution x^* of (3.24) with weights $w \in \mathbb{R}^{q_1}$, the following statements hold:

1. x^* is a weakly Pareto optimal solution for problem (3.23) if $w \geq 0$.
2. x^* is a Pareto optimal solution for problem (3.23) if $w > 0$.

For the proof of Theorem 3.3.15, we refer to Ehrgott (2005, Proposition 3.9).

In the following, we show that under convexity assumptions it is possible to find *all* Pareto optimal points by solving a weighted sum problem.

Lemma 3.3.16. Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex, and let $F_i : \mathcal{X} \to \mathbb{R}$, $i = 1, \dots, q_1$, be convex functions. Then the set $C_+(F) := \left\{(F_1(x), \dots, F_{q_1}(x))^T | x \in \mathcal{X}\right\} + \mathbb{R}^{q_1}_{0+}$ is convex.

For the proof of Lemma 3.3.16, we refer to Jahn (1986, Theorem 2.6).

Theorem 3.3.17 (Necessary Condition)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex, and let $F_i : \mathcal{X} \to \mathbb{R}$ for $i = 1, \dots, q_1$ be convex functions. Then, for each Pareto optimal solution x^* of the problem (3.23), there exist weights $w \in \mathbb{R}^{q_1}_{0+} \setminus \{0_{\mathbb{R}^{q_1}}\}$ such that x^* is an optimal solution of the weighted sum problem (3.24).

Proof. Using the convexity of the objective function, Lemma 3.3.16 shows that the set $C_+(F)$ mentioned in the lemma is convex. Using this property, the result follows directly from Jahn (1986, Theorem 5.4). □

We conclude that if the convexity assumption of Theorem 3.3.17 holds for the multi-criteria optimization problem (3.23) of class (3.20), we can apply the theorem to prove that the solutions of the single-objective optimization problem (3.24) for all combinations of weights describe the entire Pareto frontier of the multi-criteria optimization problem (3.23). Although it should be noted that the accuracy of the Pareto frontier is determined by the discretization of the weights. This strategy is applied to the optimal multivariate and multi-domain allocation problem in Chapter 4.

Chapter 4

Optimal Multivariate and Multi-domain Allocation

4.1 Motivation and issues

The aim of modern surveys is to provide accurate information for several variables of interest, simultaneously. In light of urban audits or regional policies, these figures need to be made available not only for the population itself, but also for certain areas and strata. This results in a stratified sampling design (see Section 2.3) with several stratification levels defined by both region and content, which possibly yields a vast number of cross-classification strata.

With regard to the number of variables of interest and the required estimations on several stratification levels, the allocation of a fixed sample size $n_{\text{s}} \in \mathbb{N}$ (see Section 2.5) to these strata is challenging for a number of reasons. Due to a limited total sample size and several possibly conflicting goals, we also need to make decisions regarding the priorities while also trying concurrently to achieve the optimum of combining all the requirements. With this notion, the development of an optimal allocation method, which considers both various possibly complementary or conflicting variables of interest (*multivariate*) as well as multiple, regional, and context-specific stratification levels (*multi-domain*), is the major research question of this chapter. Consequently, such a problem is denoted as *optimal multivariate and multi-domain allocation* (`MMDopt`). The stratified random sampling design provides an adequate basis that allows the integration of further optimization techniques while practical settings with various constraints might be considered.

In accordance with the notation presented in Section 2.1, we assume a finite population $\mathcal{U}$ of size N with disjoint cross-classification strata $h = 1, \dots, H$. For q_1 variables of interest y_i ($i = 1, \dots, q_1$), the population totals are denoted by τ_{y_i}. In StrRS, the unbiased HT estimator $\hat{\tau}_{y_i}^{\text{StrRS}}$ for the total of variable y_i is given by Equation (2.16). Its variance can be calculated by Equation (2.17). As optimal allocation is based on the minimization of variances of HT estimators, external auxiliary data is required for the formulation of the objective functions, which is requested to be highly correlated with the variables of interest. In practice, this data may be provided by other survey, registers, or adequate proxies. In this thesis, we tacitly assume that this information is either available in the form of adequate proxies for the stratum-specific variances S_{ih}^2 $(i = 1, \dots, q_1$ and $h = 1, \dots, H)$ or can be computed using the auxiliary data. A further investigation of proxy quality and its implications in real applications is discussed in the application study in Section 4.6, whereas the theoretical analysis of this topic based on robustness and sensitivity is beyond the scope of this thesis. However, it should be be mentioned that we assume the strata to be fixed before the sampling process. Thus, we exclude optimal stratification strategies, such as the *cum* $\sqrt{f}$ *rule* suggested by Dalenius and Hodges (1959) and the *equal aggregate* σ *rule* proposed by Wright (1983).

As starting points for `MMDopt`, we employ the optimal allocation of Tschuprow (1923) and Neyman (1934) in Equation (2.48) as well as the box-constrained optimal allocation of Gabler et al. (2012) and Münnich et al. (2012c) in Equation (2.49). Box-constraints for the stratum-specific sample sizes are vital for the consistency of our model. Upper constraints $M_h \leq N_h$ need to be introduced to avoid overallocation in strata where n_h exceeds N_h. A further reduction of M_h allows for the control of sampling fractions, for example avoiding highly different response burdens in various strata or areas. In addition, M_h prevents a stratum-specific full census, which in general may be undesirable in several applications. A full census in some strata is worthy of discussion particularly in the context of business surveys, as the response burden should not be too much different for all companies (see also the judgment of the German Federal Administrative Court; BVerwG, 03/15/2017, 8 C 6.16). As zero sample sizes in single strata lead to biased estimates and the variance computations of the total estimates require stratum-specific sample sizes of at least two units, a lower constraint $m_h \geq 2$ is applied. Altogether, the optimal univariate allocation problem under box-constraints defined in (2.49) is given by

$$\begin{aligned} &\min_{n \in \mathbb{R}_+^H} \ \text{Var}(\hat{\tau}_y^{\text{StrRS}}) \\ &\text{s.t.} \sum_{h=1}^{H} n_h = n_{\text{s}} \\ &\qquad 2 \leq m_h \leq n_h \leq M_h \leq N_h \ \ \forall h = 1, \dots, H \end{aligned} \tag{4.1}$$

for variable y. In the multivariate generalization of the optimal allocation problem (4.1), q_1 variances $\text{Var}(\hat{\tau}_{y_1}^{\text{StrRS}}), \dots, \text{Var}(\hat{\tau}_{y_{q_1}}^{\text{StrRS}})$ of the total estimates of variables $y_1, \dots, y_{q_1}$ are considered simultaneously. This results in a multi-criteria optimization problem with q_1 objectives

$$\begin{aligned} &\min_{n \in \mathbb{R}_+^H} \ \left(\text{Var}(\hat{\tau}_{y_1}^{\text{StrRS}}), \dots, \text{Var}(\hat{\tau}_{y_{q_1}}^{\text{StrRS}})\right) \\ &\text{s.t.} \sum_{h=1}^{H} n_h = n_{\text{s}} \\ &\qquad 2 \leq m_h \leq n_h \leq M_h \leq N_h \ \ \forall h = 1, \dots, H. \end{aligned} \tag{4.2}$$

Trivially, this problem is also valid for a multi-domain allocation, since the H strata correspond to the cross-classifications of all stratification levels (see Figure 2.1 for an example).

Optimal multivariate allocation problems have been widely discussed in literature over the past few decades, starting in the 1950s. Dalenius (1953) discussed this problem in detail and distinguished two solution strategies. In the first approach, one or more of the variances $\text{Var}(\hat{\tau}_{y_i}^{\text{StrRS}})$,

$i = 1, \ldots, q_1$, are bounded from above and treated as constraints of an optimization problem, in which the total sample size (or the cost of the survey) is minimized. This leads to a univariate allocation problem with nonlinear constraints. In the second strategy, the variances are minimized simultaneously subject to linear size (or cost) constraints. This perspective leads to a multi-objective optimization problem with conflicting objectives, which requires an appropriate mathematical theory. In particular, an adequate notion of optimality, such as Pareto optimality, is essential. Moreover, the problem has to be transformed into a form that is solvable by optimization algorithms. Most of the literature dealing with multivariate allocation covers either one of these two formulations.

The first variant was used by Chatterjee (1968), Chatterjee (1972), and Huddleston et al. (1970). Optimal multivariate allocation problems were addressed in the same way in Kokan (1963) and supplemented by existence and uniqueness results in Kokan and Khan (1967). In introducing overhead costs, Ahsan and Khan (1982) discussed the problem with variance constraints for a more general objective function. More recently, Bankier (1988), Hohnhold (2009b), and Hohnhold (2009a) have published allocation techniques with more than one stratification level. These techniques are based on a compensation of the accuracy of total estimates on regional and population level and also belong to the first class of methods. Falorsi and Righi (2015) presented a generalized framework for defining the optimal inclusion probabilities in multivariate and multi-domain surveys. Falorsi and Righi (2008) and Falorsi and Righi (2016) introduced a solution method using a balanced sampling design. By combining aspects from both strategies, Kish (1976) proposed to link aspects of variance and cost minimization in a nonlinear model with the help of loss functions.

In the second solution strategy introduced by Dalenius (1953), the optimal multivariate allocation is treated in Folks and Antle (1965) as a multi-objective optimization problem with linear constraints. They discussed the mathematical theory of scalarization and the relationship between the multi-objective problem and the scalarized problem. Moreover, they proved a sufficiency result for the set of efficient (or Pareto optimal) solutions for the simple problem without box-constraints. Díaz-García and Ramos-Quiroga (2014) solved the multivariate allocation as a multi-objective problem as well, but with the help of stochastic programming instead. Khan et al. (2012) also used stochastic programming on a further model. Both methods lead to nonlinear integer optimization problems, which are hard to solve even for small instances. Khan et al. (2003) solved multivariate allocation problems by exploiting the separability of the objective function and by applying dynamic programming. While dynamic programming is a classical solution method for allocation problems (Arthanari and Dodge, 1981, Chapter 5), it is not very efficient in practice, as shown in the computational study of Bretthauer et al. (1999).

Since the aim of the developed method is to minimize the variances of the population totals of the variables of interest simultaneously, the structure of the method is based on the second strategy of Dalenius (1953), in which the problem is treated as a multi-objective optimization problem. A main drawback of classical optimal allocation is the neglection of the accuracy of stratum- or area-specific estimates. However, official statistics request a method combining an optimal allocation and a *well-balanced* allocation between different variables of interest and especially on different stratification levels. Falorsi and Righi (2008) dealt with this approach only by minimizing the costs instead of variances (first strategy of Dalenius, 1953). Thus, we may also consider a compensation of the accuracy of total estimates at stratum- or area-level and population level by bounding the variances (or standardized variances) of the stratum- or area-specific estimates on the one hand, and by using compensated objective functions on the other hand. The variances of the stratum-specific estimates $\hat{\tau}_{y_i h}^{\text{StrRS}}$ for variable of interest y_i can be controlled by the lower bounds m_h $(h = 1, \ldots, H)$, since the sampling design is defined by the strata. However, area-specific estimates $\hat{\tau}_{y_i l_r}^{\text{StrRS}}$ for areas l_r in stratification level $r = 1, \ldots, R$ require the consideration of additional nonlinear inequality constraints for problem (4.2), since areas are defined as unions of strata (see Section 2.3), i.e. each individual area also contains a StrRS design. These constraints allow the users to define and include upper bounds for variances or standard errors for various areas l_r and variables of interest y_i, which are denoted by $\text{Vmax}_{(i,r,l_r)} \in \mathbb{R}_+$. The constraints are given by

$$\text{Var}(\hat{\tau}_{y_i l_r}^{\text{StrRS}}) \leq \text{Vmax}_{(i,r,l_r)}. \tag{4.3}$$

We refer to these constraints as *restrictions for regional efficiency* hereinafter. Their aim is to prevent extremely high estimation errors of area-specific estimates rather than to equalize the estimation errors over all areas. Consequently, the focus still lies on the simultaneous minimization of the population total variances, which however is restricted in order to avoid unreliable estimates for critical areas and/or variables of interest, e.g. to comply with administrative laws or internal quality assurances. If equalized or compensated estimation errors over all areas are desired, the better choice is to include a weighting of the respective parts of the objective function. This yields a compensated allocation, which is further discussed in Subsection 4.2.5 and which has been applied in the German Census 2011 (cf. Münnich et al., 2012a, pp. 31 ff.).

In addition to the inequality constraints (4.3), we replace the equality constraint $\sum_{h=1}^{H} n_h = n_{\text{s}}$ of problem (4.2) by a more general formulation of q_2 affine-linear equality constraints

$$An = b \tag{4.4}$$

with $A := [1, A_2, \ldots, A_{q_2}]^T \in \mathbb{R}^{q_2 \times H}$ and $b := (n_{\text{s}}, b_2, \ldots, b_{q_2})^T \in \mathbb{R}^{q_2}$ to ensure more flexibility. This results in the optimal multivariate and multi-domain optimization problem

$$\begin{aligned} \min_{n \in \mathbb{R}_+^H} \; & \left(\text{Var}(\hat{\tau}_{y_1}^{\text{StrRS}}), \ldots, \text{Var}(\hat{\tau}_{y_{q_1}}^{\text{StrRS}})\right) \\ \text{s.t. } & An = b \\ & \text{Var}(\hat{\tau}_{y_i l_r}^{\text{StrRS}}) \leq \text{Vmax}_{(i,r,l_r)} \;\; \text{for some } r \in \{i, \ldots, R\}, l_r \in \{1, \ldots, L_r\}, i \in \{1, \ldots, q_1\} \\ & 2 \leq m_h \leq n_h \leq M_h \leq N_h \;\; \forall h = 1, \ldots, H, \end{aligned} \tag{4.5}$$

which is henceforth referred to as `MMDopt` for the case without restrictions for regional efficiency and `MMDopt.reg` (`reg`: regional restrictions) for the case with restrictions for regional efficiency (4.3). All strategies using the multi-objective perspective of the problem need to use scalarization techniques to combine the variances for the variables in one single objective function. The selection of a scalarization technique can be interpreted as the choice of a suitable decision-making function (Schaich and Münnich, 1993 and Díaz-García and Cortez, 2006). The optimal allocation then highly depends on the concrete choice of a scalarization function. Schaich and Münnich (1993) proposed p-norms $(p = 1, 2, 4, 8, \ldots)$ of the objectives as well as the limiting case, where the maximum of the objectives is minimized (*min-max*). In this study, we address both the p-norms and the min-max approach. As discussed in Sections 4.2 and 4.3, the weighted 1-norm (i.e. the weighted sum) is mainly focused on, since it coincides with the theory of Pareto optimality presented in Section 4.2.4 under some assumptions. We have suggested a similar approach for a simpler framework in Friedrich et al. (2018).

Due to the scalarization, we take the second of the two perspectives of Dalenius (1953) and treat the multivariate allocation problem as a multi-objective problem. We extend the theoretical results in Folks and Antle (1965) by giving a (necessary and sufficient) characterization of *all* Pareto optimal points in analogy to Subsection 3.3.5. Moreover, we compute the set of Pareto optimal solutions, also called Pareto frontier, for the refined problem formulation which allows decision-makers to choose an individually specified preference from this set. In order to support the decision-maker in the final choice of the preferred allocation, a strategy is suggested in Section 4.4 to determine one specific solution of the set of all Pareto optimal solutions.

With regard to the algorithmic strategy, solving allocation problems under the box-constraints (4.1) via a special and efficient Lagrangian approach may yield non-differentiable points (Münnich et al., 2012b and Wagner, 2013). Thus, many standard algorithms, such as classical Newton techniques, may fail to provide the correct optimal solution. To avoid convergence issues, we propose using the semismooth Newton method introduced in Section 3.2. Since stratum-specific sample sizes are integer values, we also refer to the integer optimal allocation techniques published in Friedrich et al. (2015), which avoids rounding.

Each scalarization for optimal multivariate allocation contains an additive linking of variances. Due to the scaling of units of the variables of interest, the variances have to be standardized for comparability. In Schaich and Münnich (1993), a standardization using the coefficient of variation (see Equation (2.13)) is proposed. Here, we present an alternative solution that extends this technique. Additionally, both standardization techniques are analyzed in great detail under various circumstances.

Finally, it is of great importance for practical applications that large problem instances can be solved in an appropriate time. Our methods solve optimal multivariate allocation problems with several thousand strata within seconds and are reliable tools when dealing with real-world data. This stands in contrast with other algorithms for multivariate allocation problems, which are generally computationally tractable for only a small number of strata. The computationally efficient solution of large optimal multivariate allocation problems supplements the theoretical discussion and is certainly another central innovation of our methods.

In Section 4.2, we concentrate on the mathematical foundation of the `MMDopt(.reg)` problem (4.5) with various decision-making functions and standardization techniques. Moreover, we establish the link between the multivariate allocation problem and the theory of Pareto optimization presented in Section 4.2.4. In this way, the theory of multi-criteria optimization of Section 3.3 is essential. Primarily, four variants of the box-constrained optimal multivariate and multi-domain allocation problem (4.5) are derived depending on the decision-making function and the restrictions:

1. `MMDopt` with weighted sum or 1-norm scalarization,
2. `MMDopt.reg` with weighted sum or 1-norm scalarization (restrictions for reg. efficiency),

3. `MMDopt` with p-norm scalarization for $p > 1$, and
4. `MMDopt.reg` with p-norm scalarization for $p > 1$ (restrictions for regional efficiency).

In Section 4.3, the first two variants are analyzed, and a solution strategy based on semismooth Newton methods (henceforth denoted with `SSN` and `SSN.reg` for the solution of `MMDopt` and `MMDopt.reg` respectively) is developed. In taking advantage of the special structure of the problems, the general KKT-system can be equivalently rewritten as a significantly lower dimensional nonlinear system of equations, which can efficiently be solved by the `SSN` method. As shown in Section 4.6, the algorithms are fast enough to solve even large problem instances in an appropriate computing time. The scalarization techniques applied in the third and fourth variants lead to problem formulations that prohibit the usage of a lower dimensional reformulation (the objective functions are not separable). Therefore, an alternative solution strategy for these variants based on a projected inexact quasi-subgradient method (`GTM`) is addressed in Section 4.4, which still allows the solution of these problems in an appropriate time.

In Section 4.6, the statistical accuracy, numerical efficiency, and practicability of the developed methods is discussed. This is carried out within the framework of an application study based on a partly-synthetic household dataset of Germany, as presented in Section 2.6. Advantages and opportunities as well as issues and limits of the developed methods are addressed.

4.2 The multivariate and multi-domain allocation problem

4.2.1 Mathematical problem formulation

To introduce different variants of the multivariate and multi-domain optimal allocation problem (4.5) and to develop numerical solvers of these variants, a mathematical analysis of their structure and properties has to be established. In accordance with the notations and definitions of Section 2.1, the population is given by $\mathcal{U} = \{1, \dots, N\}$. In analogy to the notation in Section 2.3, the stratified random sampling design is defined by H exhaustive and disjoint cross-classification strata $\mathcal{U}_h$ with $\mathcal{U} = \bigcup_{h=1}^{H} \mathcal{U}_h$ and stratum size N_h. The strata are constructed based on R stratification levels consisting of exhaustive and disjoint areas respectively. The L_r areas on stratification level $r \in \{1, \dots, R\}$ are denoted with $\mathcal{U}_{l_r}^{(r)}$ ($l_r \in \{1, \dots, L_r\}$) and the Equation (2.32) holds for all $r = 1, \dots, R$. In order to express the areas with the aid of the strata, we define the index set

$$\mathcal{H}_{l_r}^{(r)} := \left\{ h \in \{1, \dots, H\} : \mathcal{U}_h \subseteq \mathcal{U}_{l_r}^{(r)} \right\} \tag{4.6}$$

consisting of all indices $h = 1, \dots, H$, for which the strata $\mathcal{U}_h$ is a subset of the area $\mathcal{U}_{l_r}^{(r)}$. In this way, the area l_r on stratification level r can formally be defined as union of strata $h \in \mathcal{H}_{l_r}^{(r)}$, i.e.

$$\mathcal{U}_{l_r}^{(r)} := \bigcup_{h \in \mathcal{H}_{l_r}^{(r)}} \mathcal{U}_h. \tag{4.7}$$

Equation (2.32) implies that for each stratification level r, the population $\mathcal{U}$ is given by the union of all areas $\mathcal{U}_{l_r}^{(r)}$ of the respective level r. Expression (4.7) warrants the consistency between areas and strata, i.e. each area is a union of one or more strata. Due to this, a StrRS design can be observed in each area, such that the computations of variances and quality measures defined in Section 2.1 is valid. The stratification structure is exemplarily illustrated in Figure 2.1 for $R = 3$, $L_1 = 2$, $L_2 = 3$, $L_3 = 3$, and $H = 10$. Regarding the RIFOSS dataset (see Section 2.6), stratification levels could be federal states, NUTS2-regions, NUTS3-regions, SMPs (all define regional levels), and the classes of household size (level by content; see Figure 2.2). The strata then are constructed as cross-classifications of these areas.

In consistency with Equation (2.26), the stratum-specific variances are denoted with S_{ih}^2 for all variables y_i and all strata h. If no information is available to compute S_{ih}^2, it has to be estimated using adequate proxies obtained from auxiliary data or previous surveys. The robustness with regard to proxy quality is further discussed in the context of the simulation study in Subsection 4.6.5.

In accordance with Equations (2.36) and (2.50), the objective functions of problem (4.5) are denoted with

$$F : \mathbb{R}_+^H \to \mathbb{R}_{0+}^{q_1}, n \mapsto \Big(F_1(n), \ldots, F_{q_1}(n)\Big)^T \tag{4.8}$$

with strictly convex component functions

$$F_i(n) = \sum_{h=1}^{H} \left(\frac{d_{ih}}{n_h} - e_{ih} \right) = d_i^T \left(\frac{1}{n} \right) - \mathbb{1}^T e_i. \tag{4.9}$$

In that regard, we denote $\left(\frac{1}{n}\right) := \left(\frac{1}{n_1}, \ldots, \frac{1}{n_H}\right)^T \in \mathbb{R}_+^H$, and the components are given by

$$\begin{aligned} d_i &:= (d_{i1}, \ldots, d_{iH})^T\,,\ d_{ih} = S_{ih}^2 N_h^2 \ \ \forall i = 1, \ldots, q_1 \text{ and } h = 1, \ldots, H \ \text{ and} \\ e_i &:= (e_{i1}, \ldots, e_{iH})^T\,,\ e_{ih} = S_{ih}^2 N_h \ \ \forall i = 1, \ldots, q_1 \text{ and } h = 1, \ldots, H. \end{aligned} \tag{4.10}$$

Without loss of generality, we can assume $d_{ih} > 0$ (otherwise the respective addend can be omitted). The q_2 affine-linear equality constraints are expressed as

$$A := \begin{bmatrix} 1 & \vdots & & \vdots \\ \vdots & A_2 & \ldots & A_{q_2} \\ 1 & \vdots & & \vdots \end{bmatrix}^T \in \mathbb{R}^{q_2 \times H} \ \text{ and } \ b := (n_{\text{s}}, b_2, \ldots, b_{q_2})^T \in \mathbb{R}^{q_2}, \tag{4.11}$$

which ensures the compliance with the total sample size n_{s}. By indicating the inequality constraints, we assume that restrictions for regional efficiency, i.e. bounds for the variances, are desired for a subset of all area on all stratification levels. Let

$$\mathcal{H}_{\text{restr}} \subseteq \Big\{(i, r, l_r) : i \in \{1, \ldots, q_1\},\ l_r \in \{1, \ldots, L_r\},\ r \in \{1, \ldots, R\}\Big\} \tag{4.12}$$

contain these triples (i, r, l_r) for which there exists a maximal variance $\text{Vmax}_{(i,r,l_r)} \in \mathbb{R}_+$ concerning variable y_i for area l_r on stratification level r. In particular, the triple $(2, 1, 3) \in \mathcal{H}_{\text{restr}}$ corresponds to a quality restriction of variable of interest y_2 in area 3 on stratification level 1.

For the sake of notational simplicity, the q_3 elements of $\mathcal{H}_{\text{restr}}$ are consecutively numbered from 1 to q_3, i.e. $(i, r, l_r)_1$ to $(i, r, l_r)_{q_3}$. Formally, the restriction for regional efficiency are then denoted with

$$\text{Var}(\hat{\tau}_{y_i l_r}^{\text{StrRS}}) \leq \text{Vmax}_{(i,r,l_r)_j} \tag{4.13}$$

for all $(i, r, l_r)_j \in \mathcal{H}_{\text{restr}}$. In that regard, $\text{Var}(\hat{\tau}_{y_i l_r}^{\text{StrRS}})$ denote the area-specific variances of the HT estimator for variable y_i (see Equation 2.36), i.e.

$$\text{Var}(\hat{\tau}_{y_i l_r}^{\text{StrRS}}) = \sum_{h \in \mathcal{H}_{l_r}^{(r)}} \frac{S_{ih}^2 N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) = \sum_{h \in \mathcal{H}_{l_r}^{(r)}} \frac{d_{ih}}{n_h} - \sum_{h \in \mathcal{H}_{l_r}^{(r)}} \left(S_{ih}^2 N_h\right). \tag{4.14}$$

The value $\text{Vmax}_{(i,r,l_r)_j}$ is the maximal value for restriction $(i, r, l_r)_j$, i.e. the minimal quality which has to be achieved for the HT estimate of variables y_i in area l_r on stratification level r. To express (4.13) in matrix notation, a matrix $D \in \mathbb{R}^{q_3 \times H}$ and a vector $c \in \mathbb{R}^{q_3}$ are defined as

$$D := \begin{bmatrix} d_{i1} \cdot \mathbb{1}_{\left((i,r,l_r)_1 \in \mathcal{H}_{\text{restr}} \,\wedge\, 1 \in \mathcal{H}_{l_r}^{(r)}\right)} & \dots & d_{iH} \cdot \mathbb{1}_{\left((i,r,l_r)_1 \in \mathcal{H}_{\text{restr}} \,\wedge\, H \in \mathcal{H}_{l_r}^{(r)}\right)} \\ \vdots & & \vdots \\ d_{i1} \cdot \mathbb{1}_{\left((i,r,l_r)_{q_3} \in \mathcal{H}_{\text{restr}} \,\wedge\, 1 \in \mathcal{H}_{l_r}^{(r)}\right)} & \dots & d_{iH} \cdot \mathbb{1}_{\left((i,r,l_r)_{q_3} \in \mathcal{H}_{\text{restr}} \,\wedge\, H \in \mathcal{H}_{l_r}^{(r)}\right)} \end{bmatrix} \tag{4.15}$$

with columns $D_1, \dots, D_H$ and

$$c = \begin{pmatrix} c_1 \\ \vdots \\ c_{q_3} \end{pmatrix} := \begin{pmatrix} \text{Vmax}_{(i,r,l_r)_1} + \sum_{h \in \mathcal{H}_{l_r}^{(r)}} (S_{ih}^2 N_h) \\ \vdots \\ \text{Vmax}_{(i,r,l_r)_{q_3}} + \sum_{h \in \mathcal{H}_{l_r}^{(r)}} (S_{ih}^2 N_h) \end{pmatrix}. \tag{4.16}$$

Each row in D corresponds to one restriction for regional efficiency. In this way, those of the H components of the row are non-zero, which correspond to strata that are a subset of the respective area. Using (4.15) and (4.16), an equivalent formulation of (4.13) for all $j = 1, \dots, q_3$ is given by

$$D\left(\frac{1}{n}\right) \leq c. \tag{4.17}$$

The following example illustrates the structure of the nonlinear inequality constraints.

Example 4.2.1. Let the example presented in Figure 2.1 be given and assume that restrictions for regional efficiency are given for variable y_1 and on stratification level $r = 2$ (3 areas) as well as for variable y_2 and on stratification level $r = 1$ (2 areas). Then, the structure of matrix $D \in \mathbb{R}^{5 \times 10}$ in (4.15) is given by

$$D = \left[\begin{array}{cccccccccc} d_{21} & d_{22} & & & & d_{26} & & & & \\ & & d_{23} & d_{24} & & & d_{27} & d_{28} & & \\ & & & & d_{25} & & & & d_{29} & d_{2\,10} \\ \hline d_{11} & d_{12} & d_{13} & d_{14} & d_{15} & & & & & \\ & & & & & d_{16} & d_{17} & d_{18} & d_{19} & d_{1\,10} \end{array}\right],$$

where each row corresponds to one restriction for regional efficiency, and each column represents one cross-classification stratum.

Quality restrictions required for cross-classification strata can be controlled by adjusting the predefined lower box-constraints $m \in \mathbb{R}^H_+$ for the stratum-specific sample sizes $n \in \mathbb{R}^H_+$. This is possible in contrast to the areas, since each stratum contains a SRS design, such that the variance formula $\text{Var}(\hat{\tau}^{\text{SRS}}_{y_i h})$ defined in (2.25) can be applied. We assume that the maximal variances $\widetilde{\text{Vmax}}_{i1}, \dots, \widetilde{\text{Vmax}}_{iH} \in \mathbb{R}_+$ for stratum-specific total estimates for variable y_i are given. If no quality restriction is required for a stratum h, we set $\widetilde{\text{Vmax}}_{ih} \to \infty$ for all $i = 1, \dots, q_1$. Formally, the stratum-specific quality restrictions can be expressed by

$$\text{Var}(\hat{\tau}^{\text{SRS}}_{y_i h}) = \frac{S^2_{ih} N^2_h}{n_h} \left(1 - \frac{n_h}{N_h}\right) \leq \widetilde{\text{Vmax}}_{ih} \tag{4.18}$$

for stratum h and variable y_1, and it can be equivalently rewritten as

$$\frac{S^2_{ih} N^2_h}{\widetilde{\text{Vmax}}_{ih} + S^2_{ih} N_h} \leq n_h.$$

Then, the adjusted lower box-constraints m_h can be computed as

$$m_h \leftarrow \max \left(m_h \, , \max_{i=1,\dots,q_1} \left(\frac{S^2_{ih} N^2_h}{\widetilde{\text{Vmax}}_{ih} + S^2_{ih} N_h} \right) \right) \; \forall h = 1, \dots, H. \tag{4.19}$$

After summarizing expressions (4.8) to (4.19), problem (4.5) is equivalent to

$$\begin{aligned} \min_{n \in \mathbb{R}^H_+} \; & \big(F_1(n), \dots, F_{q_1}(n)\big) \\ \text{s.t.} \; & A\, n - b = 0 \\ & D\, n^{-1} - c \leq 0 \\ & m \leq n \leq M. \end{aligned} \tag{4.20}$$

The objective functions defined in (4.9) can be simplified to

$$F_i(n) = \sum_{h=1}^{H} \left(\frac{d_{ih}}{n_h} \right), \tag{4.21}$$

since the constant second term $\mathbb{1}^T e_i$ in (4.9) does not affect the optimization. As $d_{ih} > 0$ and $n_\text{s} < N$ (otherwise we have a full census), we can assume $F_i(n)$ to be strictly positive for all $i = 1, \dots, q_1$ and $n \in \mathcal{X}$. The feasible set of (4.20) is given by

$$\mathcal{X} = \left\{ n \in \mathbb{R}^H_+ : An = b, \; Dn^{-1} \leq c, \; n \geq m, \; n \leq M \right\}. \tag{4.22}$$

The feasible set $\mathcal{X}$ is convex, since all equality constraints $(An - b)$ are affine-linear and the inequality constraints $(Dn^{-1} - c)$ are convex as sum of convex functions. Moreover, the box-constraints can also be interpreted as constant (and therefore convex) inequality constraints. Then, the convexity of $\mathcal{X}$ is given by Remark 3.1.5. Since the objective functions of problem (4.20) defined in (4.21) are strictly convex (as a sum over strict convex functions), the problem can be called a convex multi-objective optimization problem. Moreover, the existence of a solution of problem (4.20) is given by Theorem 3.3.11, since the feasible set $\mathcal{X}$ is closed and bounded. Necessary and sufficient optimality conditions are given by Theorems 3.3.12 and 3.3.13.

Remark 4.2.2. In practice, the restrictions for regional efficiency (4.17) for area-specific estimates are rarely indicated with maximal variances. Instead, they are indicated by relative values such as the relative variances, relative standard deviations, or coefficients of variation. Consequently, the values $\mathrm{Vmax}_{(i,r,l_r)_j}$ for the areas given in (4.16) and $\widetilde{\mathrm{Vmax}}_{ih}$ for the strata applied in (4.18) have to be adjusted according to the definition of the respective measure. The structures of the constraints remains unchanged.

To complete this subsection, it has to be emphasized that the additional linear equality constraints and nonlinear inequality constraints can be defined individually by the user. Provided that the feasible set of the programming problem is not empty, it is generally possible to restrict the quality of an estimate for each combination of variables of interest and combination of strata.

4.2.2 Standardization

In an optimal multivariate allocation problem of the form of Equation (4.20), several variables of interest are considered simultaneously. Thus, the resulting optimization problem has several possibly conflicting objective functions. In that regard, the correlation between the variables of interest, the variable types, as well as the purpose of the survey are decisive factors. Since the variables of interest may be of different types and scales, the variances of the respective estimators are likewise of different scales. Thus, the objective functions, which contain the variance functions, have to be standardized in order to make the objectives comparable and linkable. The relevance of standardization is accentuated by Table 4.1, which lists quantiles of three selected variables on household level. In comparing quantiles of the variables among each other, it can be noted that the scale is completely different. For instance, AGE4.1 contains values with a maximum of 705, and PEN includes values of up to 1.9 million. Thus, an additional linking of these variables without standardization would be conducive. For more information on the dataset, we refer to Section 2.6.

Table 4.1: Quantiles of household-level values and variances of the variables EDI, PEN, and AGE4.1.

Name	Quantiles							Variance
		50%	60%	70%	80%	90%	100%	
EDI	...	1 384	1 892	1 460	3 220	4 673	82 987	$25 \cdot 10^7$
PEN	...	0	0	582	1 187	2 189	1 921 720	$53 \cdot 10^5$
AGE4.1	...	0	0	0	1	2	705	3.75

Regarding the objective functions (4.8), Schaich and Münnich (1993) suggested to replace the variance of the estimators by the coefficient of variation to receive additively comparable values. In order to retain the mathematical properties of the variance function, we use the squared coefficient of variation

$$\mathrm{cv}^2(\hat{\tau}_{y_i}^{\mathrm{StrRS}}) := \frac{\mathrm{Var}(\hat{\tau}_{y_i}^{\mathrm{StrRS}})}{\tau_{y_i}^2} \tag{4.23}$$

with the population total τ_{y_i} of variable y_i according to (2.13). This standardization is referred to as *(cv)-standardization*. In that regard, the objective functions (4.21) are weighted with the standardization factors $\frac{1}{\tau_{y_i}^2}$. In contrast to the coefficient of variation (without squaring), the (cv)-standardization maintains the properties of the original objective functions, as they are only adjusted by scale. The principal effect of the simple and the squared coefficient of variation is similar. However, the squaring may lead to small differences in some settings. The coefficient of variation is a suitable instrument for the standardization, since it enables the comparison of the variances of population total estimates for various types of variables in contrast to the unstandardized variances (cf. Schaich and Münnich, 1993). Formally, it is a relative term in contrast to the absolute variance. A drawback in using the squared coefficient of variation is the requirement for the population totals τ_{y_i} of variables y_i, which are generally not given in advance as they are the objectives of the survey. Thus, adequate proxies need to be utilized. As (4.23) are ratios, their estimations are even more demanding than when only using the proxies for the stratum-specific variances contained in $\text{Var}(\hat{\tau}_{y_i}^{\text{StrRS}})$. Moreover, whenever the total of a variable is close to zero or if a minimum of a variable the lower than zero, the information content of the coefficient of variation is distorted (see Table 4.2).

Beyond the (cv)-standardization, an alternative standardization technique is proposed, which we refer to as *(opt)-standardization*. Instead of total values, the (opt)-standardization is based on variances as standardization factors. A multivariate allocation of conflicting variables of interest is also accompanied by a quest for a compromise. In this way, the amount of costs to be paid for each separate variable, i.e. the accuracy-decrease to be tolerated, in order to achieve an optimal multivariate allocation has to be determined. The accuracy-decrease of each variable is measured compared to the unique optimal univariate allocations, which are theoretically best for each separated variable, since the respective variances are minimized separately. Therefore, the variances of the total estimators under the optimal univariate allocations are used as standardization factors. Thus, the (opt)-standardization for the variable of interest y_i is given by

$$\text{opt}(\hat{\tau}_{y_i}^{\text{StrRS}}) := \frac{\text{Var}(\hat{\tau}_{y_i}^{\text{StrRS}})}{\text{Var}_{y_i}^{\text{opt}}}, \tag{4.24}$$

where $\text{Var}_{y_i}^{\text{opt}}$ is the variance of the total estimator for variable y_i under the optimal univariate allocation, which may be computed with the box-constraint optimal allocation by Münnich et al. (2012c) given in (2.49). This standardization technique reflects the relative loss for each variable under the consideration when using the compromise allocation rather than the single variable optimized allocation. In contrast to (cv), an advantage of this technique is that the total τ_{y_i} of variable y_i is not required. Moreover, if the stratum-specific variances S_{ih}^2 need to be estimated, the uncertainty and blur of this estimation is symmetrically present in the numerator and denominator of the objectives, and thus it may be eliminated. Hence, a standardization by the optimal univariate variances may result in a more robust optimal multivariate allocation. This is analyzed in the application study in Subsection 4.6.2. Moreover, the (opt)-standardization is also consistent, if variables comprise negative values, which is not the case for the (cv)-standardization.

Both standardization techniques are compared in Table 4.2, where the values of the objective functions (4.9) are tabulated for a proportional allocation. The variances of the variables are presented in the second column. If the objective functions F_i are computed by (4.9) using the unstandardized variances, their scale is not additively linkable (see column 3). If one of the standardization techniques is applied (column 4 and 5), the values of the three variables are comparable among each other. In particular, variable PEN has the highest coefficient of variation (column 4), whereas the ratio of the actual variance and the variance under optimal univariate allocation is the highest for variable AGE4.1. These differences clearly affects the optimal multivariate allocation concerning the three variables. We have a closer look on the differences depending on the standardization technique in the application results in Section 4.6.

Table 4.2: Value of objective functions for standardization techniques (under proportional allocations).

name	S^2	F_i (original)	F_i (cv)-standardization	F_i (opt)-standardization
EDI	$25 \cdot 10^7$	$2.84 \cdot 10^{16}$	$3.69 \cdot 10^{-5}$	1.02
PEN	$53 \cdot 10^5$	$4.71 \cdot 10^{15}$	$6.21 \cdot 10^{-5}$	1.24
AGE4.1	3.75	$1.27 \cdot 10^9$	$3.97 \cdot 10^{-5}$	3.47

In the following, we assume the given standardization factors $\gamma_1, \ldots, \gamma_{q_1}$ for all objective functions calculated using the (cv)- or (opt)-standardization. Due to notational simplicity, we set

$$F_i(n) \leftarrow \gamma_i F_i(n) \ \forall i = 1, \ldots, q_1.$$

Thus, the functions $F_i(n)$ refer to the standardized variances $\gamma_i \text{Var}(\hat{\tau}_{y_i}^{\text{StrRS}})$ hereafter.

4.2.3 Scalarization

As already mentioned in Section 3.3, the scalarization of the vector-valued objective function of problem (4.20) is mandatory for solving the multi-criteria optimization problem and depends on the characterization of optimality discussed in Section 3.3.1. The choice of a scalarization technique is not clear in advance and coincides with the preferences of the decision-maker. Thus, the scalarization method can also be interpreted as a decision-making function. As we will show in Section 4.6.3, the choice of a decision-making function has a considerable influence on the solution of the problem. A wide range of scalarization techniques can be found in literature. For the mathematical foundation of these techniques, we refer to Jahn (1986, Chapter 5), Ehrgott (2005, Chapter 3 and 4) and Lin (2005). In the following, we make a distinction between two types of techniques that resemble the two strategies of Dalenius (1953) already mentioned in Section 4.1. In the first strategy, some of the q_1 objective functions of problem (4.20) are treated as inequality restrictions to attain a real-valued optimization problem. Consequently, only one objective function is optimized, whereas the others remain bounded. One widespread technique belonging to this strategy is the epsilon-constraint method (cf. Ehrgott, 2005, Section 4.1)

$$\begin{aligned} &\min_{n \in \mathcal{X}} F_i(n) \\ &\text{s.t. } F_j(n) \leq \text{box}_j \ \forall j \in \{1, \ldots, q_1\} \setminus \{i\} \end{aligned} \tag{4.25}$$

with a fixed $i \in \{1, \ldots, q_1\}$ and scalar values $\text{box}_j \in \mathbb{R}_+$. As the decision-maker is tasked with selecting the objectives to be considered as restrictions, the technique would require an a priori ranking of the objectives. In the context of optimal multivariate allocation, a similar version of the epsilon-constraint method is applied, in which the allocation corresponds to a cost minimization while variance restrictions are respected, which is treated in Falorsi and Righi (2015). In this way, all objective functions are shifted to the constraints, whereas the costs (e.g. expressed as the sum over the stratum-specific sample sizes) are minimized. These strategies will not be focused on in this thesis since they contradict with the assumption of a simultaneous optimization of several objectives.

By contrast, the second strategy of Dalenius (1953) is based on an additive linkage of the q_1 objectives of (4.20) to achieve a real-valued objective function $f : \mathbb{R}^H_+ \to \mathbb{R}_{0+}$. Therefore, the strategy prevents an a priori ranking or selection of the objectives. First, p-norms of the objectives ($p \in \mathbb{N}$) are proposed leading to a real-valued objective function

$$f(n) := \left\| \Big(F_1(n), \ldots, F_{q_1}(n)\Big)^T \right\|_p = \left(\sum_{i=1}^{q_1} \Big(F_i(n)\Big)^p \right)^{\frac{1}{p}} \tag{4.26}$$

with $f : \mathbb{R}^H_+ \to \mathbb{R}_+$. Due to the definition of the objectives in (4.8), $F_i \geq 0$ holds for all i, such that the absolute value $| \cdot |^p$ can be omitted. This strategy is among others discussed in Schaich and Münnich (1993) and complies with the characterization of optimality in multi-criteria oprimization of class (norm) presented in Subsection 3.3.1. The bigger p is, the more attention is given to higher values of F_i. Thus, if p is high, the allocation focusses more on variables with larger standardized variances. Schaich and Münnich (1993) and Lin (2005) studied the particular case of $p = \infty$, which is equivalent to the *min-max* method (see class (max) presented in Subsection 3.3.1). Then the objective function is given by

$$f(n) := \max_{i=1,\ldots,q_1} F_i(n) \tag{4.27}$$

with $f : \mathbb{R}^H_+ \to \mathbb{R}_+$. In that regard, the allocation concentrates on the minimization of the variable with the largest standardized variance. In case of the (cv)-standardization, the maximal squared coefficient of variation of all variables is minimized. This leads to an allocation, where the coefficients of variation are well-compensated between the variables of interest. By contrast, the (opt)-standardization should be chosen if a well-balanced accuracy-decrease compared to the optimal univariate allocations is required. In practice, the min-max method is a very popular method, as there is no risk to neglect the variables which are more critical for the allocation.

Besides the p-norm and min-max method, the most intuitive and common scalarization technique is the *weighted sum method*, with which each objective is weighted and the weighted objectives are cumulated (Jahn, 1986, Chapter 3). The vector-valued objective function of problem (4.20) then becomes a real-valued objective $f : \mathbb{R}^H_+ \to \mathbb{R}_+$ of the form

$$f(n) := \sum_{i=1}^{q_1} w_i F_i(n) \tag{4.28}$$

with non-negative weights $w_1, \ldots w_{q_1} \geq 0$ and $\sum_{i=1}^{q_1} w_i = 1$. Using equal weights $w_i = \frac{1}{q_1}$ for all variables, the minimization of the weighted sum (4.28) is equivalent to the minimization of the 1-norm in (4.26). Generally, the weighted sum method depends on an a priori choice of the weighting scheme, i.e. the decision-maker needs to rank the variables of interest similarly to the epsilon-constraint method (4.25). To avoid this, we focus on solving the weighted sum problem with all possible combinations of weights, subject to a discretization accuracy of the weights. In Subsection 3.3.1, the weighted sum approach is described with the optimization class (EF). In accordance with this characterization, and in analogy to Subsection 4.2.4, the discretization and the special structure of the problem allows us to compute the whole Pareto frontier, i.e. to find *all* Pareto optimal solutions. Thus, the decision-makers are able to choose their own preferred allocation *a posteriori* instead of doing an inevitable a priori weighting or ranking of the variables of interest. In conclusion, the decision is based on more reliable information which allows for more flexibility.

Apart from the statistical properties and advantages of the developed allocation tool, an efficient and robust numerical solution of the scalarized and standardized version of problem (4.20) is a vital part of this thesis. For the choice of a numerical solver, the properties of the objective function are a decisive factor. In the case of the weighted sum scalarization (4.28), the objective function f is continuously differentiable, strictly convex, and separable, since f is a (weighted) sum over continuously differentiable, strictly convex, and separable functions F_i defined in (4.21). These properties are essential for the fast algorithms presented in Section 4.3. If the alternative scalarization methods are used, f changes and may lose some of these properties. In particular, the objective function f in (4.26) is continuously differentiable and strictly convex, but it is only separable if $p = 1$. If f is not separable as in the case $p \neq 1$, special attention must be paid to the selection of the solution algorithm. This problem is tackled in Section 4.4. Furthermore, for the min-max method, the objective (4.27) is not continuously differentiable. However, many classical optimization methods (such as the Newton method) rely on differentiability and are not applicable in this case. An alternative solution for this case is derived in Section 4.4, where the non-separable case $p \neq 1$ is traced back to the separable case of the weighted sum scalarization ($p = 1$).

4.2.4 Weighted sum and Pareto optimization

As already analyzed in Subsection 3.3.5, the scalarization by the weighted sum coincides with the theory of Pareto optimality introduced in Subsection 3.3.2. When optimizing the conflicting objectives in (4.20), the *Pareto frontier* describes the set of all efficient solutions (cf. Definition 3.3.8). Efficient solutions are characterized by all points for which one objective function F_i can only be improved by diminishing another. Therefore, the Pareto frontier provides a suitable characterization of all the points that decision-makers should consider in a multi-criteria optimization problem. However, it is not advisable to choose an allocation that is not an element of the Pareto frontier, since in this case at least one objective function F_i can be improved without causing any further costs (i.e. without diminishing another objective function).

Moreover, the Pareto frontier describes the optimal solutions independent of the weighting, i.e. independent of the ranking of the variables of interest. Instead of determining the ranking in advance, the developed method allows users to select a preferred solution among all Pareto optimal points *after* the optimization step. Advantages of this procedure include the ability to optimize without a known priority ranking of the variables of interest, the robustness of the solution with respect to the weights, and the possibility to use additional information at the time of decision, such as variance structures or sensitivity analyses.

The entire frontier of Pareto optimal solutions for the multivariate allocation problem (4.20) is described mathematically in Subsection 3.3.5 and extend by the results of Folks and Antle (1965). By Theorem 3.3.15, it can be proved that each optimal solution of the weighted sum reformulation for an arbitrary choice of weights is a Pareto optimal solution for (4.20). Moreover, since F_i are (strictly) convex functions and the feasible set $\mathcal{X}$ defined in (4.22) is a convex set, the following statement holds by applying Lemma 3.3.16 and Theorem 3.3.17. If the weighted sum problem is solved for all possible combinations of weights, *all* Pareto optimal solutions of the original problem are obtained (only subject to the discretization of the weights). In this way, the whole Pareto frontier of the multivariate allocation problem (4.20) is computed. We refer to Section 4.6 for numerical results.

4.2.5 Generalizations of the allocation method

Compensated optimal allocation

Generally, optimal allocation techniques are based on the minimization of the (standardized) variances of the HT estimates for the *population* total of the variables of interest. In this way, although a good quality for some *area-specific* estimates might be of interest as well, the accuracy of these estimates is basically neglected. This effect is also observed in the application study in Subsection 4.6.4, where high estimation errors are observed with regard to a few particularly small areas and strata. Using `MMDopt`, the restrictions for regional efficiency can be added as inequality restrictions to the optimization problem in (4.20). Another approach was implemented in the German Census 2011 for the box-constrained optimal univariate allocation by Gabler et al. (2012) and Münnich et al. (2012c). In this case, the population was divided into regional sampling points and classes of address sizes (Münnich et al., 2012a, pp. 22 ff.). Thus, the stratified sampling design was defined by cross-classification strata composed of the sampling points and classes of address sizes. To achieve a *compensated* accuracy in each sampling point, the 2-norm of the *weighted* RRMSEs of sampling point-specific estimates was minimized instead of the variance of the population total estimate given by (2.50) (cf. Münnich et al., 2012a, pp. 31 ff.). This results in the following compensated objective function for the box-constrained optimal univariate allocation for $j = 1, \ldots, J$ classes of address sizes, $h = 1, \ldots, H$ sampling points and sampling point-specific weighting coefficients v_h:

$$\begin{aligned}\left\|\text{RRMSE}(\hat{\tau}_y^{\text{StrRS}})\right\|_2 := &\sum_{h=1}^{H} v_h \text{RRMSE}(\hat{\tau}_{yh}^{\text{StrRS}})^2 \\ = &\sum_{h=1}^{H} v_h \frac{\text{Var}(\hat{\tau}_{yh}^{\text{StrRS}})}{\tau_{yh}^2} \\ = &\sum_{h=1}^{H} \frac{v_h}{\tau_{yh}^2} \sum_{j=1}^{J} \frac{N_{hj}^2 S_{hj}^2}{n_{hj}} \left(1 - \frac{n_{hj}}{N_{hj}}\right).\end{aligned} \tag{4.29}$$

In the German Census 2011, the weighting coefficients v_h are neglected. If $v_h = \tau_{yh}^2$, then (4.29) is equal to the standard expression of the variance in (2.50). Other values for v_h are possible, such as the results computed by an a priori given function of a minimal allowed precision. This alternative approach may lead to a more compensated regional-specific accuracy and is includable to `MMDopt` with little effort. Nevertheless, this approach is omitted in the application study due to two reasons. At first, several test simulations yielded more compensated and more accurate estimates for the sampling points, while the efficiency of the estimates on other stratification levels significantly decreased. `MMDopt` is developed to generate accurate estimates on several stratification levels simultaneously. To achieve this, the restrictions for regional efficiency tend to be the more suitable instrument, since unusable inaccurate outliers for the area-specific estimates are omitted, but the variances of the population totals are still minimized. Furthermore, the focus of `MMDopt` is on the several stratification levels and, in particular on the optimal multivariate allocation. Thus, the compensated allocation would generate undesirable side effects. Nevertheless, using (4.29) is possible in the `MMDopt` framework, and it may be useful in some applications.

Optimal allocation with GREG-type objective function

Optimal allocations such as the Neyman-allocation (Neyman, 1934), the box-constrained optimal univariate allocation (Münnich et al., 2012c), and our multivariate extensions `MMDopt` are based on the minimization of the sum of the variances of the stratum-specific HT estimates $\mathrm{Var}(\hat{\tau}_{yh}^{\mathrm{SRS.HT}})$ of all strata h given by (2.25). To achieve efficient estimates with the aid of known auxiliary data, the GREG estimator (2.19) is a very popular instrument. As discussed in Lohr (2009, pp. 74 ff.), the variance of the GREG estimator for StrRS can be approximated by

$$\mathrm{Var}(\hat{\tau}_{y_i}^{\mathrm{StrRS,GREG}}) \approx \sum_{h=1}^{H} \mathrm{Var}(\hat{\tau}_{y_ih}^{\mathrm{SRS.HT}}) \cdot \left(1 - \vartheta_h^2\right), \tag{4.30}$$

where ϑ_{ih} is the stratum-specific coefficient of correlation between the variable of interest y_i and the auxiliary variable. Equation (4.30) is only valid, if ϑ_{ih} is computed separately for each stratum and SRS is applied in each stratum. Otherwise, a more general definition of ϑ has to be used (cf. Krug et al., 2001, Equation (7.1.39)). However, the expression in (4.30) is only valid for *one* auxiliary variable in stratum h and the contained approximation may be inaccurate in specific situations, for instance in small strata. Alternatively, the variance formula for the GREG estimator under StrRS as presented in Section 2.3 can be used, i.e.

$$\mathrm{Var}(\hat{\tau}_{y_i}^{\mathrm{StrRS,GREG}}) = \sum_{h=1}^{H} \frac{S_{\mathrm{e}_{ih}}^2 N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right), \tag{4.31}$$

where $S_{\mathrm{e}_{ih}}^2$ is computed as a proxy to (2.29) for the variables of interest y_i $(i = 1, \dots, q_1)$. This is valid for more than one auxiliary variable. Thus, an allocation tailored to the GREG estimator is also covered by the developed `MMDopt` method. Although it is theoretically possible, the minimization of a variance function of the GREG is unusual in optimal allocation, since it requires the availability of auxiliary data on individual level at the design stage in order to compute proxies for $S_{\mathrm{e}_{y_ih}}^2$. Due to this and in order to avoid side effects, we limit ourselves to the consideration of the variances of the HT estimator in the application study.

Additional cost functions

Up until this point, we have solely considered the equality constraint $\sum_{h=1}^{H} n_h = n_{\text{s}}$ corresponding to an a priori fixed total sample size n_{s}. Nevertheless, our approach behaves similarly when using different types of restrictions. Cost restrictions are frequently added, as described in Falorsi and Righi (2016, Example 1). Instead of n_{s}, we assume an a priori given total cost limit of $c_{\text{full}} \in \mathbb{R}_{0+}$. Let the fixed costs per stratum be defined by $c_{h_{\text{fix}}} \in \mathbb{R}_{0+}$ and the cost per unit sampled in stratum h by $c_{h_{\text{var}}} \in \mathbb{R}_{0+}$. Then, the restriction is given by

$$\sum_{h=1}^{H} \left(c_{h_{\text{fix}}} + n_h c_{h_{\text{var}}} \right) = c_{\text{full}}, \tag{4.32}$$

which is equivalent to

$$\sum_{h=1}^{H} n_h c_{h_{\text{var}}} = c_{\text{full}} - \sum_{h=1}^{H} c_{h_{\text{fix}}}. \tag{4.33}$$

Frequently, the fixed costs are not defined per stratum, i.e. $\sum_{h=1}^{H} c_{h_{\text{fix}}}$ may be replaced by a value $c_{\text{fix}} \in \mathbb{R}_{0+}$ representing the total fixed costs of the survey. With $n_{\text{s}} := c_{\text{full}} - \sum_{h=1}^{H} c_{h_{\text{fix}}}$ and an appropriate redefinition of A in (4.4), we prove the compatibility of our approach under consideration of generalized cost restrictions of the form defined in (4.32).

Various levels of sampling units

Besides the cost restrictions, the fact that the restrictions refer to the sampling units of the population is a fundamental assumption in the previous derivations. In relation to this point, a contradictory example is the German Census 2011 (cf. Münnich et al., 2012a, pp. 31 ff.), where the addresses of Germany are the sampling units, while $n_{\text{s,pers}}$ refers to the maximal number of persons that can be drawn. Then, $\sum_{h=1}^{H} n_h = n_{\text{s}}$ can be rewritten as

$$\sum_{h=1}^{H} n_h \cdot \text{pPA}_h = n_{\text{s,pers}}, \tag{4.34}$$

where $\text{pPA}_h \in \mathbb{R}_{0+}$ is the average number of persons living in one address in stratum h. Since pPA_h is an average number, only the expected number of persons within the sample will be equal to $n_{\text{s,pers}}$. The precision of the expected value will be high if the structure of the strata is associated with the number of persons living in one address. For example, in the German Census 2011, the strata are built according to different classes of address sizes. Moreover, a reformulation of the box-constraints of problem (4.5) can be handled in analogy to the previous reformulation in (4.34), i.e.

$$m_{\text{pers}_h} \leq n_h \cdot \text{pPA}_h \leq M_{\text{pers}_h} \; (h = 1, \ldots, H), \tag{4.35}$$

where m_{pers_h} and M_{pers_h} are the lower and upper bounds for the number of persons to be drawn in the strata $h = 1, \ldots H$ respectively. In some surveys, sampling fractions are used instead of absolute numbers n_{s}, m_h, and M_h. These can be trivially included and transformed to the standard form by a multiplication with N and N_h respectively.

4.3 Algorithmic solution of weighted sum allocation problems

In this section, two efficient algorithms for the numerical (continuous) solution of the weighted sum scalarized `MMDopt` problem (4.20) with and without additional restrictions for regional efficiency are presented, namely algorithms `SSN` and `SSN.reg`. Both algorithms are based on a Lagrangian approach developed in Münnich et al. (2012c), where an optimal univariate allocation problem with box-constraints is considered. The approaches are accompanied by a significant reduction of the dimension of the underlying problem enabled by the special structure of the problem. After a scalarization with the weighted sum and standardization with the techniques described in Section 4.2.2, the strategy applied in the univariate case is also applicable to the multivariate problem. The main characteristic of the algorithm is the possibility to express the stratum-specific sample sizes n_h as a function depending on the Lagrange multipliers by transforming the KKT optimality conditions. The strict convexity and separability of the scalarized objective function f in (4.28) is crucial for the correctness and applicability of both algorithms.

For the sake of generality, the following derivation is conducted to solve the weighted sum scalarized problem (4.5), i.e. the optimal multivariate allocation problem with q_1 variables of interest, q_2 affine-linear equality constraints, and q_3 convex inequality constraints. Nevertheless, it is also valid for the common practical case $q_2 = 1$, where $\sum_{h=1}^{H} n_h = n_\text{s}$ is the only equality constraint, and $q_3 = 0$, where there are no restrictions for regional efficiency.

For an integer solver for the case without inequality constraints ($q_3 = 0$) and only one equality constraint ($q_2 = 1$), we refer to a Greedy-based solution strategy developed in Friedrich et al. (2015) and Friedrich (2016). A further comparison of continuous and integer solvers can be found in Friedrich et al. (2018).

4.3.1 Existence and uniqueness of the solution

In applying the weighted sum scalarization method (4.28), the vector-valued objective function F of problem (4.20) changes in accordance with (4.21) to

$$f(n) = \sum_{i=1}^{q_1} w_i F_i(n) = \sum_{i=1}^{q_1} w_i \left(\sum_{h=1}^{H} \frac{d_{ih}}{n_h} \right), \tag{4.36}$$

which can be simplified to

$$f(n) = \left(\sum_{i=1}^{q_1} w_i \sum_{h=1}^{H} \frac{d_{ih}}{n_h} \right) = \left(\sum_{h=1}^{H} \frac{\sum_{i=1}^{q_1} w_i d_{ih}}{n_h} \right).$$

Using the function $\varphi : \mathbb{R}_+^H \to \mathbb{R}_+^H,\ n \mapsto n^{-1}$, the objective function can be written as

$$f(n) = \sum_{h=1}^{H} \frac{d_h(w)}{n_h} = d(w)^T \varphi(n) \tag{4.37}$$

with functions

$$d_h : \mathbb{R}^{q_1} \to \mathbb{R}_+ \ , \ w \mapsto \sum_{i=1}^{q_1} w_i d_{ih} \ \ (h = 1, \ldots, H) \tag{4.38}$$

and $d(w) = \Big(d_1(w), \ldots, d_H(w)\Big)^T$. Then, the weighted sum scalarized problem (4.20) can be rewritten as

$$\begin{aligned} &\min_{n \in \mathbb{R}_+^H} f(n) \\ &\quad \text{s.t. } h(n) = 0 \ \ , \ g(n) \leq 0 \ \ , \ m \leq n \leq M. \end{aligned} \tag{4.39}$$

with functions

$$\begin{aligned} &f : \mathbb{R}_+^H \to \mathbb{R}_+, \ f(n) = \sum_{h=1}^{H} \frac{d_h(w)}{n_h} = d(w)^T \varphi(n), \\ &h : \mathbb{R}_+^H \to \mathbb{R}^{q_2}, \ h(n) = An - b, \ \text{and} \\ &g : \mathbb{R}_+^H \to \mathbb{R}^{q_3}, \ g(n) = Dn^{-1} - c = D\varphi(n) - c. \end{aligned}$$

In that regard, the functions d_h is dependent on the weights w of the weighted sum and are component-wise defined by (4.38). Moreover, $A \in \mathbb{R}^{q_2 \times H}$ (4.11), $D \in \mathbb{R}^{q_3 \times H}$ (4.15), $b \in \mathbb{R}^{q_2}$ (4.11), $c \in \mathbb{R}^{q_3}$ (4.16), and $n, m \in \mathbb{R}_+^H$. Thus, the feasible set $\mathcal{X}$ is given by (4.22).

In comparing problem (4.39) with the problem formulation in the publications Gabler et al. (2012) and Münnich et al. (2012c), we can identify the following extensions:

1. More than one equality constraint can be included.
2. Nonlinear inequality constraints can be considered (restrictions for regional efficiency).
3. The objective function depends on the weights w, as problem (4.39) considers an optimal *multivariate* allocation with a weighted sum scalarization; i.e. problem (4.39) has to be solved for all combinations of weights.

Thus, our method can be called a generalization of the box-constrained optimal univariate allocation of Gabler et al. (2012) and Münnich et al. (2012c).

The existence, optimality conditions, and uniqueness of the solution of the optimization problem are shown by applying a Lagrangian approach. The Lagrangian function is defined as $L : \mathbb{R}_+^H \times \mathbb{R}^{q_2} \times \mathbb{R}^{q_3} \times \mathbb{R}^H \times \mathbb{R}^H \to \mathbb{R}$ with

$$L(n, \lambda, \beta, \mu, \kappa) = d(w)^T \varphi(n) + \lambda^T (An - b) + \beta^T (D\varphi(n) - c) + \mu^T (n - M) + \kappa^T (m - n)$$

and Lagrangian multipliers $\lambda \in \mathbb{R}^{q_2}, \beta \in \mathbb{R}^{q_3}, \mu \in \mathbb{R}^H$, and $\kappa \in \mathbb{R}^H$. Following the theory of Section 3.1, corresponding (necessary and sufficient) optimality conditions for problem (4.39) can be defined by solving the KKT-system of Equations (3.4) to (3.6).

Theorem 4.3.1. A vector $n^* \in \mathbb{R}^H$ is a solution of problem (4.39) if and only if the Slater-condition is satisfied and there exist Lagrangian multipliers $\lambda^* \in \mathbb{R}^{q_2}$, $\beta^* \in \mathbb{R}_{0+}^{q_3}$, $\mu^* \in \mathbb{R}_{0+}^H$, and $\kappa^* \in \mathbb{R}_{0+}^H$ such that

$$\nabla_n \Big(d(w)^T \varphi(n^*)\Big) + A^T \lambda^* + \Big(\nabla_n (D\varphi(n^*) - c)\Big)^T \beta^* + \mu^* - \kappa^* = 0$$

$$An^* - b = 0$$

$$\Big(D\varphi(n^*\Big) - c) \leq 0 \tag{4.40}$$

$$\beta_j^*(D_j^T\varphi(n^*) \ - c_j) = 0 \, \forall j \in \{1, \dots, q_3\} \tag{4.41}$$

$$\mu_h^*(n_h^* - M_h) = 0 \ \ \forall h \in \{1, \dots, H\}$$

$$\kappa_h^*(m_h - n_h^*) = 0 \ \ \forall h \in \{1, \dots, H\}.$$

Proof. See Geiger and Kanzow (2002, Theorems 2.45 and 2.46). The necessity holds due to the assumptions that the Slater-condition is satisfied (see Definition 3.1.6). Since f and g are convex and continuously differentiable functions and h is a continuously differentiable affine-linear function, the feasible set $\mathcal{X}$ is convex (see Remark 3.1.5), so that the conditions are also sufficient. □

The KKT-system in Theorem 4.3.1 contains the nonlinear inequality constraints (4.40) and (4.41). Moreover, the Lagrangian multipliers $\beta^* \in \mathbb{R}_{0+}^{q_3}$, $\mu^* \in \mathbb{R}_{0+}^H$, and $\kappa^* \in \mathbb{R}_{0+}^H$ have to be non-negative. By applying a numerical solver, these inequality conditions are challenging as the KKT-system does not represent a system of equations. An alternative approach was introduced in Section 3.1, where the KKT-system can be equivalently rewritten using a NCP-function (see (3.7)). Thus, an alternative version of Theorem 4.3.1 is given in Theorem 4.3.2, in which the inequalities (4.40) and (4.41) are rewritten as equations using NCP functions. Moreover, the corresponding Lagrangian multipliers $\beta^* \in \mathbb{R}_{0+}^{q_3}$ do not have to be non-negative.

Theorem 4.3.2. A vector $n^* \in \mathbb{R}^H$ is a solution of problem (4.39) if and only if the Slater constraint qualification condition is satisfied and there exist Lagrangian multipliers $\lambda^* \in \mathbb{R}^{q_2}$, $\beta^* \in \mathbb{R}^{q_3}$, $\mu^* \in \mathbb{R}_{0+}^H$ and $\kappa^* \in \mathbb{R}_{0+}^H$ such that

$$\nabla_n\Big(d(w)^T\varphi(n^*)\Big) + A^T\lambda^* + \Big(\nabla_n(D\varphi(n^*) - c)\Big)^T\beta^* + \mu^* - \kappa^* = 0 \tag{4.42}$$

$$An^* - b = 0 \tag{4.43}$$

$$\min\Big(- \Big(D_j^T\varphi(n^*) - c_j\Big), \beta_j\Big) = 0 \ \ \forall j \in \{1, \dots, q_3\} \tag{4.44}$$

$$\mu_h^*(n_h^* - M_h) = 0 \ \ \forall h \in \{1, \dots, H\} \tag{4.45}$$

$$\kappa_h^*(m_h - n_h^*) = 0 \ \ \forall h \in \{1, \dots, H\}. \tag{4.46}$$

For the proof of Theorem 4.3.2, we refer to Theorems 3.1.7 and 4.3.1.

In contrast to Theorem 4.3.1, the KKT-system (4.42) to (4.46) exclusively consists of equality condition. In addition, the Lagrangian multipliers $\beta^* \in \mathbb{R}^{q_3}$ do not have to be positive numbers, which allows a solution strategy where β is iteratively updated. Nonetheless, Equation (4.44) is non-smooth, but we will see that this is no issue with regard to the applied solution strategy. In the following, we transform the KKT-system given by (4.42) to (4.46) in a way similar to the derivations in Münnich et al. (2012c). Some necessary properties are among others the separability of f, the affine-linearity of h, and the special structure of g, which is similar to the structure of f.

Lemma 4.3.3. Under the given assumptions of Theorem 4.3.2, Equation (4.42) is equivalent to

$$\left.\begin{array}{ll} 0 \geq (d_h(w) + D_h^T\beta^*)\varphi_h'(M_h) + A_h^T\lambda^*, & \text{if } n_h^* = M_h \\ 0 = (d_h(w) + D_h^T\beta^*)\varphi_h'(n_h^*) + A_h^T\lambda^*, & \text{if } n_h^* \in (m_h, M_h) \\ 0 \leq (d_h(w) + D_h^T\beta^*)\varphi_h'(m_h) + A_h^T\lambda^*, & \text{if } n_h^* = m_h \end{array}\right\} \text{ for } h = 1, \ldots, H \tag{4.47}$$

with component-wise defined

$$\varphi(n) = \Big(\varphi_1(n_1), \ldots, \varphi_H(n_H)\Big)^T = \left(\frac{1}{n_1}, \ldots, \frac{1}{n_H}\right)^T.$$

We note, that $\varphi_h(n_h)$ is continuously differentiable with derivative $\varphi_h'(n_h) = -1/n_h^2$. Since φ_h' is strongly monotonically increasing, the inverse

$$\varphi_h'^{-1}(z_h) = \sqrt{-1/z_h} \tag{4.48}$$

exists.

Proof. A closer look at Equations (4.45) and (4.46) as well as using the non-negativity of $\mu^* \in \mathbb{R}^H_{0+}$ and $\kappa^* \in \mathbb{R}^H_{0+}$ lead to following three equivalences for the optimal stratum-specific sample sizes n_h^* $(h = 1, \ldots, H)$ depending on the box-constraints:

$$\begin{array}{lcl} n_h^* = M_h & \Leftrightarrow & \mu_h^*(n_h^* - M_h) = 0,\ \mu_h^* \geq 0 \text{ and } \kappa_h^*(m_h - n_h^*) < 0,\ \kappa_h^* = 0 \\ n_h^* = m_h & \Leftrightarrow & \mu_h^*(n_h^* - M_h) < 0,\ \mu_h^* = 0 \text{ and } \kappa_h^*(m_h - n_h^*) = 0,\ \kappa_h^* \geq 0 \\ n_h^* \in (m_h, M_h) & \Leftrightarrow & \mu_h^*(n_h^* - M_h) < 0,\ \mu_h^* = 0 \text{ and } \kappa_h^*(m_h - n_h^*) < 0,\ \kappa_h^* = 0 \end{array}$$

for all $h = 1, \ldots, H$. Furthermore, each component of Equation (4.42) is equal to

$$\Big(d_h(w) + D_h^T\beta^*\Big)\varphi_h'(n_h^*) + A_h^T\lambda^* + \mu_h^* - \kappa_h^* = 0\ \forall h = 1, \ldots, H.$$

This is equivalent to the following three cases for $h = 1, \ldots, H$:

$$\begin{array}{ll} 0 \geq (d_h(w) + D_h^T\beta^*)\varphi_h'(M_h) + A_h^T\lambda^*, & \text{if } n_h^* = M_h \\ 0 = (d_h(w) + D_h^T\beta^*)\varphi_h'(n_h^*) + A_h^T\lambda^*, & \text{if } n_h^* \in (m_h, M_h) \\ 0 \leq (d_h(w) + D_h^T\beta^*)\varphi_h'(m_h) + A_h^T\lambda^*, & \text{if } n_h^* = m_h \end{array}$$

for all $h \in 1, \ldots, H$, which completes the proof. □

It should be noted here that the resembling structures of the objective function f and the function of the nonlinear inequality constraints g are necessary conditions for the verification of Lemma 4.3.3. If Equation (4.47) is revisited, the optimality conditions can be rewritten by using $\lambda \in \mathbb{R}^{q_2}$ and $\beta \in \mathbb{R}^{q_3}$ as dependent variables, and then the vector n can be defined depending on the choice of λ and β. To achieve this, a function $n : \mathbb{R}^{q_2+q_3} \to \mathbb{R}^H_+$ is component-wise defined as

$$n_h(\lambda, \beta) = \left\{\begin{array}{ll} M_h, & \text{if } -\frac{A_h^T\lambda}{d_h(w)+\beta^T D_h} \geq \varphi_h'(M_h) \\ \varphi_h'^{-1}\left(-\frac{A_h^T\lambda}{d_h(w)+\beta^T D_h}\right), & \text{if } \varphi_h'(m_h) < -\frac{A_h^T\lambda}{d_h(w)+\beta^T D_h} < \varphi_h'(M_h) \\ m_h, & \text{if } -\frac{A_h^T\lambda}{d_h(w)+\beta^T D_h} \leq \varphi_h'(m_h) \end{array}\right. \tag{4.49}$$

$$= \text{Proj}_{[m_h, M_h]}\left(\varphi_h'^{-1}\left(-\frac{A_h^T\lambda}{d_h(w) + \beta^T D_h}\right)\right)$$

for all $h = 1, \dots, H$ and $\varphi_h^{'-1}$ defined in (4.48). Combining (4.49) with (4.43) and (4.44) results in a nonlinear system of equations of dimension $q := q_2 + q_3$ that only depends on the Lagrangian multipliers λ and β. This system is defined by

$$\Phi(\lambda, \beta) := \begin{cases} An(\lambda, \beta) - b \\ \min\Big(- \Big(D\varphi\big(n(\lambda, \beta)\big) - c\Big), \beta\Big) \end{cases} = 0 \tag{4.50}$$

with $\Phi : \mathbb{R}^q \to \mathbb{R}^q$ and the component-wise defined minimum function. The goal is to solve the nonlinear system of equations in (4.50) of dimension $q = q_2 + q_3$ instead of the optimization problem in (4.39) of dimension H. Common solution strategies for nonlinear optimization such as Lagrange-Newton, SQP, and trust region methods do not use the special structure of this problem. All these strategies contain an iterative algorithm, where the optimized variable $n \in \mathbb{R}_+^H$ and the Lagrangian multipliers need to be updated in each iteration (see Section 3.1 for further details). In comparison to these methods, solving (4.50) would be associated with a reduction of the dimension to $(q_2 + q_3)$, which is independent of the dimension H of the optimization problem (4.39). Moreover, as H is comparatively high and $H \gg q_2$ as well as $H \gg q_3$, the computational burden is generally expected to be significantly reduced compared to standard solvers for nonlinear optimization. This is particularly the case for selected applications in official statistics, such as nationwide household or business surveys. The following theorem proofs the equivalence of both problem formulations.

Theorem 4.3.4. A vector $n^* \in \mathbb{R}_+^H$ is the unique solution of the optimization problem (4.39) if and only if there exists multipliers $\lambda^* \in \mathbb{R}^{q_2}$ and $\beta^* \in \mathbb{R}^{q_3}$ such that $n(\lambda^*, \beta^*)$ defined in (4.49) satisfies

$$\Phi(\lambda^*, \beta^*) = 0. \tag{4.51}$$

Proof. If $(n^*, \lambda^*, \beta^*)$ are given such that (4.47) holds, we compute $n(\lambda^*, \beta^*)$ as defined in (4.49). By means of the three given cases in Lemma 4.3.3, it can then be observed that $n(\lambda^*, \beta^*) = n^*$. By contrast, if for some (λ^*, β^*) the vector $n(\lambda^*, \beta^*)$ satisfies (4.51), then we can easily verify that $\big(n(\lambda^*, \beta^*), \lambda^*, \beta^*\big)$ also satisfies (4.47). This completes the proof. □

Finally, Theorem 4.3.4 verifies that the weighted sum scalarized MMDopt problem (4.20) of dimension H, both with and without additional restrictions for regional efficiency, can be solved via solving the significantly lower dimensional nonlinear system of equations (4.50) of dimension $q = q_2 + q_3$. An appropriate algorithm is suggested in Subsection 4.3.2.

Sensitivity analysis

In the previous paragraphs, the equivalence of the nonlinear system of equations in (4.51) to the weighted sum scalarized problem (4.20) was proved. Moreover, we have shown in Subsection 4.2.4, that solving (4.51) for all possible combinations of weights enables the computation of the whole Pareto frontier of optimization problem (4.20). Regarding this notion, it is desirable to analyze the sensitivity of the solution of the weighted sum scalarized MMDopt concerning

1. the stratum-specific variances S_{ih}^2 for all strata $h = 1, \dots, H$ and all variables y_i $(i = 1, \dots, q_1)$, which are used as input data (as seen in (4.10)), and

2. the combinations of weights $w = (w_1, \dots w_{q_1})^T \in [0,1]^{q_1}$.

The sensitivity and robustness of `MMDopt` concerning the stratum-specific variances S^2_{ih} is of great importance if these values are not known but are instead estimated in advance. Thus, the question is how the uncertainty in S^2_{ih} influences the solution of the optimal allocation computed by the method `MMDopt`. To analyze this, the derivative of the variance formula of variables y_i

$$\text{Var}(\hat{\tau}^{\text{StrRS}}_{y_i}) = \sum_{h=1}^{H} \frac{S^2_{ih} N^2_h}{n_h(\lambda, \beta)} \left(1 - \frac{n_h}{N_h}\right) \tag{4.52}$$

concerning S^2_{ih} needs to be computed. Using this, $n_h(\lambda, \beta)$ is computed by (4.49). The evaluation of the derivatives is a challenging task, as S^2_{ih} is comprised in $d_h(w)$ and $d_h(w)$ is comprised in the distinction of cases of the components of $n_h(\lambda, \beta)$. This entails the computation of the derivative of the projection functions, which is highly non-trivial. Hence, we omit the analytical sensitivity analysis in this thesis, but we do practically investigate the robustness of `MMDopt` concerning uncertainty deviations of S^2_{ih} in the application study (see Subsection 4.6.5).

A sensitivity analysis concerning the weights $w = (w_1, \dots, w_{q_1})^T$ would reveal the same problems stated above. Thus, this analysis is also omitted, but it can be done numerically by analyzing the *gradient* between the different points of the Pareto frontiers (see Section 4.4).

4.3.2 Semismooth Newton method

Theorem 4.3.4 verifies the equivalence and enables solving $\Phi(\lambda, \beta) = 0$ instead of (4.39). However, a standard Newton method cannot be applied, since Φ is not continuously differentiable due to both the projection in (4.49) and the minimum function in the second component of Φ. Alternatively, Münnich et al. (2012c) choose a fixed-point algorithm to solve $\Phi = 0$ for the special case of $q_2 = 1$ and $q_3 = 0$, i.e.

$$\Phi(\lambda) := \sum_{h=1}^{H} n_h(\lambda) - n_s = 0. \tag{4.53}$$

This approach turns out to be computationally efficient, as an optimization problem of dimension $H \approx 20\,000$ is solved within a fraction of a second. To make use of this increase in efficiency, we suggest using a `SSN` method (see Algorithm 1) to solve (4.51), which was proposed by Münnich et al. (2012b) in the context of calibration. Therefore, we need to verify the semismoothness (cf. Definition 3.2.4 and Qi and Sun, 1993) of Φ.

Theorem 4.3.5. The function Φ defined in (4.50) is semismooth.

Proof. According to Qi and Sun (1993), the minimum and maximum functions are strongly semismooth. Since $\text{Proj}_{[m_h, M_h]}(x) = \min\left\{M_h, \max\{m_h, x\}\right\}$ for $x \in \mathbb{R}$, the projection is semismooth as a composition of semismooth functions (Lemma 3.2.5, Item 5). In the same manner, $A_i^T n(\lambda, \beta) - b_i$ $(i = 1, \dots, q_2)$ and $\min\{-(D_j^T \varphi(n^*) - c_j), \beta_j\}$ $(j = 1, \dots, q_3)$ are semismooth. Then, since all components of Φ are semismooth (and are also Lipschitz-continuous), Φ is semismooth due to Item 3 of Lemma 3.2.5. □

Finally, the application of the SSN method in solving MMDopt and MMDopt.reg with a weighted sum scalarization (4.39) is possible. Due to Theorem 3.2.7, the nonlinear system of equations in (4.50) converges superlinearly to the unique optimal solution (λ^*, β^*) if the elements of the B-subdifferential $H \in \partial_B \Phi(\lambda^*, \beta^*)$ is regular. In applying Theorem 4.3.4, the unique optimal solution of the H-dimensional problem (4.39) can be computed via solving the $(q_2 + q_3)$-dimensional problem (4.49). To summarize, the allocation problem of dimension H is solved using a significantly lower dimensional compensatory problem given by (4.50). Due to the different structure of building the Jacobian for $\Phi(\lambda)$ in the MMDopt case ($q_3 = 0$) and for $\Phi(\lambda, \beta)$ in the MMDopt.reg case, the appropriate semismooth Newton algorithms are separately implemented and are called SSN and SSN.reg respectively.

In the practical applications, two assumptions need to be satisfied in order to achieve a robust convergence of the SSN and SSN.reg methods. On the one hand, the initial iterate (λ^0, β^0) needs to be close enough to the solution and on the other hand, all $H \in \partial_B \Phi(\lambda^*, \beta^*)$ must be regular. To guarantee this, we propose a range of procedures that can *optionally* be included before and within the iterations:

1. The initial value for λ is inherited from the box-constrained optimal allocation of Münnich et al. (2012c), computed by a fixed-point iteration.

2. Due to the different scales of the first part of equations $An(\lambda, \beta) - b$ and the second part of the equations $\min\{-(D\varphi(n(\lambda, \beta)) - c), \beta\}$, the equations need to be standardized before the iteration. This supports a similar scale of the components of the Newton update as well as a good condition of the elements of the B-subdifferential of Φ, which has to be computed in each iteration.

3. A preconditioned element of the B-subdifferential of Φ facilitates a good condition of the linear system in the Newton step.

4. An Armijo step-size strategy (see Algorithm 2) prevents step-sizes that are too large within the iterations.

Since stratum-specific sample sizes need to be integer numbers (a fraction of a person cannot be drawn in a sample), in general the results of SSN need to be rounded. Therefore, a simple rounding strategy is suggested in Algorithm 3, which leaves the sum of all stratum-specific sample sizes unchanged. For a more detailed discussion of the integrity of the solutions, we refer to Subsection 4.6.7.

Algorithm 3 Smart rounding with unchanged sum (smart.round)

```
Input: Vector $n \in \mathbb{R}^H_+$ with real numbers and $n_s = \sum_{h=1}^H n_h \in \mathbb{N}$
  compute $\tilde{n} = \lfloor n \rfloor$
  compute $\eta = n - \tilde{n} \in \mathbb{R}^H$
  compute $\delta_{\text{diff}} = \sum_{h=1}^H \eta_h$
  select $\delta_{\text{diff}}$ components $j \in \text{Comp} \subseteq \{1, \ldots, H\}$, which correspond to biggest values of $\eta$
  set $\tilde{n}_j \leftarrow \tilde{n}_j + 1 \; \forall j \in \text{Comp}$
Return: Solution $n \leftarrow \tilde{n}$
```

4.4 Algorithmic solution depending on decision-making function

In many practical applications, the decision-maker will likely use a decision-making function to choose a specific justifiable optimal multivariate allocation. Thus, the computation of the whole Pareto frontier of the underlying vector-optimization problem

$$\text{(VP)} \quad \min_{n \in \mathcal{X}} F(n) := \Big(F_1(n), \ldots, F_{q_1}(n)\Big) \tag{4.54}$$

is sufficient, as described in Section 4.3. In a second step, the decision-maker needs to choose *one* of the Pareto optimal solutions that is considered preferable. To select the preferable Pareto optimal solution, the decision-maker will likely rely on certain decision-making functions. Some of these have also been introduced as scalarization techniques in Subsection 4.2.3. In particular, a p-norm of the objectives (4.26) and a min-max approach (4.27) are commonly used, which are both consistent with the theory of multi-criteria optimization presented in Subsection 3.3.1. The higher p is, the more attention is given to the objectives with higher values. Depending on the choice of the standardization technique presented in Section 4.2.2, these values correspond with variances (without standardization), coefficients of variation ((cv)-standardization), or relative losses of accuracy compared to optimal univariate allocations ((opt)-standardization). The effects of different p-norms are demonstrated in the following example.

Example 4.4.1. This example shows the different solutions of the multi-objective optimization problem, which is scalarized by a p-norm:

$$\min_{n \in [1,10]} \left\| \Big(F_1(n), F_2(n)\Big)^T \right\|_p$$

with the two selected functions (no variance functions, for illustration only)

$$F_1 : \mathbb{R} \to \mathbb{R} : n \mapsto (n-4)^2 + 3 \quad \text{and} \quad F_2 : \mathbb{R} \to \mathbb{R} : n \mapsto \frac{100}{n+2}$$

on the closed feasible set $n \in [1, 10]$. Figure 4.1 shows that, if p increases, the solution (represented as a red dot) approaches the intercept point of F_1 and F_2. In that regard, the Pareto optimal solutions are illustrated in blue. For the min-max approach, the solution would be the intercept point. Thus, the higher the value of p, the more attention is paid to reduce the highest value of the objectives.

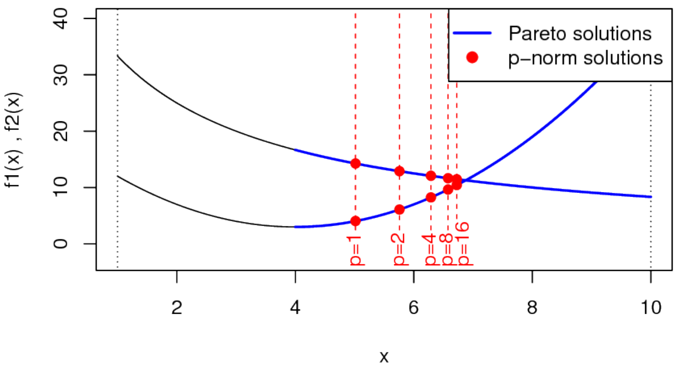


Figure 4.1: Example using the p-norm as decision-making function.

Figure 4.1 shows the possibly high influence of the scalarization technique on the solution of the problem. However, with the exception of $p = 1$ the structure of the p-norm scalarized problem

$$\text{(P.OP)} \quad \min_{n \in \mathcal{X}} \sum_{i=1}^{q_1} F_i(n)^p, \tag{4.55}$$

disables the highly efficient solution strategy applied to the weighted sum scalarized problem

$$\text{(Ws.OP)} \quad \min_{n \in \mathcal{X}} \sum_{i=1}^{q_1} w_i F_i(n) \tag{4.56}$$

via the SSN algorithm, which was detailed in Section 4.3. More precisely, Lemma 4.3.3 is not fulfilled due to the non-separability of the objective function of the p-norm scalarized problem (P.OP). Other possible solvers for the non-separable optimization problem (P.OP) are inefficient in general, as they do not make use of a strategy to reduce the dimension of the KKT-system to $(q_2 + q_3)$, which is generally significantly lower than the number of strata H and is also independent of H.

In the following, an alternative solution strategy for $p < \infty$ is presented for the p-norm scalarized problem (P.OP) based on the algorithmic solver of the weighted sum scalarized problem (Ws.OP) in Section 4.3. The strategy relies on a connection of (P.OP) and (Ws.OP), which is used to significantly shrink the set of feasible solutions for (P.OP). In addition, we show that the alternative strategy is also applicable to solve the min-max problem (i.e. $p = \infty$) with a certain precision.

In a first step, we need to prove a property of strictly convex functions with a strictly positive image set. Lemma 3.3.16 and Theorem 3.3.17 show that solving (Ws.OP) for all possible combinations of weights ($w_i \geq 0$, $\sum_{i=1}^{q_1} w_i = 1$) yields *all* Pareto optimal solutions of the original problem (VP). Their proofs are based on the strict convexity of the components of F, namely F_i $(i = 1, \ldots, q_1)$, and the convexity of the bounded feasible set $\mathcal{X}$ (see Subsection 4.2.1). Using these properties, the following lemma holds for a general real-valued function f.

Lemma 4.4.2. Let $\mathcal{X} \subseteq \mathbb{R}_+^H$. Let the strictly convex function $f : \mathcal{X} \to \mathbb{R}_+$ be twice continuously differentiable with a strictly positive image space (i.e. $f(n) > 0 \; \forall n \in \mathcal{X}$). Then, f^p is also strictly convex for an integer number $p \geq 1$.

Proof. Since f is strictly convex, the Hessian $\nabla^2 f(n)$ is positive definite for all $n \in \mathcal{X}$. Due to the power gradient rule, the gradient of f^p is given by

$$\nabla f^p(n) = p f^{p-1}(n) \nabla f(n).$$

Then, the following equation holds for the hessian of f^p:

$$\nabla^2 f^p(n) = p(p-1) f^{p-2}(n) \nabla f(n) (\nabla f(n))^T + p f^{p-1}(n) \nabla^2 f(n).$$

To show that $\nabla^2 f^p(n)$ is positive definite on $\mathcal{X}$, let $z \in \mathcal{X}$. Then

$$z^T \nabla^2 f^p(n) z = \underbrace{p(p-1) f^{p-2}(n)}_{\geq 0} \underbrace{z^T \left(\nabla f(n) (\nabla f(n))^T \right) z}_{\geq 0 \;\; (*)} + \underbrace{p f^{p-1}(n)}_{>0} \underbrace{z^T \nabla^2 f(n) z}_{>0} > 0.$$

The expression $(*)$ holds, as $\nabla f(n)(\nabla F(n))^T$ is a rank-1 matrix and is therefore positive semidefinite. Thus, the hessian $\nabla^2 f^p(n)$ is positive definite for all $n \in \mathcal{X}$, which completes the proof. □

The elements of the image space of the strictly convex functions F_i defined in (4.21) are strictly positive for all $i = 1, \dots, q_1$, since variances or coefficients of variation are strictly positive in general. Consequently, F_i^p is also strictly convex for $p \geq 1$ due to Lemma 4.4.2. Thus, the relationship between (VP) and (Ws.OP) that solving (Ws.OP) yields all Pareto optimal solutions of (VP) can be similarly applied to the p-norm scalarized problem (P.OP) and the corresponding vector-optimization problem

$$(\text{VP.P}) \quad \min_{n\in\mathcal{X}} \Big(F_1(n)^p, \dots, F_{q_1}(n)^p\Big). \tag{4.57}$$

For (VP) and (Ws.OP), the connection between solving all (Ws.OP) problems and the Pareto frontier of (VP) has been proved in Lemma 3.3.16 and Theorem 3.3.17. Accordingly, the solution of the p-norm scalarized problem (P.OP) is *one* Pareto optimal solution of (VP.P), since (P.OP) can be expressed as a weighted sum scalarized (VP.P) problem and equal weights.

Moreover, the Pareto frontier of (VP) is equal to the Pareto frontier of (VP.P). This is because of the equivalence

$$\min_{n\in\mathcal{X}} F_i(n) \quad \Leftrightarrow \quad \min_{n\in\mathcal{X}} F_i(n)^p, \tag{4.58}$$

which holds due to strict convexity and strict positiveness of F_i as well as the convexity and boundedness of $\mathcal{X}$. Thus, the solution of the p-norm scalarized problem (P.OP) is an element of the Pareto frontier of the original multi-criteria optimization problem (VP), which is illustrated in Figure 4.2. This fact offers the opportunity to shrink the feasible set $\mathcal{X}$ of the p-norm scalarized problem (VP.P) to the set of Pareto optimal solutions of the original multi-criteria problem (VP), which is given by

$$\mathcal{X}_{\text{OPT}} := \Big\{ n \in \mathcal{X} : n \in \mathbb{R}_+^H \text{ is Pareto optimal solution of (VP)} \Big\}. \tag{4.59}$$

$$\begin{tikzcd}
(\text{VP}) \ \min\limits_{n\in\mathcal{X}} \Big(F_1(n), \dots, F_{q_1}(n)\Big) \arrow[r, Leftrightarrow] \arrow[d, Leftrightarrow] & (\text{Ws.OP}) \ \min\limits_{n\in\mathcal{X}} \sum\limits_{i=1}^{q_1} w_i F_i(n) \quad \forall w_i \in [0,1], \ \sum w_i = 1 \\
(\text{VP.P}) \ \min\limits_{n\in\mathcal{X}} \Big(F_1(n)^p, \dots, F_{q_1}(n)^p\Big) & (\text{P.OP}) \ \min\limits_{n\in\mathcal{X}} \sum\limits_{i=1}^{q_1} F_i(n)^p \arrow[l, Rightarrow] \arrow[u, Rightarrow]
\end{tikzcd}$$

("$\Longleftrightarrow$" means, that the sets of optimal solutions is equal)

("$\Longrightarrow$" means, that the optimal solution is element of the set of optimal solutions)

Figure 4.2: Connection between multi-objective problems and scalarized problems.

Hence, if the *whole* set of Pareto optimal solutions $\mathcal{X}_{\text{OPT}}$ of (VP) is computed via the SSN algorithm (see Algorithm 1) with the solution strategy presented in Section 4.3, only the specific element which is the minimal solution of (P.OP) has to be determined in $\mathcal{X}_{\text{OPT}}$. Each element of $\mathcal{X}_{\text{OPT}}$ corresponds to a specific vector of weights $w \in [0,1]^{q_1}$ with $\sum_{i=1}^{q_1} w_i = 1$. Hence, only the combination of weights, which is assigned to the optimal solution of the p-norm scalarized problem (P.OP) has to be determined.

To determine the combination of weights assigned to the optimal solution of the p-norm scalarized problem (P.OP), we propose an iterative solver based on a projected inexact quasi-subgradient method (`GTM`). In that regard, the iterates of the method are the weights corresponding to the elements of the Pareto frontier of the original multi-criteria problem (VP). Its Pareto frontier is given by

$$\mathcal{Y}_{\text{OPT}} := \Big\{ \big(F_1(n), \dots, F_{q_1}(n) \big)^T \in \mathbb{R}^{q_1} : n \in \mathcal{X}_{\text{OPT}} \Big\} \tag{4.60}$$

as a subset of the image space $\mathbb{R}^{q_1}_+$. The major advantage of the `GTM` algorithm compared to the direct solving of (P.OP) is the dimension of the feasible set of the underlying problem. Since the iterates are the weights $w \in [0,1]^{q_1}$ of (Ws.OP) and q_1 is small (e.g. having values of about $q_1 = 3$), the computational burden is significantly reduced compared to direct solvers, which would have to deal with the high dimension H and the non-separability of problem (P.OP).

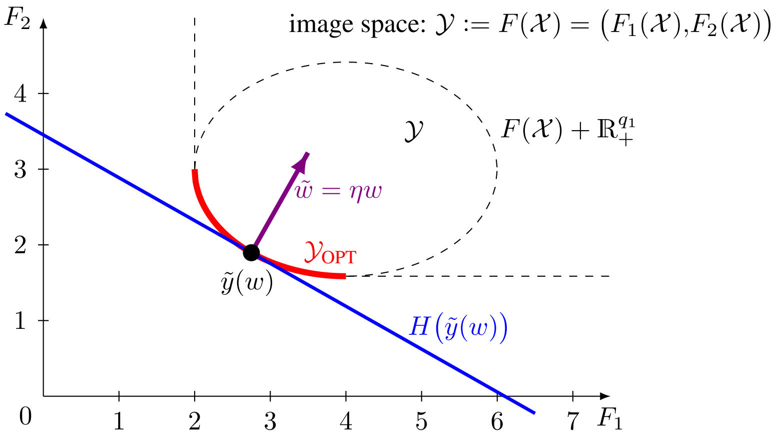
To apply `GTM` to solve the p-norm scalarized problem (P.OP), some preliminary studies have to be considered with regard to the Pareto frontier $\mathcal{Y}_{\text{OPT}}$ of the original problem (VP). As proved in Lemma 3.3.16, the epigraph $C_+(F)$ of F is convex. By applying Definition 3.3.6 and the argumentation in the proof of Theorem 3.3.9, we can define a hyperplane

$$H\big(\tilde{y}(w)\big) := \Big\{ x \in \mathbb{R}^{q_1} : \tilde{w}^T \big(x - \tilde{y}(w)\big) = 0 \Big\} \tag{4.61}$$

for each Pareto optimal solution $\tilde{y}(w) \in \mathcal{Y}_{\text{OPT}}$ with the condition

$$\tilde{y}(w) \in H\big(\tilde{y}(w)\big) \cap \mathcal{Y}_{\text{OPT}} \tag{4.62}$$

as a tangent to $\mathcal{Y}_{\text{OPT}}$ (see Figure 4.3). Thus, $\tilde{w} \in \mathbb{R}^{q_1}$ is the normal vector to $H(\tilde{y}(w))$, which corresponds with the vector of weights w associated with the Pareto optimal solution $\tilde{y}(w)$ (i.e. $\tilde{w} = \eta w$, $\eta > 0$; cf. Craft et al., 2006, p. 3401 and Bokrantz and Forsgren, 2013, p. 379). Moreover, the Pareto frontier here is a subset of the boundary of the convex epigraph of F defined by $F(\mathcal{X}) + \mathbb{R}^{q_1}_+$, which generally only holds under convexity assumptions.



F_2
image space: $\mathcal{Y} := F(\mathcal{X}) = \big(F_1(\mathcal{X}), F_2(\mathcal{X})\big)$
4
3
2
1
0
$\mathcal{Y}$
$F(\mathcal{X}) + \mathbb{R}^{q_1}_+$
$\tilde{w} = \eta w$
$\mathcal{Y}_{\text{OPT}}$
$\tilde{y}(w)$
$H\big(\tilde{y}(w)\big)$
1
2
3
4
5
6
7
F_1

Figure 4.3: Pareto frontier $\mathcal{Y}_{\text{OPT}}$ of (VP) for $q_1 = 2$.

To apply a `GTM` algorithm to solve the p-norm scalarized problem (P.OP), we define the functions

$$G_i : \mathbb{R}_+^{q_1} \to \mathbb{R}_+, w \mapsto F_i\Big(n(w)\Big) = y_i(w) \quad (i = 1, \ldots, q_1). \tag{4.63}$$

These functions depend on the weights of (Ws.OP), where $n(w) \in \mathcal{X}_{\text{OPT}}$ is the Pareto optimal solution of (VP) associated with the weights w in (Ws.OP). Moreover, $y_i(w)$ is the component i of the corresponding element belonging to the Pareto frontier, i.e. $(y_1(w), \ldots, y_{q_1}(w))^T \in \mathcal{Y}_{\text{OPT}}$. Thus, function G_i maps each combination of weights to the corresponding objective value $F_i\Big(n(w)\Big) = y_i(w)$, which is an element of the Pareto frontier of the weighted sum scalarized problem (Ws.OP). Using (4.63), we can define the function

$$G^p : \mathbb{R}_+^{q_1} \to \mathbb{R}_+, w \mapsto \sum_{i=1}^{q_1} G_i^p(w) = \sum_{i=1}^{q_1} F_i\Big(n(w)\Big)^p = \sum_{i=1}^{q_1} y_i(w)^p \tag{4.64}$$

as the sum of the components to the power of p of the elements of the Pareto frontier of (VP) depending on the weights of (Ws.OP). It has to be noted here that each evaluation of G^p requires the solution of one (Ws.OP)-problem using the `SSN` algorithm, in particular

$$y_i(w)^p = F_i\Big(n(w)\Big)^p = F_i\Big(\operatorname*{arg\,min}_{n \in \mathcal{X}} \sum_{i=1}^{q_1} w_i F_i(n)\Big) \text{ for all } i = 1, \ldots, q_1. \tag{4.65}$$

Overall, we have shown that the solution of the H-dimensional p-norm scalarized problem (P.OP) is equivalent to the solution of the q_1-dimensional problem

$$\begin{aligned} \min_{w \in [0,1]^{q_1}} \; & G^p(w) = \sum_{i=1}^{q_1} F_i\Big(n(w)\Big)^p \\ \text{s.t.} \; & \sum_{i=1}^{q_1} w_i = 1. \end{aligned} \tag{4.66}$$

In looking at problem (4.66), it is an optimization problem with a not necessarily differentiable objective function and a convex feasible set

$$W := \Big\{ w \in [0,1]^{q_1} : \sum_{i=1}^{q_1} w_i = 1 \Big\}. \tag{4.67}$$

An intuitive approach to solve problem (4.66) is to solve (Ws.OP) for a high resolution of the weights (e.g. up to three decimals) and choose the solution $n(w^*) \in \mathbb{R}_+^H$ which is minimal for G^p. Such an enumerative approach has the major disadvantage in that it requires a huge computational burden. For a resolution of the weights of 0.002 and $q_1 = 3$, the problem (Ws.OP) needs to be solved $125\,751$ times. Moreover, tests have shown that this high computational effort is still accompanied with a not negligible rounding error.

To omit an enumerative strategies, `GTM` is developed to utilize a different approach based on a *descent algorithm*. Generally, descent algorithms are given by an iterative procedure

$$w^{k+1} = w^k + \alpha_k d^k, \tag{4.68}$$

where $d^k \in \mathbb{R}^{q_1}$ is a descent direction at the iterate w^k, and α_k is a predetermined step-size in iteration k (cf. Spellucci, 1993, p. 97). One of the most prominent examples of a descent algorithm is the gradient method for continuously differentiable functions, where the descent directions are determined by the negative gradient (cf. Ruszczynski, 2006, pp. 218 ff.). Since properties such as Lipschitz-continuity and differentiability are not given or known for G^p and the feasible set of (4.66) is bounded, the application of a classical gradient method is therefore not possible. Instead, the use of a projected subgradient method is possible, with this method given as

$$w^{k+1} = \text{Proj}_W\Big(w^k + \alpha_k d^k\Big), \tag{4.69}$$

with $d^k = -\|s^k\|^{-1}s^k$, and s^k is an element of the B-subdifferential of $G^p(w^k)$ (cf. Geiger and Kanzow, 2002, pp. 366 ff. and Algorithm 6.50). However, the subgradient of G^p is not given explicitly since each evaluation of G^p requires the solution of one (Ws.OP)-problem. Alternatively, a subgradient s^k at the iterate w^k can be approximated by a hyperplane containing q_1 surrounding points, as shown in Figure 4.4. Thus, the GTM algorithm may be characterized as an inexact projected subgradient method. Nevertheless, we can only assume convergence, if G^p would be a convex function with regard to the weights w. To weaken this assumption Hu et al. (2015) proposed such methods for the case where G^p is required to be quasi-convex and proved convergence results. Under this circumstances, an *inexact projected quasi-subgradient method* can be applied for (4.69), in which the descent direction is computed by

$$d^k = -\Big(\|\tilde{s}^k\|^{-1}\tilde{s}^k + r^k\Big). \tag{4.70}$$

In that regard, $\tilde{s}^k$ is an element of the so-called ν_k*-quasi-subdifferential* $\partial^*_{\nu_k}G^p(w^k)$ given by

$$\partial^*_{\nu_k}G^p(w^k) := \Big\{z : z^T(u - w^k) \leq 0 \; \forall u \in \{v \in \mathbb{R}^{q_1} : G^p(v) \leq G^p(w^k) - \nu_k\}\Big\} \tag{4.71}$$

with a certain noise $r^k \in \mathbb{R}^{q_1}$ and error $\nu_k \in \mathbb{R}$ in iteration k. Due to the highly technical level, we will not present the theory of Hu et al. (2015) in greater detail in this thesis. Indeed, this is not necessary for our settings, given that we interpret the noise r^k and the error ν_k as the deviation of the approximated descent direction s^k from the (non-computable and therefore

Algorithm 4 Inexact projected quasi-subgradient method GTM for solving MMDopt ($p \neq 1$)

```
Input: Choose a high resolution of weights (e.g. $0.0001$), choose a starting combination
       of weights $w^0 \in [0,1]^{q_1}$, set $k = 0$ and $d^0 >$ tol.
  while $\|d^k\| \leq$ tol
    Compute $q_1$ combinations of weights $w^{(1)}, \dots, w^{(q_1)} \in [0,1]^{q_1}$ adjoining $w^k$.
    Solve (Ws.OP) for $w^{(1)}, \dots, w^{(q_1)}$; solutions are given by $n(w^{(1)}), \dots, n(w^{(q_1)}) \in \mathbb{R}^H$.
    Compute $(q_1 - 1)$-dimensional hyperplane in $\mathbb{R}^{q_1}_+$, which contains the points
      $G^p\Big(n(w^{(1)})\Big), \dots, G^p\Big(n(w^{(q_1)})\Big) \in \mathbb{R}^{q_1}$.
    Compute $d^k = -\|s^k\|^{-1}s^k$, where $s^k$ is an approximation of the element $\tilde{s}^k$ of the
      $\nu_k$ quasi-subdifferential $\partial^*_{\nu_k}G^p(w^k)$ of $G^p(w^k)$ with the noise $r^k$ and error $\nu_k$.
    Compute an appropriate step-size $\alpha_k > 0$ depending on $d_k$.
    Compute the next solution $w^{k+1} = w^k + \alpha_k d^k$.
  end while
Return: Solution $n^* = n(w^k)$.
```

only theoretical) element $\tilde{s}^k \in \partial^*_{\nu_k} G^p(w^k)$. We note, that the approximated descent direction s^k is computed by the q_1 surrounding points (see Figure 4.4). Following this, a pseudo-code of GTM is shown in Algorithm 4. It starts with an initial combination of weights $w^0 \in W$ with $F(n(w^0)) \in \mathcal{Y}_{\text{opt}}$. Thereafter, in each iteration k the gradient of G^p in w^k is approximated by q_1 surrounding dots. In a second step, an appropriate step-size $\alpha_k > 0$ is determined. Then, the iterate w^k is updated. The algorithm stops, if a predefined tolerance is reached.

For a graphical illustration of GTM for $q_1 = 3$ and variable dummies V1,V2, and V3, we refer to Figure 4.4, where five iterations (It.) of GTM are plotted exemplarily. The triangle represents the space of all combination of weights. Each grey dot represents one combination of weights with a scaling resolution of 0.1. The weight for one specific variable is marked on the respective axis. Since the weights need to sum up to 1.0, this results in 66 combinations of weights. Thus, if the value of $G^p(w)$ is assigned to the respective dot and is illustrated by a related color, the triangle can be interpreted as a heatmap of G^p depending on the weights w.

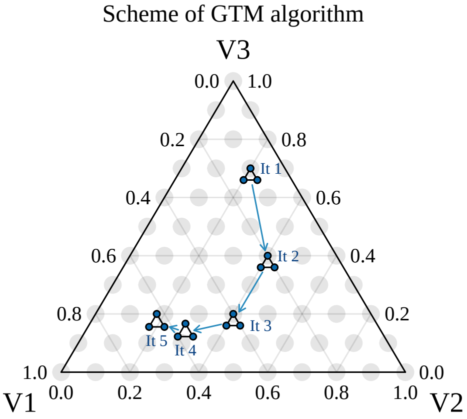


Figure 4.4: Iterative GTM method to solve (P.OP) for $q_1 = 3$.

In general, a projected subgradient method converges to a local optimum of the objective function. Under the assumption of quasi-convexity for the function G^p depending on the weights w, some global convergence results for GTM can be stated in analogy to Hu et al. (2015), which are dependent on the sequences of the noises $\{r^k\}$, the errors $\{\nu^k\}$, and the step-sizes $\{\alpha_k\}$. In a first step, the quasi-convexity of G^p is proved in the following lemma.

Lemma 4.4.3. The function G^p defined in (4.64) is quasi-convex on the convex set W defined in (4.67) for $p \geq 1$.

Proof. Due to Definition 3.1.4, we need to prove that for any $w, u \in W$ and any $\alpha \in [0,1]$

$$G^p\Big(\alpha w + (1-\alpha)u\Big) \leq \max\Big(G^p(w), G^p(u)\Big).$$

Without loss of generality, let $G^p(u) \geq G^p(w)$. The corresponding points of the Pareto frontier of the weights w and u are given by $y(w), y(u) \in \mathcal{Y}_{\text{OPT}}$. As $\Big(\alpha w + (1-\alpha)u\Big)$ describe the connecting line between w and u, the following equation holds:

$$\Big(\alpha w + (1-\alpha)u\Big) \in \bigtimes_{i=1}^{q_1} \Big[\min\{w_i, u_i\}, \max\{w_i, u_i\}\Big].$$

Specifically, the connection line is located in the q_1-dimensional rectangle defined by the component-wise minima and maxima of w and u. Since the epigraph $C_+(F)$ of F is convex (see Lemma 3.3.16) and its boundary contains the Pareto frontier $\mathcal{Y}_{\text{OPT}}$, the Pareto frontier describes a convex curve in the image space $\big(F_1(\mathcal{X}), \dots, F_{q_1}(\mathcal{X})\big)$ of the multi-criteria problem (VP) (see Figures 4.3 and 4.5 for the case $q_1 = 2$). Thus, the connecting line between $y(w)$ and $y(u)$ lies above $\mathcal{Y}_{\text{OPT}}$. Each point of the connecting line can be characterized by $\mu y(w) + (1-\mu)y(u)$ with $\mu \in [0,1]$. Moreover, if $a \leq b$, it follows that $a^p \leq b^p$ for some scalar values $a, b \geq 0$.

Then, there exist $\delta \geq 0$ and a specific $\mu \in [0,1]$ such that the following holds:

$$\begin{aligned}
G^p\Big(\alpha w + (1-\alpha)u\Big) &= \sum_{i=1}^{q_1} F_i\Big(n\Big(\alpha w + (1-\alpha)u\Big)\Big)^p && (4.72)\\
&\leq \left(\sum_{i=1}^{q_1} F_i\Big(n\Big(\alpha w + (1-\alpha)u\Big)\Big)^p\right) + \delta && (4.73)\\
&= \left(\sum_{i=1}^{q_1} \mu F_i\Big(n(w)\Big)^p + (1-\mu)F_i\Big(n(u)\Big)^p\right) && (4.74)\\
&= \mu G^p(w) + (1-\mu)G^p(u) && (4.75)\\
&\leq \mu G^p(u) + (1-\mu)G^p(u) && (4.76)\\
&\leq G^p(u). && (4.77)
\end{aligned}$$

The step from (4.73) to (4.74) holds, since the addends of (4.74) define a specific point on the connecting line between the elements of the Pareto frontier $y(w)$ and $y(u)$. This is fully determined by the choice of μ, as illustrated in Figure 4.5 for $q_1 = 2$. The step from (4.75) to (4.76) holds due to the assumption $G^p(u) \geq G^p(w)$. Thus, we have shown the quasi-convexity of G^p for $p \geq 1$.

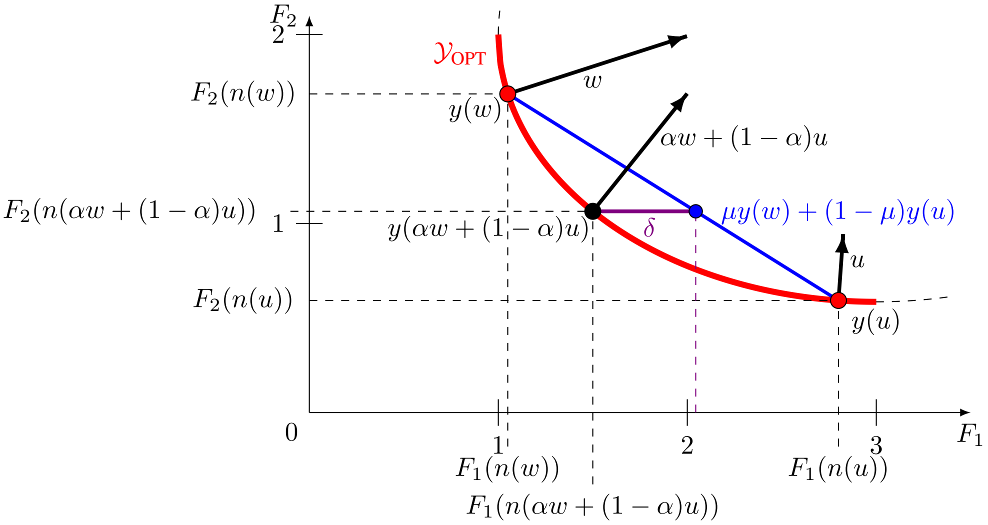


Figure 4.5: Illustration of Pareto frontier $\mathcal{Y}_{\text{OPT}}$ of (VP) for $q_1 = 2$.

□

Since the quasi-convexity of G^p is proved in Lemma 4.4.3, a closer look to theoretical results concerning the choice of the step-size parameters α_k is necessary to present convergence results for GTM. As each functional evaluation of G^p requires the solution of an optimization problem (Ws.OP), the elements of the subgradient of $G^p(w^k)$ are not explicitly given. As a consequence, this does not permit the application of a common step-size rule, such as the Armijo-rule (cf. Algorithm 2), which is based on several function evaluations of G^p. Therefore, other strategies needs to be considered in determining the step-size parameters α_k for each iteration.

Commonly, strategies for the determination of the step-size parameters are given for the general setting of descent methods for the minimization of the Fréchet-differentiable function $f : \mathbb{R}^{q_1} \to \mathbb{R}$ on a feasible set $\mathcal{X}$. In this sense, a vector $\tilde{d}^k$ is called a descent direction if $\nabla f(x^k)^T \tilde{d}^k \leq 0$ (cf. Sachs and Sachs, 2011). For the inexact gradient method, the steepest descent direction is computed by the negative gradient of $f(x^k)$ with a perturbation r^k, i.e.

$$\tilde{d}^k = -\nabla f(x^k) + r^k. \tag{4.78}$$

In order to determine the next iterate

$$x^{k+1} = x^k + \alpha_k \tilde{d}^k, \tag{4.79}$$

an appropriate choice of the step-size parameter α_k guarantee convergence and may significantly increase the convergence rate. For the differentiable case, some suitable strategies are proposed by Sachs and Sachs (2011) for general Fréchet-differentiable functions $f : \mathcal{X} \to \mathbb{R}$ for $\mathcal{X} \subseteq \mathbb{R}^{q_1}$, which still guarantee certain convergence statements. To summarize the statements of Sachs and Sachs (2011, Theorem 2) and Sachs and Sachs (2011, Corollary 1), an inexact gradient method given by

$$x_{k+1} = x^k + \alpha_k d^k \;\text{ with }\; d^k = -\nabla f(x^k) + r^k \tag{4.80}$$

for the minimization of the Fréchet-differentiable function f on the feasible set $\mathcal{X}$ converges, if (besides some additional assumptions) the sequence of step-size parameters α_k for all iterations satisfies the following conditions:

$$\sum_{k=1}^{\infty} \alpha_k^2 < \infty \;\text{ and }\; \sum_{k=1}^{\infty} \alpha_k = \infty. \tag{4.81}$$

For a more detailed analysis, we refer to Sachs and Sachs (2011).

Nevertheless, differentiability of the objective function G^p is not given in our case. Thus, convergence results for GTM need to consider this fact. Hu et al. (2015) presented some convergence results for GTM for the solution w^* of problem (4.66) with quasi-convex objective function G^p under the following assumptions.

(A1) The feasible set W of problem (4.66) is compact.

(A2) G^p satisfies the Hölder condition of order $\bar{p} > 0$ with modulus $\kappa > 0$ on $\mathbb{R}^{q_1}$, i.e.

$$\left\| G^p(w) - G^p(w^*) \right\| \leq \kappa \left\| w - w^* \right\|^{\bar{p}} \;\text{ for all } w \in \mathbb{R}^{q_1}.$$

(A3) The noise and errors are bounded, i.e. there exist some $B, \nu \geq 0$ such that

$$\|r^k\| \leq B \;\text{ for all } k \geq 0 \;\text{ and }\; \limsup_{k\to\infty} \nu_k = \nu.$$

Using these assumptions, two convergence results are given in Hu et al. (2015) for two different step-size strategies.

Theorem 4.4.4. Let assumptions (A1) to (A3) hold. Then, for a sequence $\{x^k\}$ generated by the GTM algorithm with the *constant* step-size $\alpha_k \geq 0$, we have

$$\liminf_{k\to\infty} G^p(w^k) \leq G^p(w^*) + \kappa \left(Bd + \frac{\alpha_k}{2}(1+B)^2 \right)^{\bar{p}} + \nu,$$

where $\|w^k - w\| \leq d$ for all $k \geq 0$ and $w \in W$.

For the proof of Theorem 4.4.4, we refer to Hu et al. (2015, Theorem 3.1).

Theorem 4.4.5. Let assumptions (A1) to (A3) hold. Then, for a sequence $\{x^k\}$ generated by the GTM algorithm with a sequence of step-sizes $\{\alpha_k\}$ given by

$$\alpha_k > 0, \;\; \lim_{k\to\infty} \alpha_k = 0 \;\text{ and }\; \sum_{k=1}^{\infty} \alpha_k = \infty,$$

we have

$$\liminf_{k\to\infty} G^p(w^k) \leq G^p(w^*) + \kappa (Bd)^{\bar{p}} + \nu.$$

For the proof of Theorem 4.4.5, we refer to Hu et al. (2015, Theorem 3.2).

Theorems 4.4.4 and 4.4.5 show convergence for problem (4.66) to the optimal value within some tolerance given in terms of errors ν_k and noises r^k and depending on the step-size strategy. Since W is bounded and closed, and therefore compact, a convergence of the GTM algorithm can be assumed under the above mentioned assumptions (A1) to (A3) if the approximations of the descent directions s^k correspond to the theory of the ν_k-quasi-subdifferential $\partial^*_{\nu_k} G^p(w^k)$ and its parameters ν_k and r^k. To underpin the analytically proved global convergence of GTM under specific assumptions, a global convergence has been observed in all of the tested applications without a specific predetermination of the assumptions of Theorems 4.4.4 and 4.4.5.

A detailed analysis of the performance of GTM is given in the application study in Section 4.6.6. The accuracy of the optimal allocation $n^* = n(w^*)$ strongly depends on two factors, namely the resolution of the weights while solving (4.66) with the GTM method and also the provided tolerance while solving problem (Ws.OP) with the SSN method. To avoid intolerable high approximation errors, both values should be chosen small enough. The practical applicability of this approach is tested in Section 4.6.3. It turns out that for the common case $q_1 = 3$, approximately 20 iterations are required. Thus, solving a p-norm scalarized multivariate allocation problem requires approximately 60 solutions of the weighted sum scalarized problem via the SSN method. Since this method is extremely fast, the performance of the GTM is also very efficient. A detailed analysis of the numerical performance is given in Section 4.6.6.

The proposed method GTM is only valid for $p < \infty$. The search for an appropriate and computationally efficient solution strategy for the min-max case $p = \infty$ turned out to be more challenging. The Lagrangian approach proposed for the weighted sum case (or $p = 1$) in Section 4.3 also failed due to the non-differentiability of the objective function (4.27). Thus, Theorem 4.3.1 of the optimality conditions does not hold. Nonetheless, a quite accurate approximation of the solution of the min-max approach can be computed by GTM if p is large enough. This is due to the fact that the solution of a p-norm scalarized MMDopt converges to the solution of the min-max scalarized MMDopt for $p \to \infty$ (cf. Lin, 2005). Several simulations with various scenarios and parameters have shown that a choice of $p = 128$ leads to approximations, which are generally precise enough for almost all real applications.

4.5 Summary of methods

Prior to the analysis of the developed optimal multivariate and multi-domain allocation method in an extensive application study in Section 4.6, the methods and algorithms are summarized in the following.

In general, the MMDopt method is a very flexible method to apply an optimal allocation. In a first step, it can be selected whether the optimal allocation should comply with box-constraints for the stratum-specific sample sizes and quality restrictions for regional estimators. Thereafter, one of the two standardization techniques presented in Subsection 4.2.2 ((cv)- or (opt)-standardization) has to be chosen. In the next step, it is left to the user to select a characterization of the optimality (see Section 3.3) to solve the multi-criteria optimization problem of the form (4.5). In the context of this work, scalarization approaches are presented in Subsection 4.2.3. On the one hand, a weighted sum scalarization can provide the prerequisite for the determination of the entire Pareto frontier of the original problem. This approach enables the calculation of the set of all optimal solutions according to the definition of Pareto optimality presented in Subsection 3.3.2. In order to obtain a specific optimal allocation, it can be chosen between p-norm scalarization and min-max scalarization. The choice of the parameter $p \in \mathbb{N}$ is based on the preferences of the user.

Beside the influence on the solution of the allocation problem, the selection of the scalarization technique also determines the applicability of the corresponding solution algorithms. In general, it is always possible to use a standard solver for restricted optimization problems, which are briefly described in Section 3.1. In order to avoid the direct solution, which may be expensive in high dimensions (i.e. for problems with a high number of strata), several alternative solution strategies are proposed in this thesis depending on the choice of the scalarization technique. Using a weighted sum scalarization, the original problem can be transformed into a significantly lower dimensional nonlinear system of equations by transforming the optimality conditions. The resulting lower dimensional problem can then be solved in a time-efficient way with the semismooth Newton method (SSN) presented in Section 3.2. Necessary conditions for this strategy are the separability and convexity of the objective function of the standardized and scalarized problem as well as the convexity of the feasible set. In addition, the similarity of the structures of the objective function and the functions of the regional quality restrictions is inevitable; both the objective function and the functions of the restrictions are built by variance functions (see Section 4.3). By using a p-norm scalarization, the separability of the objective function is not given, so that a direct application of the SSN method is not possible. However, the discussion in Section 4.4 has shown that the p-norm scalarized problem can be traced back to the weighted sum problem. Hence, the application of the SSN method is still possible. This approach is only verifiable due to the strict convexity and strict positivity of the variance functions. Thus, the p-norm scalarized problem can also be solved efficiently with a linear dependency between the computing time and the dimension H of the original problem by using a projected inexact quasi-subgradient method (GTM). When using the min-max scalarization instead, the scalarized objective function loses the property of the continuous differentiability. Thus, the presented methods are not applicable. However, detailed simulations have shown that, by selecting a high value of p in the p-norm scalarization, a solution for the min-max scalarization is generally precise enough for almost all applications.

Finally, it can be summarized that the presented methods represent different ways to solve optimal multivariate and multi-domain allocation problems with several constraints in a time-efficient way by exploiting the specific structure of the problems. Besides, a linear dependency between the computation time and the dimension H of the original problem (number of the strata of the sampling design) can be observed. As the specific structures are exploited, the applicability of the developed approaches has some limits.

For the sake of clarity and comprehensibility, the methods and algorithms are illustrated in a graphical overview in Figure 4.6. In that regard, blue boxes indicate statistical methods, orange boxes represent numerical algorithms, and green boxes show the resulting output.

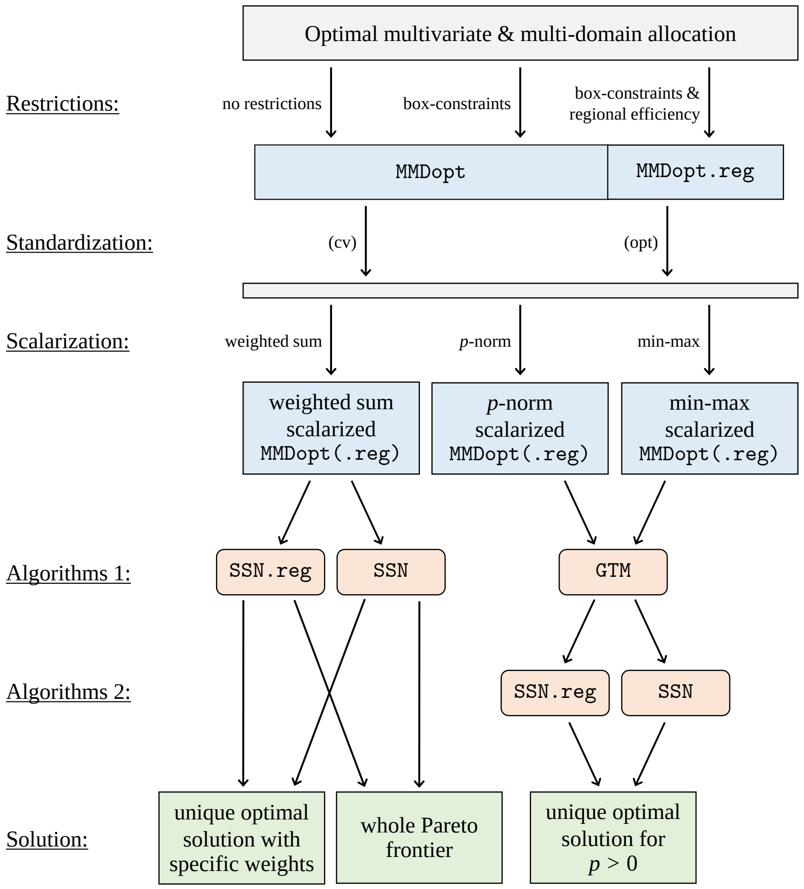

Optimal multivariate & multi-domain allocation
Restrictions:
no restrictions
box-constraints
box-constraints & regional efficiency
MMDopt
MMDopt.reg
Standardization:
(cv)
(opt)
Scalarization:
weighted sum
p-norm
min-max
weighted sum scalarized MMDopt(.reg)
p-norm scalarized MMDopt(.reg)
min-max scalarized MMDopt(.reg)
Algorithms 1:
SSN.reg
SSN
GTM
Algorithms 2:
SSN.reg
SSN
Solution:
unique optimal solution with specific weights
whole Pareto frontier
unique optimal solution for p > 0

Figure 4.6: Summary of the optimal multivariate and multi-domain allocation method MMDopt.

4.6 Application study and results

4.6.1 Framework

The application study is based on the synthetic RIFOSS dataset introduced in Section 2.6. We consider the household structure and confine ourselves to the federal states of Hesse, North Rhine-Westphalia, Rhineland-Palatinate, and Saarland. This leads to a population size of $11\,121\,631$ households accommodating $30\,077\,329$ individuals. The sampling design is based on strata built as cross-classifications of sampling points (SMP – 784 regional areas) and classes of household sizes (HHS – 8 classes). This results in $784 \cdot 8 = 6\,272$ cross-classification strata, which are simply called $\mathtt{strata}$. In addition to the population ($\mathtt{pop}$) estimates and stratum-specific estimates, regional estimates are also evaluated for $\mathsf{NUTS2}$-regions ($\mathsf{NUTS2}$; 12), $\mathsf{NUTS3}$-regions ($\mathsf{NUTS3}$; 121), and for the SMPs (SMP; 784). Both NUTS2- and NUTS3-regions are unions of SMPs. The overall sampling fraction is fixed to 1%, i.e. the total sample size is given by $n_{\mathrm{s}} = 111\,216$. The lower bound for each cross-classification strata h is set to $m_h = 2$ (in order to ensure the computation of variances), the upper bound is set to the stratum size $M_h = N_h$ (in order to avoid an overallocation). The size distribution of the cross-classification strata in terms of the number of included households is shown in Table 4.3, where a strong homogeneity of the strata can be observed with regard to their size. The smallest stratum consists of two and the largest one contains over $120\,000$ households, which complicates the allocation.

Table 4.3: Quantiles of the size of the $6\,272$ cross-classification strata.

0%	20%	40%	60%	80%	100%
2	335	602	$1\,004$	$1\,851$	$123\,652$

In the following evaluations, six variables are considered, which are split into auxiliary variables (used for the allocation) and variables of interest. The stratum-specific totals and variances of the auxiliary variables are assumed to be known at the sampling stage. By contrast, the values of the variables of interest are unknown, as their estimation is the aim of the survey. The variables household income per person ($\mathsf{INC.PP}$), number of people over 65 ($\mathsf{AGE7.7}$), and number of children under 15 ($\mathsf{AGE7.1}$) are used as variables of interest. The major goal is to gain accurate total estimates for these variables at regional level as well as at population level. Generally, *good* proxies for the stratum-specific variances are assumed to be known in advance and can be used as input data for the allocation. In practice, these proxies may be gained in various ways, such as by evaluating previous surveys, using obtainable register data, or using highly correlated auxiliary variables. To meet this in an application framework, some proxies could be computed under the consideration of certain trends and random perturbations in the variables of interest. Alternatively, we prefer to use other variables (i.e. auxiliaries), which are pairwise highly correlated with the variables of interest as input data for the allocation. On the one hand, this allows a detailed analysis of the functionality of the developed allocation method by comparing the accuracy of the estimates of the *auxiliary variables*. On the other hand, this ensures an extensive and realistic evaluation of the quality of the estimates of the *variables of interest* in comparison to other allocation techniques. Therefore, the variables equivalized disposable household income (EDI), value of pensions (PEN), and number of people under 20 (AGE4.1) are used as (auxiliary) variables for the MMDopt method (i.e. $q_1 = 3$). Since their stratum-specific variances are assumed to be known, these variables are used for the allocation process (i.e. $q_1 = 3$).

In order to achieve accurate estimates for the variables of interest, the auxiliaries and variables of interest are pairwise highly correlated. In particular, their correlations are given by

- $\text{cor}\Big(\mathsf{PEN}, \mathsf{AGE7.7}\Big) = 0.72$,
- $\text{cor}\Big(\mathsf{EDI}, \mathsf{INC.PP}\Big) = 0.97$, and
- $\text{cor}\Big(\mathsf{AGE4.1}, \mathsf{AGE7.1}\Big) = 0.92$.

Besides these, the correlation between the three auxiliary variables are small, which is challenging on the one hand but helps to emphasize the advantages of MMDopt on the other hand. If the three auxiliaries would be highly correlated, MMDopt yields similar results compared to an optimal univariate allocation, i.e. the application of an optimal multivariate allocation method such as MMDopt would be senseless. The correlation structures of the auxiliaries are shown in Figure 4.7, where we make a distinction between correlations within strata, SMPs, NUTS3-, and NUTS2-regions. Each dot in the boxplots represents the correlation within one specific region. The correlations over the whole population are plotted as red vertical lines. Since the correlations are located around zero and most often do not exceed values of ± 0.2, we expect strongly different optimal *univariate* allocations for each of the three variables. In such a case, we suppose that the greatest improvements can be observed by applying the developed MMDopt method. Nevertheless, as we see later, the usage of highly correlated variables will not lead to disadvantages compared to other standard allocation techniques.

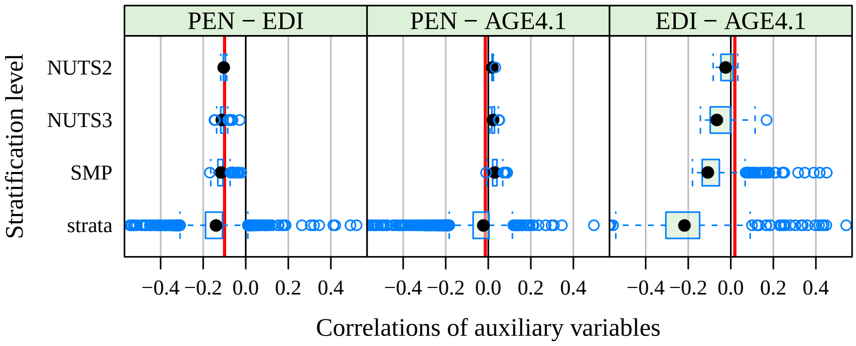


Figure 4.7: Boxplots of the correlations of the auxiliary variables for the population (vertical red lines), NUTS2, NUTS3, SMP, and strata.

To evaluate the results and analyze strengths and weaknesses of the developed method MMDopt, we compare the results among others with the following allocation methods:

- prop proportional allocation,

- `uni.[var]` optimal univariate allocation with box-constraints for variable `[var]`, and
- `Cmin` minimal costs allocation, which is to minimize n_s subject to variance restrictions (in analogy to the first strategy of Dalenius, 1953; $n_s \approx 111\,216$ is experimentally achieved by an adjustment of the restrictions).

Approaches of the type of `Cmin` are very common in practice, nevertheless the `Cmin` methods completely neglect the minimization of variances or coefficients of variations, which is one of the key points of `MMDopt`. As we will see in the analysis of the application, the minimization of the variance of the total population estimator has indisputable advantages. The methods denoted with `uni.[var]` are based on the box-constrained optimal univariate allocation published in Gabler et al. (2012), Münnich et al. (2012c), and Münnich et al. (2012a), where only one variable is considered in the allocation process. This tends to be insufficient in some applications, especially in modern surveys with conflicting aims.

We compare the HT estimates under various settings by means of the relative root mean squared error (RRMSE). Since the design-based HT estimates are unbiased in stratified random sampling per definition, the comparison of their RRMSEs is equivalent to the comparison of their variances (Lohr, 2009, Chapter 2). In particular, the RRMSE of an estimator equals its relative standard deviation. Since the RIFOSS dataset is fully available, the values to be estimated are known and will be used for the direct computation of the RRMSEs. Thus, a Monte-Carlo simulation study can be omitted in this study.

As mentioned in Subsection 4.2.5, the method is also applicable for the GREG estimator by adopting the objective functions. Nevertheless, this study focuses on HT estimates, since the aim of the study is to compare various allocation strategies with one another without generating undesirable side effects. Some of the following figures and graphs contain relative values that compare the RRMSEs, variances, or sample sizes of the developed methods in relation to the case of an independent optimal univariate allocation (`uni.[var]`) of the three auxiliary variables. This allows a straight comparison of advantages and disadvantage of `MMDopt`.

The application study is divided into several parts. The functionality of `MMDopt` concerning the weighted sum optimal allocation is analyzed in Subsection 4.6.2. From there, the resulting RRMSEs of population total and area-specific estimates are shown in dependence of the predefined weights. Moreover, these results are assembled to build the whole Pareto frontier, as proved in Subsection 4.2.4. The standardization methods (cv) and (opt) are compared among one another in each setting. Subsequently, the different decision-making functions (namely p-norms and the min-max approach) are compared in Subsection 4.6.3. The evaluations of Section 4.4 enable solving these problems for various decision-making functions with the `GTM` algorithm. Since the inclusion of quality restrictions for regional estimates is omitted up to this point, the effects of these restrictions on various stratification levels (cf. (4.3)) are illustrated and particularly compared with existing allocation methods in Subsection 4.6.4. A detailed sensitivity and robustness analysis concerning inaccurate input data is presented in Subsection 4.6.5. Finally, the numerical performance of the algorithms is investigated in Subsection 4.6.6. Due to the conventions determined of Chapter 1, the results of the auxiliaries are generally shown as plots with headers shaded in blue and green. By contrast, red and orange shades are chosen for the variables of interest. Supplementary figures are additionally shown in the Appendix B.2.

4.6.2 Weighted sum optimal allocation problem

In this subsection, we first focus on the weighted sum method with predefined weights. As illustrated in Subsection 4.2.4, this strategy facilitates the computation of the whole set of Pareto optimal solutions. In Figure 4.8, the RRMSEs of NUTS3-specific estimates are plotted for ten selected combinations of weights for the *auxiliary variables*, which are used for the allocation (for settings with (cv)- and (opt)-standardization). In Figure 4.9, the RRMSEs of the corresponding estimates on NUTS3 level of the *variables of interest* are plotted.

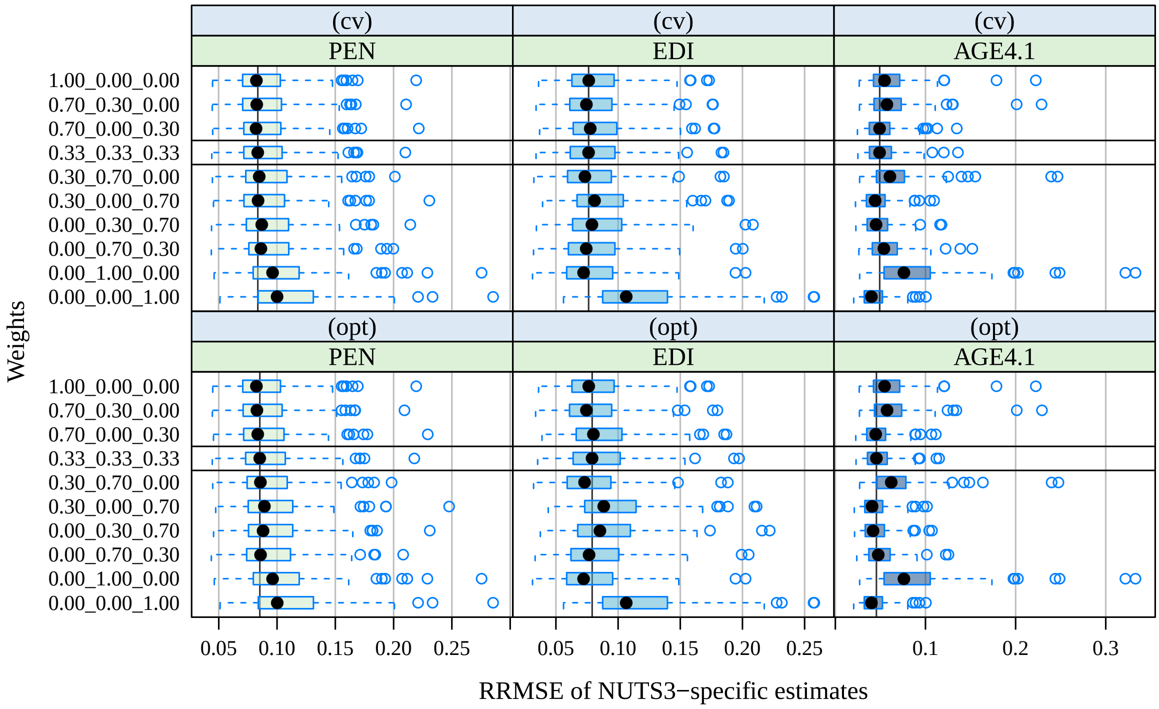


Figure 4.8: RRMSE of the NUTS3-specific total estimates for ten combinations of weights with (cv)- and (opt)-standardization (auxiliaries).

The setting in row 4 corresponds to an equal weighting. In most cases, a stronger weighting of a variable coincides with a lower RRMSE of the estimate for the NUTS3-specific total in comparison to the equal weighting case. Since the auxiliaries are desired to be pairwise highly correlated with the variables of interests, the statement holds for the auxiliaries as well as for the variables of interest. Nevertheless, the connection between correlation structure and efficiency is not a general statement, and it depends on a few factors, including the correlation structure of the data. The settings in the row 1, 9, and 10 are equal to the optimal univariate allocations with respect to one of the three auxiliary variables. The errors of the estimates for the other variables are comparably high especially in these cases since they are assigned with a weight of zero. This is a strong argument for the necessity of an optimal *multivariate* allocation. In comparing the standardizations, (cv) and (opt) do not lead to significant differences. By using the (opt)-standardization, a slightly more compensated error-increases (compared to the optimal univariate allocations) over all variables and all NUTS3 regions can be observed than by using the (cv)-standardization. This can be seen by comparing row 4 with the optimal univariate allocations in row 1, 9, and 10. In addition, the estimates of AGE4.1 are more efficient for the (opt)-standardization, if a weight of greater than 0.0 is assigned to AGE4.1. This is due to the high influence of the respective term in the objective function (see Table 4.2; the value of AGE4.1 for the (opt)-standardization dominates). Results for stratification levels apart of the NUTS3-regions (which have been plotted here exemplarily) follow similar patterns (see Figure B.1 in the Appendix B.2).

The RRMSEs of the corresponding NUTS3-specific estimates of the variables of interest plotted in Figure 4.9 follow similar patterns. This is mostly due to the high correlation between the auxiliaries and the variables of interest. Thus, MMDopt is able to generate allocations which allow for an accurate estimation of variables of interest that are fully unknown in the selection process. In that regard, the driving force is their pairwise correlation with the auxiliaries. In general, this statement also holds for the optimal univariate allocation. Nevertheless, if the correlation between auxiliary variable and variable of interest in unknown, the user is not able to presume the advantages gained by the optimal univariate allocation. Applying MMDopt, this is rather given, because several auxiliaries are utilized in the allocation.

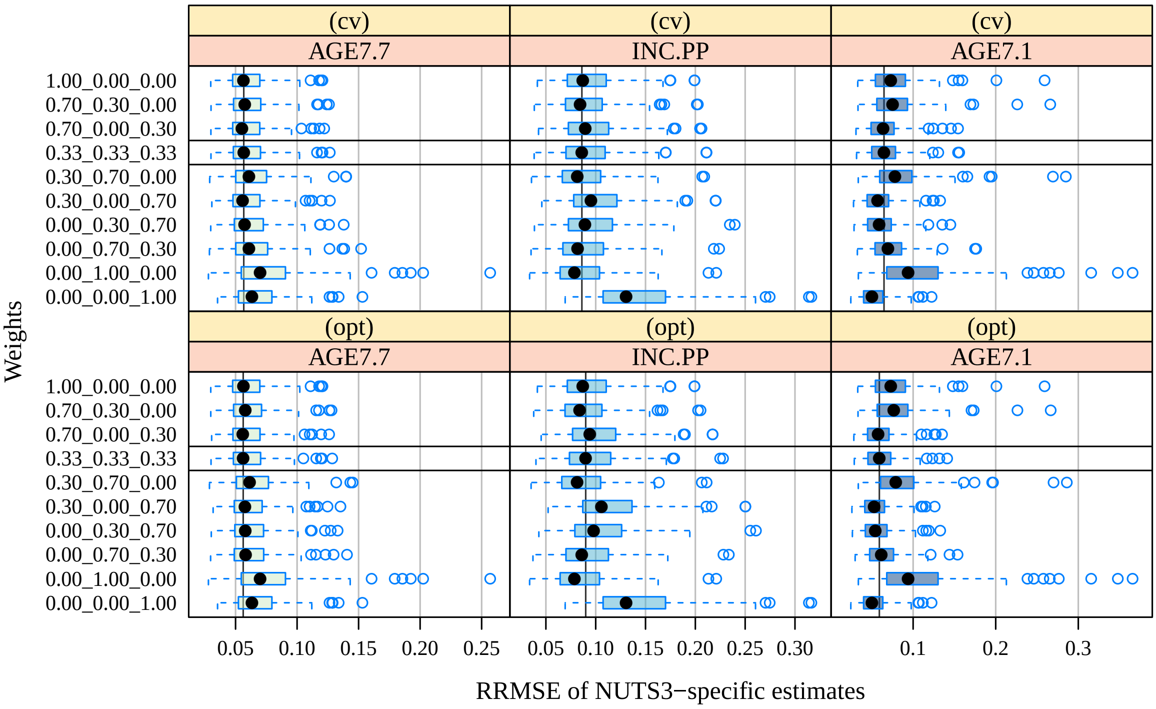


Figure 4.9: RRMSE of the NUTS3-specific total estimates for ten combinations of weights with (cv)- and (opt)-standardization (variables of interest).

In the following, the variances of the population total estimates are compared for all possible combinations of weights with a scaling resolution of 0.1. Since the weights have to sum up to 1.0, this results in 66 combinations of weights. Solving the weighted sum scalarized MMDopt problem (4.56) provides the opportunity to plot the whole Pareto frontier, even though this is subject to the resolution of the weights. To illustrate this, the increase of the RRMSEs of the population total estimates for the auxiliaries is plotted relative to the RRMSE using an optimal univariate allocation in the heatmaps in Figure 4.10 for the three auxiliary variables. The three heatmaps on the left correspond to the (cv)-standardized results, while the ones to the right correspond to the (opt)-standardization. Each dot represents one combination of weights. The weight for one specific variable is marked on the respective axis. To facilitate the reading, each variable and the values of its corresponding weight is plotted in the same color. In that regard, the weight of a dot for one variable (for instance PEN) can be determined by going along that connection line of the dot, which has the same color as the variable (red for PEN). The color of the dot represents the relative increase of the RRMSE of the estimate of the variable stated in the title of the respective plot.

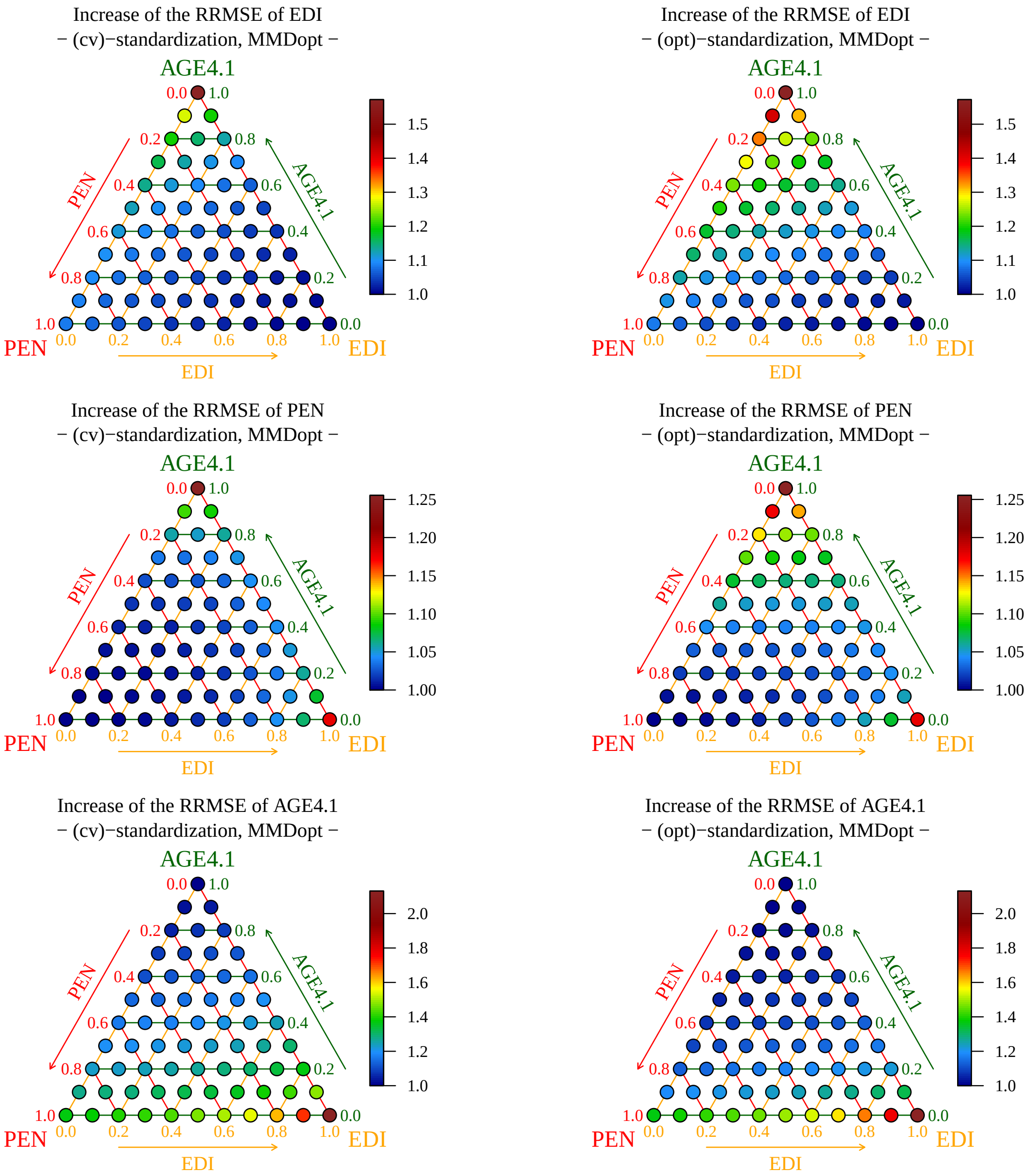


Figure 4.10: Relative increase of the RRMSE of the population total estimates under (cv)- and (opt)-standardization for 66 combinations of weights for each auxiliary variable.

As a consequence of the scaling resolution of 0.1, the dot which represents the equal weighting $w = (1/3, 1/3, 1/3)$ is not contained in the heatmaps. However, they can be approximated by the surrounding dots (we refer to the *quasi-convexity* of the Pareto frontier proved in Section 4.4). Dots with a (dark) blue color that corresponds to the lower end of the color scale are favorable because they represent combinations of weights with a lower increase of the RRMSE in comparison to the optimal univariate allocation. As an example, the smallest RRMSE for variable EDI (heatmaps in row 1) can be located at the vertex where variable EDI is given the full weight 1.00. The RRMSEs differ depending on the choice of the standardization strategy.

In analyzing the heatmaps, a full weight of 1.0 corresponding to an optimal univariate allocation with respect to the respective variable yields the most accurate estimate of the population total for the respective variable. In general, the smaller the weight, the higher the error-increase becomes. For instance, the maximum RRMSE of the estimate of variable EDI can be located at the vertex where AGE4.1 has the full weight 1.00. At that point, the RRMSE is approximately 55% higher in comparison to the optimal univariate allocation. In comparing the behavior of the relative error-increases of the estimates of variable AGE4.1, a high error-increase of over 100% can be observed if a weight of 1.00 is assigned to EDI. In comparison, the error-increase is more moderate ($\approx 40\%$) if PEN has a full weight of 1.00. These values are not directly accompanied by the correlation structures of the variables presented in Figure 4.7, where the correlation between EDI and AGE4.1 exceeds the correlation between PEN and AGE4.1. This observation shows that the behavior of optimal allocations is not directly comparable to the correlation structure of the variables used for the allocation.

Similar to Figure 4.8, the setting with (opt)-standardization (right-hand side) results in slightly more balanced error-increases. While the accuracy of the population estimate of AGE4.1 by means of roughly equal weights is better when using the (opt)-standardization, it is poorer with respect to the estimates of EDI and PEN. As the (opt)-standardization reduces the percentage increases of AGE4.1, which may assume the highest values, it may be considered as more balanced (see Table 4.2). This effect is investigated in detail in Subsection 4.6.3.

The structure of the heatmaps in Figure 4.11 is equivalent to Figure 4.10, whereas now the *cumulated* error-increase of the total estimates of the three auxiliaries is plotted. The heatmap for each standardization is derived by dividing the sum of the three respective heatmaps of Figure 4.10 by the value of 3.0. Since the RRMSEs of the estimates of all considered variables are contained, these heatmaps may give an impression of the overall quality of the allocation. Nevertheless, they do not provide any information regarding the quality of the estimates for the individual variables. In the case of (cv)-standardization, the best choice in the sense of a low cumulated error-increase is an asymmetric weighting (PEN 0.2, EDI 0.2, AGE4.1 0.6). By contrast, the setting with (opt)-standardization is more balanced, and thus the lowest cumulated increase of RRMSEs can be reached by using roughly equal weights (PEN 0.4, EDI 0.3, AGE4.1 0.3).

Due to Theorem 3.3.15, each dot in the heatmaps represents the RRMSE of one Pareto optimal solution (i.e. Pareto optimal vector of the stratum-specific sample sizes). To be precise, the dots along the edges (where at least one weight is zero) are only weakly Pareto optimal, which will not be decisive in the following analysis. The heatmaps for the variables of interest follow similar patterns and are shown in Figure B.2. By combining the heatmaps of the three auxiliaries

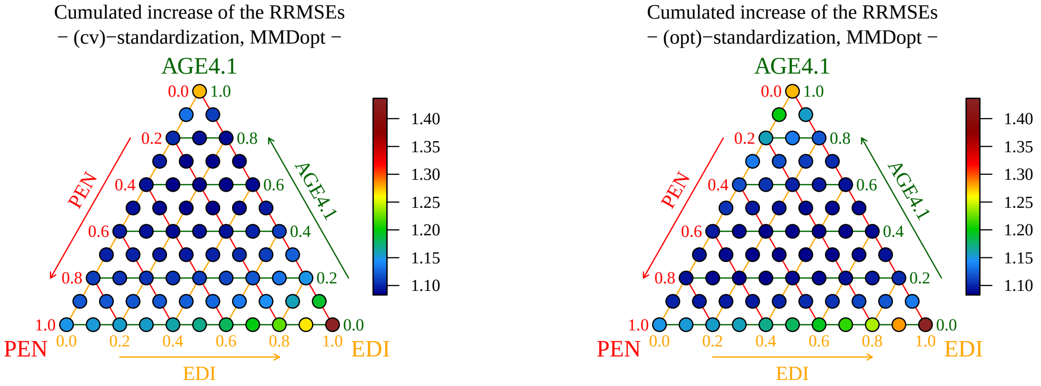


Figure 4.11: Relative cumulated increase of the RRMSE of the population total estimates under (cv)- and (opt)-standardization for 66 combinations of weights (auxiliaries).

in one plot, we can display the Pareto frontier as a subset of a three-dimensional cube, which is done in Figure 4.12. Each dot in the three-dimensional space represents the values of the objective functions of one Pareto optimal solution relative to the optimal univariate allocations. For this plot, the scaling resolution of 0.01 is chosen, resulting in $5\,152$ combinations of weights. Each of the three axes represents the RRMSE-increase of the population total estimate of the corresponding variable. The colors of the dots reflect the combination of weights (PEN red, EDI green, AGE4.1 blue). For instance, a green dot belongs to a combination of weights, where a high weight is assigned to EDI.

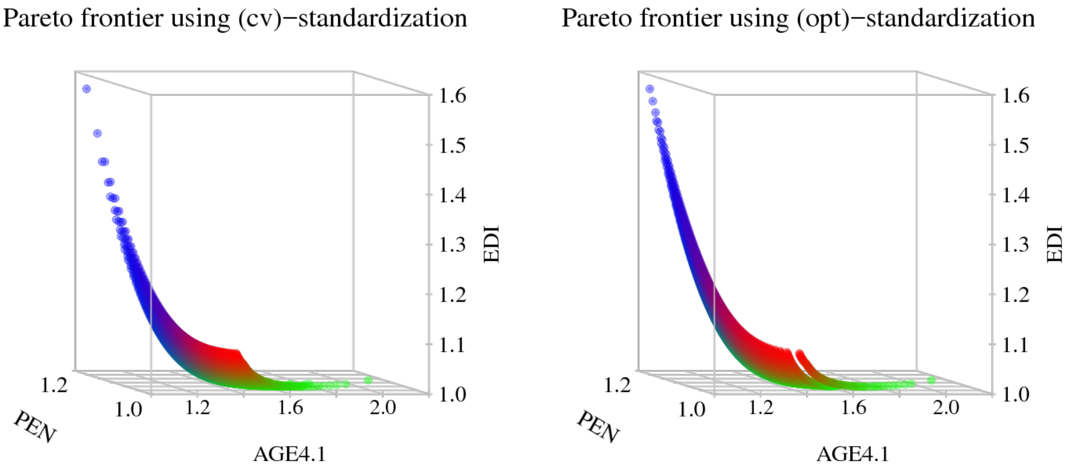


Figure 4.12: Pareto frontiers using the (cv)-standardization and (opt)-standardization.

The Pareto frontier is the set of all solutions of the underlying multi-criteria optimization problem (4.20), i.e. it contains the solution of the weighted sum scalarized problem (4.56) for all combinations of weights (cf. Theorem 3.3.15). The difference in solving the weighted sum scalarized MMDopt problem (4.56) using the (cv)- and the (opt)-standardization is the varying choice of the standardization factors γ_i $(i = 1, \dots, q_1)$, which can be interpreted as a rescaling of the weights (see Subsection 4.2.2). In other words, the only difference between (cv)- and the (opt)-standardization is the relation of another combination of weights to one specific Pareto optimal solution, which is illustrated in the heatmaps in Figure 4.11. In that regard, the shape of both Pareto frontiers in Figure 4.12 need to be equal for (cv)- and (opt)-standardization. Since the elements are related to other combinations of weights, the color scheme of both graphs differs from one another, and the dots at the upper tail of the Pareto frontier are located closer to each other when using the (opt)-standardization than when using the (cv)-standardization. This again indicates a more balanced allocation if the (opt)-standardization is used. Nevertheless, these observations are the consequence of the scaling resolution of the weights. Theoretically, the Pareto frontiers are exactly the same sets for both standardizations. Moreover, the *quasi-convexity* of the Pareto frontier (cf. Section 4.4) facilitates the approximation of the whole Pareto frontier by an interpolation of the computed dots, even for comparably high scaling resolutions of the weights. This is focused on in Subsection 4.6.3.

The evaluation of the whole Pareto frontier offers valuable support for the decision-maker to select the preferred solution among all efficient solutions. By using the weighted sum and by calculating the Pareto frontier, the decision is based on a higher level of reliable information. As the computation of the Pareto frontier requires the solution of many optimal allocation problems, it can only be realized in a practical time frame if efficient algorithms are used. The evaluations of Subsection 4.6.6 show that our algorithms are fast enough to facilitate this analysis of the Pareto frontier for `MMDopt`, even for large problem instances.

To compare `MMDopt` with other techniques of optimal allocation, the RRMSEs of the population total estimates are shown in Table 4.4, and the RRMSEs of the NUTS3-specific estimates for the auxiliaries and variables of interest are plotted in Figure 4.13. In this case, the `MMDopt` scenarios, computed with equal weights, are compared to both the *univariate* optimal allocations concerning auxiliaries PEN, EDI, and AGE4.1, and the *proportional* allocation. The univariate cases yield the most accurate estimates (underlined) for the optimized auxiliary variable as well as for the highly correlated variable of interest. However, a further consequence is found, namely that there are partly unacceptably inaccurate estimates for the other variables. By contrast, `MMDopt` leads to a compensated or balanced efficiency between all auxiliaries and variables of interest. This even holds for the population total estimates (Table 4.4), the NUTS3-specific estimates (Figure 4.13), and the other stratification levels (see Figure B.3). The efficiency of `MMDopt` is particularly demonstrated in Figure 4.13. `MMDopt` results in comparably small RRMSEs for

Table 4.4: RRMSE of population total estimates for selected allocation strategies (absolute values $\cdot 10^{-3}$).

	Auxiliaries			Variables of interest		
	PEN	EDI	AGE4.1	AGE7.7	INC.PP	AGE7.1
`MMDopt` (cv)	7.16	6.35	4.12	4.85	7.16	5.28
`MMDopt` (opt)	7.29	6.60	3.79	4.84	7.51	4.84
`uni_PEN`	7.07	6.50	4.64	4.78	7.41	5.91
`uni_EDI`	8.32	6.00	7.22	6.55	6.53	9.12
`uni_AGE4.1`	8.87	9.44	3.39	5.46	11.67	4.19
`prop`	7.83	6.04	6.26	5.95	6.66	7.86

the NUTS3-specific estimates of *all* considered variables. In looking at the three univariate cases, there is at least one variable with a comparatively inefficient estimate. This statement even holds for the proportional allocation. Thus, the efficiency with regard to MMDopt is more balanced compared to the other techniques, which improves the overall quality of the results of the survey.

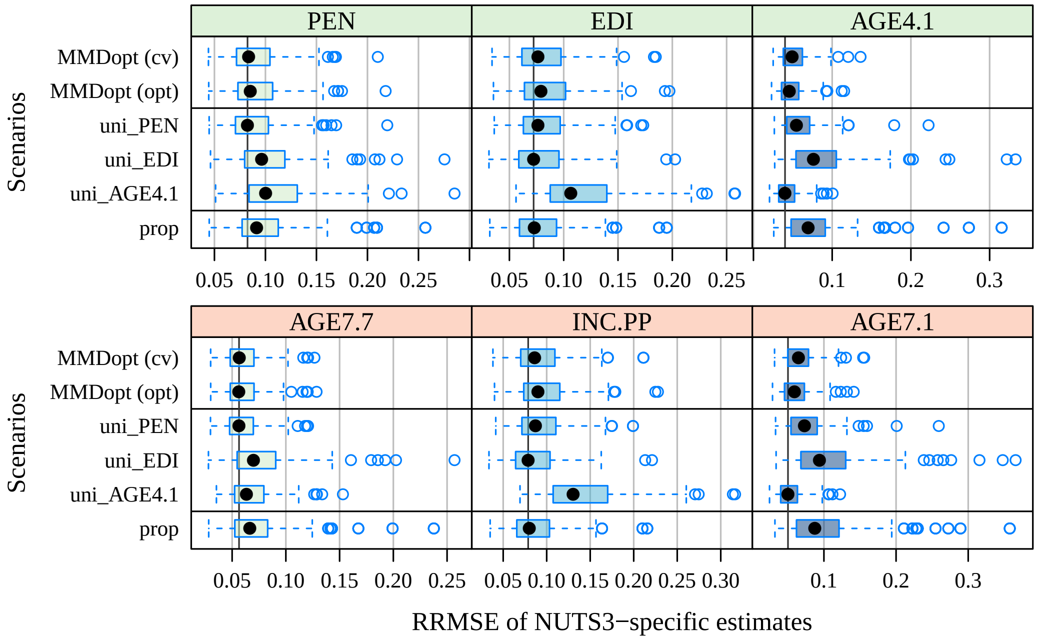


Figure 4.13: RRMSE of NUTS3-specific total estimates for selected allocation strategies (auxiliaries and variables of interest).

4.6.3 Decision-making strategy

Up until this point, we focused on the weighted sum scalarized MMDopt for generating the whole Pareto frontier. This is desirable from a theoretical point of view, since in this way each Pareto optimal solution is computed. Nevertheless, in practice decision-makers need to select their preferred solution. They may possibly make the decision on the visual basis of the heatmaps in Figures 4.10 and 4.11 as well as the Pareto frontiers in Figure 4.12. An alternative to this *graphical-based* decision-making strategy is the usage of a decision-making function, which is closely related to the scalarization techniques in Section 4.2.3. With the usage of such a function, the standardized variances are additively combined and then minimized. The most common decision-making functions are the p-norms (4.26) and the min-max method (4.27). Following the discussion in Section 4.2.3, the 1-norm as decision-making function for MMDopt is equivalent to the weighted sum scalarized MMDopt with equal weights for all variables. As addressed in Section 4.3, the weighted sum scalarized problem (and the 1-norm) can be solved with the SSN method, which is not applicable for other p-norms. This is due to the lack of separability of the objective function. However, the solution strategy of Section 4.4 based on the relation between the weighted sum scalarized problem (Ws.OP) and the p-norm scalarized problem (P.OP) can be applied to solve the `MMDopt` for p-norm scalarizations in analogy to the weighted sum scalarized `MMDopt` (see Figure 4.2). The corresponding `GTM` algorithm is given in Algorithm 4.

In the following, we compare the results of the four decision-making functions, namely 1-norm, 2-norm, 8-norm, and 64-norm, for the (cv)-standardization as well as for the alternative (opt)-standardization. As pointed out before, the 1-norm is equivalent to the weighted sum scalarization with equal weights.

In Figure 4.14, the relative increases of the RRMSEs of the population total estimates are shown for the auxiliary variables using `MMDopt` and different decision-making functions in relation to the corresponding optimal univariate allocations. Thus, a value of 1.0 corresponds to an estimate that is as accurate as the estimate based on the optimal univariate allocation. As every optimal univariate allocation is optimal with regard to the respective target variable, the RRMSEs of the multivariate allocation need to be higher than or equal to the RRMSEs of the univariate allocation. In the settings with (cv)-standardization, the error-increases are not well-balanced. Even for $p = 1$, the allocations in the (cv)-standardization case are extremely focused on PEN, whereas the other two variables are basically almost ignored. This is due to the different coefficients of variation tabulated in Table 4.2, where the relations

$$\mathrm{cv}^2(\hat{\tau}_{\mathsf{PEN}}^{\mathrm{StrRS}}) > \mathrm{cv}^2(\hat{\tau}_{\mathsf{EDI}}^{\mathrm{StrRS}}) \text{ and } \mathrm{cv}^2(\hat{\tau}_{\mathsf{PEN}}^{\mathrm{StrRS}}) > \mathrm{cv}^2(\hat{\tau}_{\mathsf{AGE4.1}}^{\mathrm{StrRS}})$$

can be observed. In the min-max case ($p \to \infty$), the error-increase of the estimate of variable PEN is even equal to zero, which means that the optimal multivariate allocation is equal to the optimal univariate allocation with respect to PEN in that case.

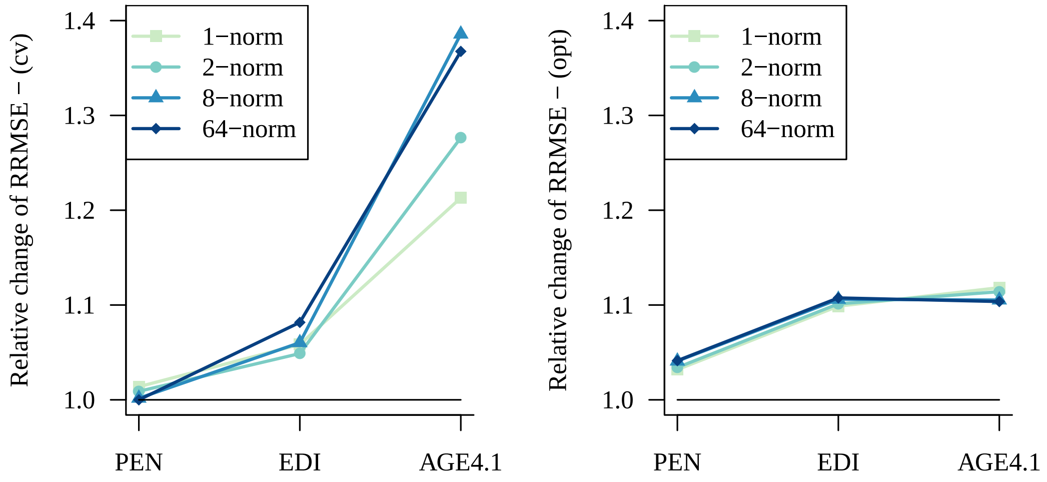


Figure 4.14: Relative increase of RRMSE of the population total estimates depending on the decision-making function (auxiliaries).

By contrast, we observe a well-balanced increase of the RRMSEs for the (opt)-standardization, since the p-norm of the relative change of the RRMSEs compared to the optimal univariate allocations is minimized. This results in a well-compensated allocation for all p-norms. In the case $p = \infty$, we obtain an almost equal increase of the RRMSE for EDI and AGE4.1, whereas the increase of PEN is slightly smaller. A similar structure can be observed for the variables of interest in Figure 4.15 due to the correlation structure. While the (opt)-standardization yields to balanced increases, the results for the (cv)-standardization are more heterogeneous.

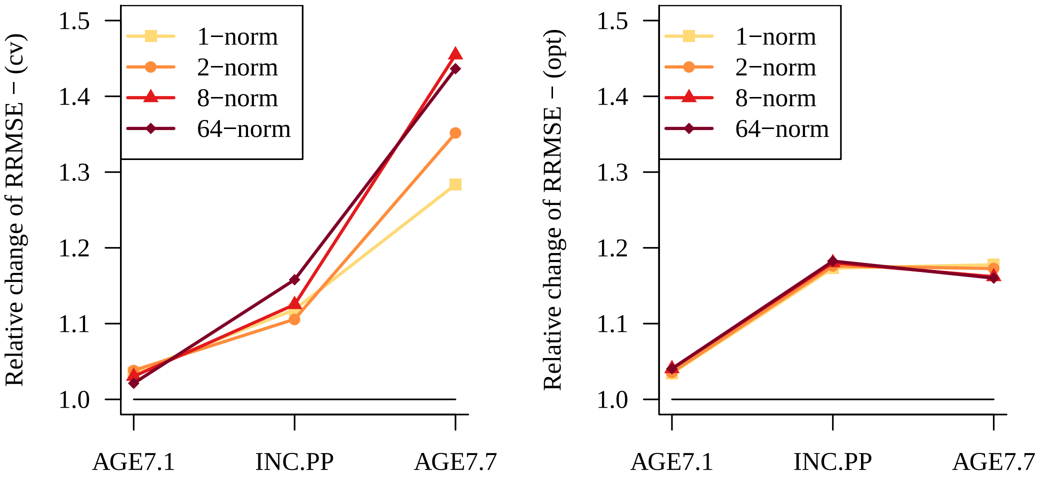


Figure 4.15: Relative increase of RRMSE of the population total estimates depending on the decision-making function (variables of interest).

In Figure 4.16, the same settings as in Figure 4.14 are plotted, but the RRMSEs of the SMP-specific total estimates are shown for the auxiliaries instead. As before, the errors are illustrated in relation to the errors resulting from the optimal univariate allocations. Dots located to the right of the vertical one-line correspond to the SMP-specific estimates with an increase of the respective RRMSE. Accordingly, dots to the left of the one-line correspond to estimates with a decrease. Again, the settings with the most balanced error-changes are these corresponding to the (opt)-standardization. In particular, higher values of p intensify the imbalance of the error-changes in the case of the (cv)-standardization, whereas the results of the (opt)-standardization are more robust. This is clearly observable for the variable AGE4.1, where high values of p while using the (cv)-standardization yield extremely high RRMSEs in some SMPs. To be precise, there are two SMPs with an error-increase of over 400% compared to the estimates under an optimal univariate allocation concerning AGE4.1. This highlights the imbalance of the (cv)-standardization. Although the *population-specific* RRMSEs in Figure 4.14 are equal to or higher than the univariate RRMSEs, the multivariate allocation also leads to error-decreases in several *SMPs*, which is illustrated by the points located to the left of the one-line in the boxplots. In general, however, the SMP-specific estimates behave similarly to the population total estimates. Results for other stratification levels are shown in Figure B.4.

To summarize, the application of a decision-making function to `MMDopt` enables the user to make a reasonable decision in favor of an application-specific optimal allocation. Figures 4.14, 4.15, and 4.16 show that the (cv)-standardization, and especially a scalarization with a larger p, can accentuate single variables, particularly the ones with estimates that have a comparably high coefficient of variation. This may be the best choice, if the goal is to achieve an allocation, in which the absolute maximal RRMSE is as small as possible. Certainly, this approach neglects the structure of specific variables; the estimation of both heterogeneous variables and these ones that have a skewed distribution may be more complicated. In contrast to this, the (opt)-standardization yields more compensated results, which may be preferable in cases in which no variable is considered as most important. In this case, not the *absolute maximal RRMSE*, but the *relative increase of the RRMSE* compared to the optimal univariate allocation is prioritized by the decision-maker. Thus, the optimization is not mainly focused on highly heterogeneous variables or these with a skewed distribution, but instead it considers each variable. Eventually, these reasons lead to the conclusion that the (opt)-standardization is more *compensated* and *balanced* compared to the (cv)-standardization. Finally, a larger p always places a higher focus on the bigger values of the objective function. Thus, if decision-makers need to focus more on high values (i.e. an absolute RRMSE with (cv) or a relative increases of the RRMSE with (opt)), they should choose a higher p.

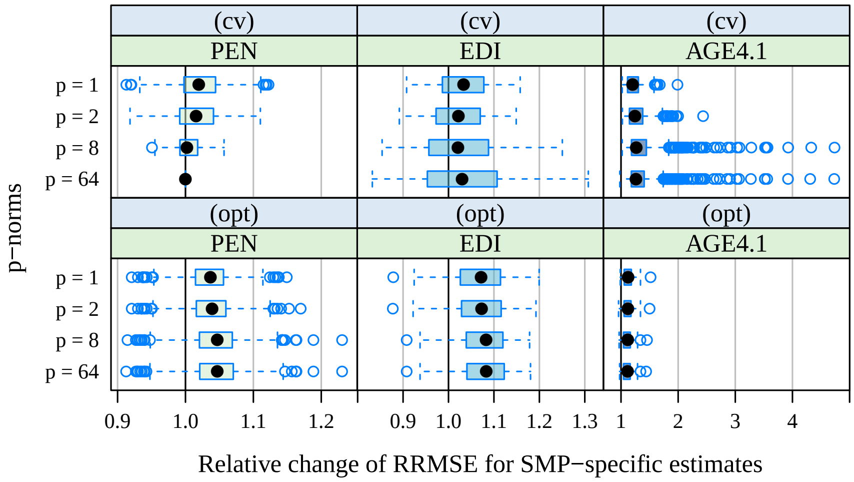


Figure 4.16: Relative change of RRMSE of the SMP-specific population total estimators depending on the decision-making function.

Finally, we remark that the development of the solution strategy of Section 4.4 using the GTM algorithm allows for solving a p-norm MMDopt in an appropriate time even for large problem instances. We take a closer look on the numerical performance of the GTM algorithm in Subsection 4.6.6.

4.6.4 Compromise allocation on different stratification levels

Generally, the focus of optimal allocation techniques lies on the minimization of the variances of the *population* estimates of the variables of interest. In this case, although a good quality for sub-populations estimates is often also required, the efficiency of *regional-specific* estimates is basically neglected. The reason is that the behavior of regional-specific estimators does not entirely coincide with the behavior of the population estimators. The quality of estimates for small regions is especially not considered by classical optimal allocation methods. This is exemplarily shown by the maps of the population in Figure 4.17. These maps illustrate the RRMSE of the NUTS3- and stratum-specific estimates based on MMDopt with (cv)-standardization, equal weights, and no assumptions for regional efficiency. Higher RRMSEs of auxiliary variable PEN for the regions are illustrated as dark blue areas of the map. Efficient estimates are assigned to the majority of NUTS3-regions as well as SMPs. Nevertheless, there are a few NUTS3-regions with a comparatively inefficient estimate and several strata with high estimation errors. Some stratum-specific RRMSEs are extreme high, since these strata are very small and highly heterogeneous. With regard to the stratification, these strata are mostly related to the class of the biggest household sizes. In looking at the maps, we note that the dataset used is synthetically generated. Thus, the analysis of the maps does not allow conclusions to be drawn about the regional estimation errors that would result from estimates based on real data.

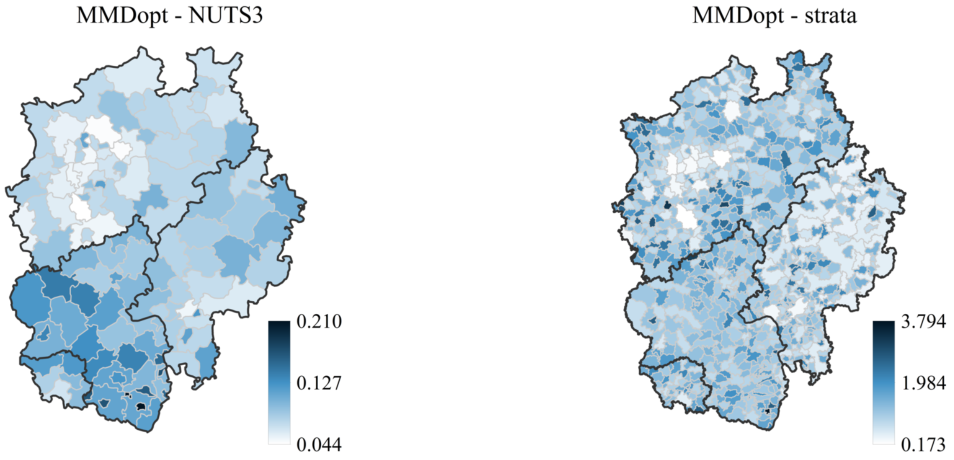


Figure 4.17: RRMSE of NUTS3- and stratum-specific estimates for PEN without restrictions for regional efficiency.

One solution approach is based on the first strategy of Dalenius (1953) (cf. Section 4.1), where the variances (of population total or area-specific estimates) are bounded from above and treated as constraints of the optimization problem, while the total sample size (or a cost function depending on the sample sizes) is minimized. This Cmin approach allows the integration of restrictions for regional efficiency, as among others addressed by Falorsi and Righi (2015) and Falorsi and Righi (2016). Nevertheless, MMDopt differentiates from the Cmin-approach, as the minimization of the variance for the population total estimates is a major key point of our analysis. As proved in Section 4.3, we are able to include restrictions for regional efficiency of the auxiliaries to the MMDopt method while maintaining the minimization of variances of the population estimates. As a consequence, better regional estimates for the variables of interest are expected due to their correlation with the auxiliary variables, and the variances of the population estimates are still minimized, in contrast to Cmin.

In the evaluation, a distinction is made between the following scenarios of MMDopt using the 1-norm scalarization. The maximal allowed RRMSEs are placed in parentheses after the scenarios for the respective stratification level and variable.

1. MMDopt: no restrictions for regional efficiency

2. `MMDopt+strata`: max. stratum-specific RRMSEs (PEN, AGE4.1: 1.5, EDI: 1.0)
3. `MMDopt+NUTS3`: maximal NUTS3-specific RRMSEs (PEN, EDI: 0.13)
4. `MMDopt+strata&NUTS3`: max. stratum-specific RRMSEs (PEN, AGE4.1: 1.5, EDI: 1.0) and maximal NUTS3-specific RRMSEs (PEN, EDI: 0.13)
5. `Cmin` (Benchmark): minimize n_s with *properly* chosen RRMSE restrictions

In Table 4.5, the increase of the RRMSEs of the population estimates for the previously defined settings is tabulated in relation to the optimal univariate allocations. Since the assumptions for regional efficiency are included, the feasible set of the optimization problem shrinks. Thus, the RRMSEs of the population estimates increase if the number of regional restrictions increases. In a certain manner, a compromise has to be made for enforcing efficient regional estimates. In the strictest scenario, the error-increase compared to the standard `MMDopt` is about 5%. This increase may be acceptable, especially since intolerable errors of regional estimates are omitted coincidently on various stratification levels as shown in Figure 4.18 for the auxiliaries. The maximum permitted errors are plotted as vertical red lines in the boxplots. Generally, all NUTS3- as well as stratum-specific estimates comply with the predefined borders in the respective scenarios. Concurrently, the RRMSEs of the other areas slightly increases, as a certain amount of the total sample size is shifted to the areas with intolerable errors. To conclude, inefficient outliers can be omitted using the restrictions for regional efficiency. In comparing the results of the variables of interest with the auxiliaries, a similar efficiency-increase of regional estimates can be observed (see Figure B.5). The results using `Cmin` also reveal a compliance with the predefined restrictions for regional efficiency (see Figure 4.18). Nevertheless, the medians of the stratum- and NUTS3-specific RRMSEs are significantly higher than using `MMDopt`, and the RRMSEs of the population total estimates are disproportionate high, as their variances are not optimized in `Cmin` (see Table 4.5). These observation are major characteristics of `Cmin`, which mostly results in estimates that both comply with predefined errors and are not optimized.

Table 4.5: Increase of the RRMSEs of population total estimates for selected allocation strategies with restrictions for regional efficiency (1.00 corresponds to the same RRMSE as in the optimal univariate allocation).

		Auxiliaries			Variables of interest		
		PEN	EDI	AGE4.1	AGE7.7	INC.PP	AGE7.1
MMDopt	(cv)	1.01	1.06	1.21	1.04	1.12	1.28
	(opt)	1.03	1.10	1.12	1.03	1.17	1.18
MMDopt+ strata	(cv)	1.03	1.06	1.23	1.05	1.45	1.30
	(opt)	1.04	1.10	1.14	1.05	1.17	1.20
MMDopt+ NUTS3	(cv)	1.02	1.06	1.22	1.04	1.12	1.29
	(opt)	1.03	1.10	1.12	1.04	1.08	1.18
MMDopt+ strata&NUTS3	(cv)	1.06	1.09	1.26	1.08	1.14	1.33
	(opt)	1.07	1.12	1.18	1.08	1.19	1.24
Cmin		1.55	1.35	2.22	1.70	1.29	2.32

The effect of increasing regional efficiency is also shown in the maps in Figure 4.19 in conformity with the maps in Figure 4.17. The dark blue shades indicating regions with high estimation errors (NUTS3-regions as well as strata) have disappeared. In other words, a minimal quality for regional-specific estimates is guaranteed. Concurrently, the quality of the estimates for the other regions does not considerably suffer from the included restrictions for regional efficiency. Although the observation is restricted to `MMDopt` with $p = 1$ and equal weights, the inclusion of restrictions for regional efficiency is also possible for `MMDopt` with other p-norms as decision-making functions (Section 4.6.3) as well as for the computation of the whole Pareto frontier (Section 4.6.2).

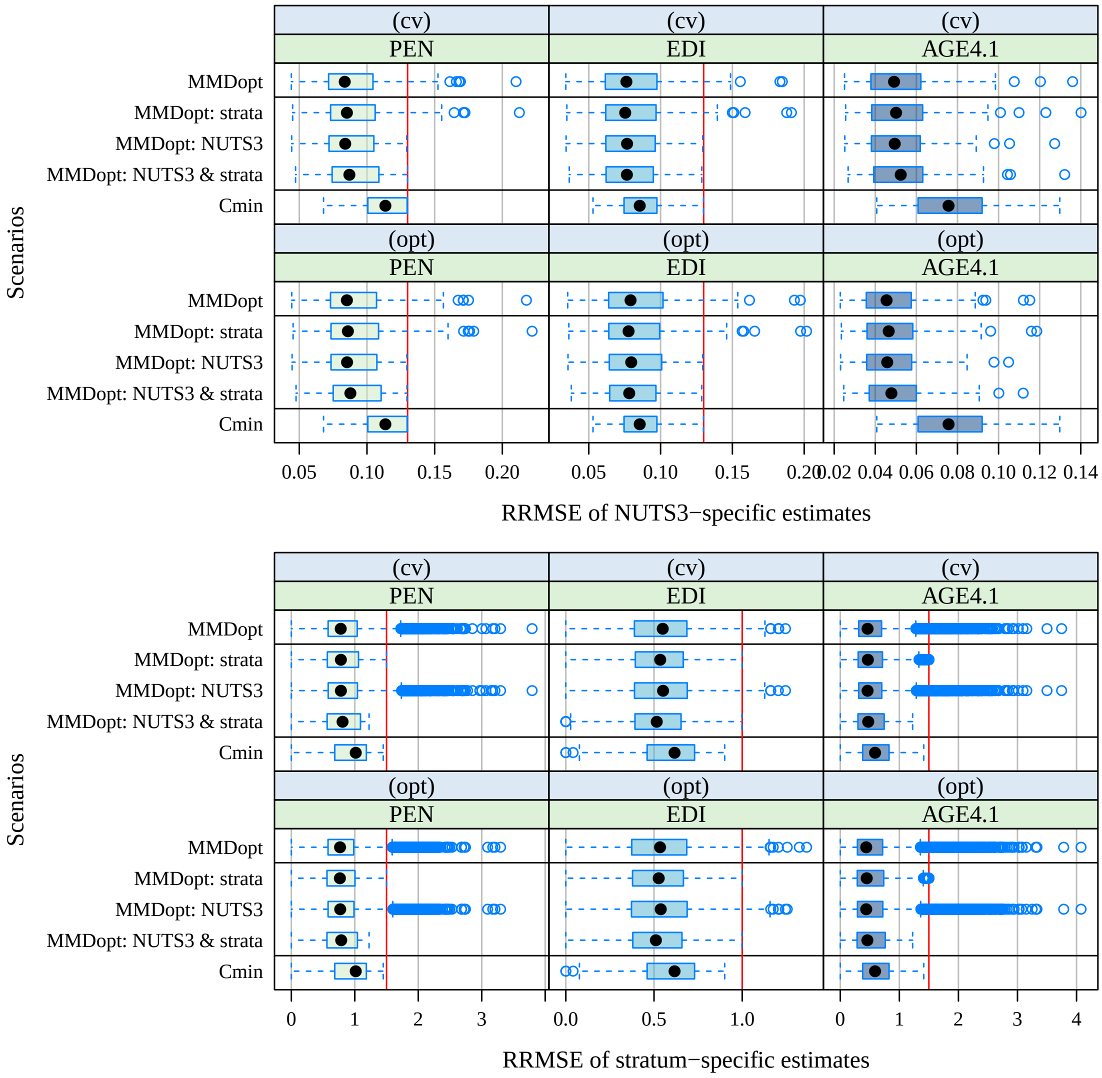


Figure 4.18: RRMSE for the NUTS3- and stratum-specific estimates with restrictions for regional efficiency.

To conclude, the possibility of including restrictions for regional efficiency can be of particular interest for official statistics. Usually, requirements for a minimal quality of estimates need to be complied with. Using MMDopt with restrictions for regional efficiency provide the technical opportunity to apply an optimal multivariate allocation under these circumstances.

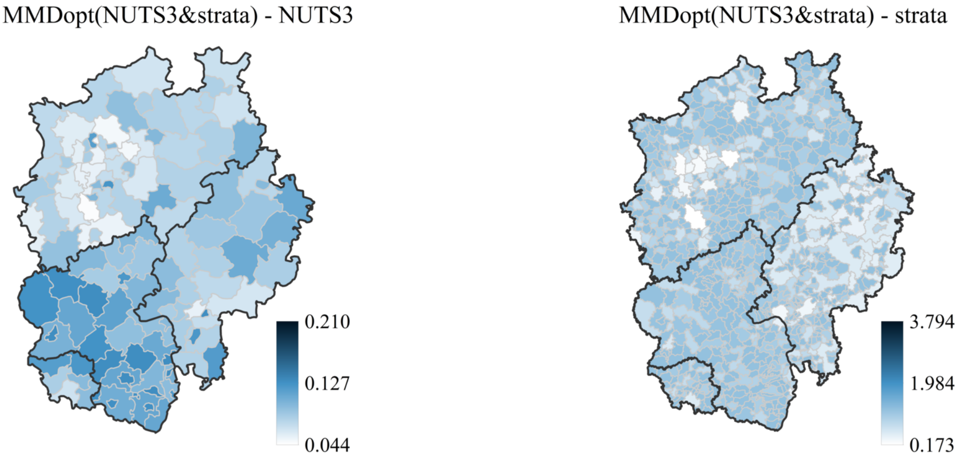


Figure 4.19: RRMSE of NUTS3- and stratum-specific estimates for PEN with restrictions for regional efficiency.

As mentioned, the inclusion of the restrictions for regional efficiency to MMDopt results in a shift of a certain amount of the sample size is to those areas for which the restrictions are active, i.e. to those areas with high RRMSEs. A comparison of the optimal stratum-specific sample sizes for some selected scenarios is shown by means of scatterplots in Figure 4.20. The stratum-specific sample-sizes n_h for the optimal univariate and Cmin allocations (columns) are contrasted with MMDopt and MMDopt.reg (i.e. MMDopt+strata&NUTS3; cf. Table 4.5) using the (opt)-standardization in order to compare both the univariate cases with MMDopt and MMDopt with MMDopt.reg. The results compared to MMDopt are plotted in row 1 and the results

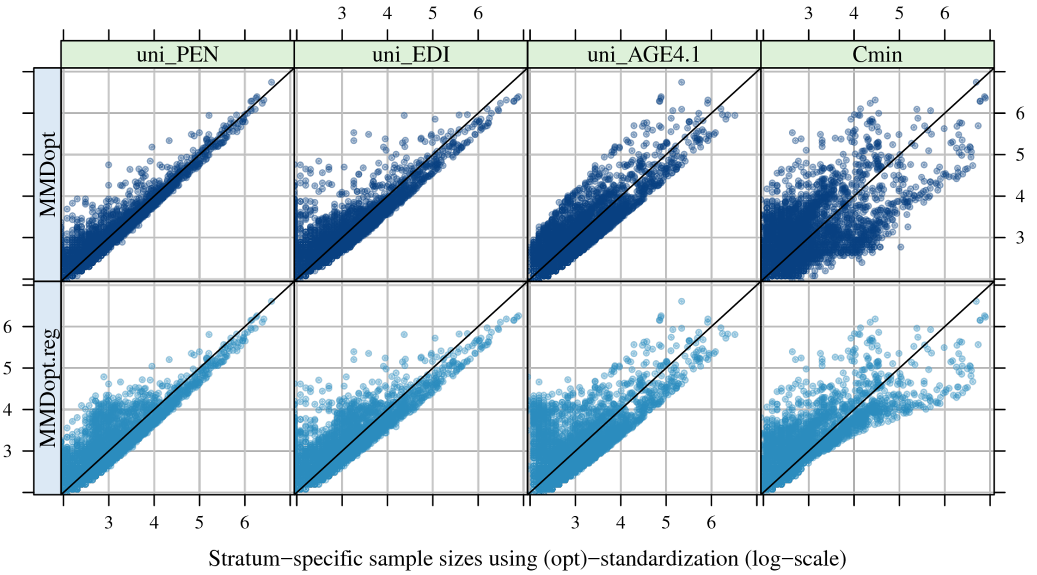


Figure 4.20: Stratum-specific sample sizes depending on allocation method in log-scale.

for `MMDopt.reg` in row 2. The stratum-specific sample sizes are plotted in log-scale, since it acquires a higher visibility in the comparison of the small stratum-specific sample sizes. If dots are located below the line of the bisector, the respective stratum-specific sample size n_h is lower for `MMDopt` or `MMDopt.reg` compared to the benchmark methods. The structure of the optimal allocations is not completely different in general. However, it is obvious that forcing the regional efficiency (row 2) yields particularly small sample sizes to be higher than without regional restrictions. This is illustrated by the higher number of points with small sample sizes (lower than $\log(n_h) \approx 4.0$, i.e. $n_h \approx 50$) above the bisector line in the `MMDopt.reg` case. To compensate this, the stratum-specific sample sizes of big strata slightly decrease. This is illustrated by points below the bisector line in case of high stratum-specific sample sizes. As a result, a moderate error-increase is observed in the big strata (in which the quality is already good enough), but instead significantly better estimates are ensured in the small strata.

In comparing `Cmin` with `MMDopt` and `MMDopt.reg`, major differences between the stratum-specific sample sizes can be observed, but these differences do not follow a specific pattern. This is due to the fact that the stratum-specific sample sizes using `Cmin` are completely determined by a predefined minimum quality in *each* stratum. Thus, the stratum-specific sample sizes are just high enough to ensure the compliance with the minimum quality. However, there is no compensatory optimization between the strata and no minimization of the variance of the population total estimator, which can also be seen in the last row of Table 4.5.

4.6.5 Robustness and sensitivity

In this subsection, the robustness and sensitivity of `MMDopt` and the resulting regional and population estimates depending on the input data are analyzed in relation to the optimal univariate allocation. As mentioned, the input data can be chosen in different ways. Firstly, *good* proxies for the stratum-specific variances may be gained by evaluating previous surveys or obtainable register data. In this case, these proxies may be affected by certain trends, disruptive effects (such as natural disasters or economic crises), or some random noise in the data. To handle this, we investigate the results of `MMDopt` for the four chosen scenarios `sen1` to `sen4` of the input data tabulated in Table 4.6. These scenarios contain SMP-specific trends (constant factor for all units of a SMP) as well as an individual normal-distributed random noise (random factor for each unit). These scenarios are computed separately for each auxiliary variable. Secondly,

Table 4.6: Scenarios of input data for sensitivity analysis.

	Uniform-distributed SMP-specific trend	Individual normal-distributed noise
`org`	unchanged original data	
`sen1`	-5% up to $+10\%$	sd $= 0.15$
`sen2`	-15% up to $+25\%$	sd $= 0.15$
`sen3`	-5% up to $+10\%$	sd $= 0.4$
`sen4`	-15% up to $+25\%$	sd $= 0.4$

the input data can be gained from highly correlated and known auxiliary variables (as done in Subsections 4.6.2, 4.6.3 and 4.6.4). From this, the question arises regarding how the height of the correlation between auxiliaries and variables of interest affects the estimations. Moreover, the relation between the correlation of the variables of interest and the correlation of the resulting stratum-specific sample sizes is of interest. We first focus on the evaluation of the results based on the five scenarios of the auxiliary variables. Scatterplots illustrating the changes of the stratum-specific totals $\tau_{y_i h}$ and the stratum-specific variances S^2_{ih} compared to the original data are shown in Figure 4.21. As expected, the changes of the stratum-specific totals (row 1) compared to the original totals are greater for the scenarios with a higher SMP-specific trend (sen2 and sen4), whereas the differences of the stratum-specific variances (row 2) are higher for the scenarios with a more distinctive noise on the individual level (sen1 and sen3).

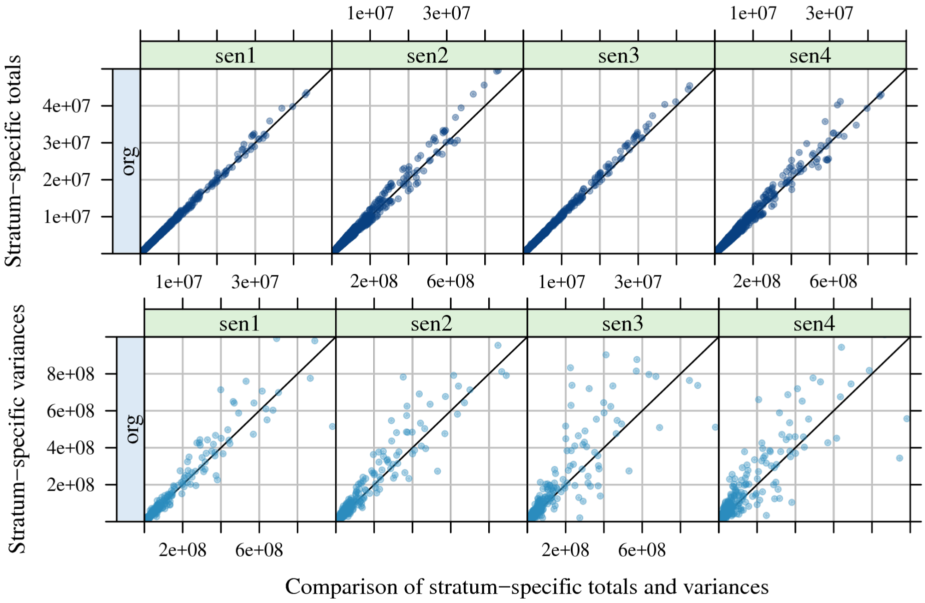


Figure 4.21: Comparison of stratum-specific totals and variances depending on the sensitivity of the data.

The effect on the estimates is shown in Figure 4.22, in which the relative changes of the RRMSEs of the SMP-specific estimates for the four scenarios in comparison to the scenario with the original data are plotted. Thus, dots to the left of the red one-line represent SMP-regions with a decreased RRMSE compared to the case with original data; likewise, dots to the right represent SMP-regions with an increased RRMSE. The change of the population estimates is plotted as a vertical blue line within each boxplot. The results of the optimal univariate allocations (uni_[var]) are compared with MMDopt and MMDopt.reg using the scenario with equal weights and the (cv)-standardization. Scenarios sen2 and sen4 comprise higher differences in the SMP-specific trends, and they contain greater changes in the efficiency compared to sen1 and sen3 for using uni_[var] and MMDopt. Moreover, MMDopt is more robust compared to the optimal univariate allocations, as the boxplots of MMDopt are tighter than these ones of the optimal univariate allocation. An explanation may be the usage of more than one auxiliary variable in MMDopt. The changes in the data may be balanced out over the various auxiliaries, which is not possible in the case of optimal univariate allocations as only one variable in considered.

If restrictions for regional efficiency are added (MMDopt.reg), the level of robustness as well as the efficiency of the estimates decreases significantly. Since the restrictions are based on incorrect data, the feasible set is skewed, which results in a loss of efficiency. This efficiency loss is particularly higher if the noise on individual level is higher (sen3 and sen4).

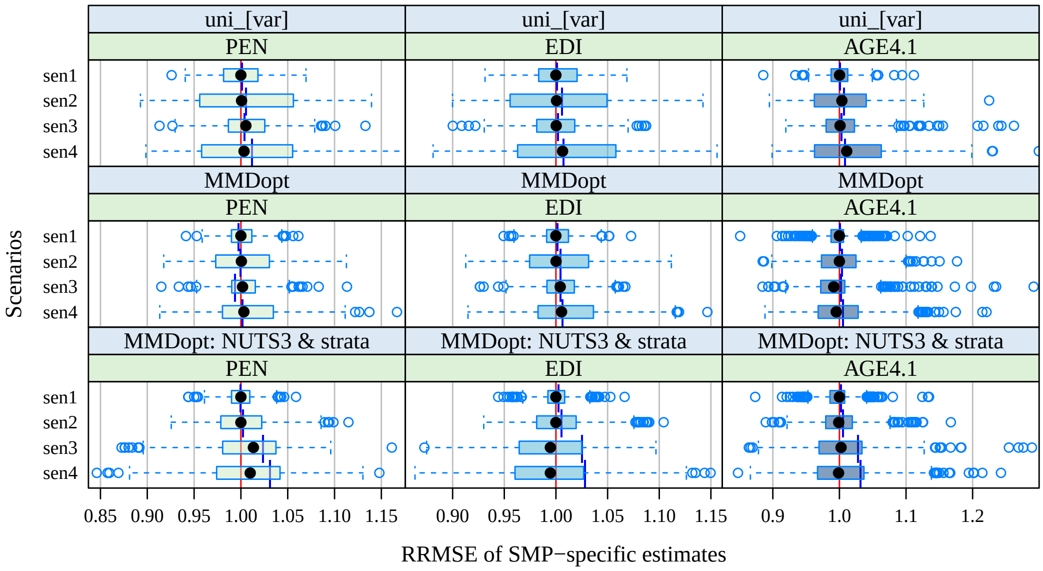


Figure 4.22: Relative change of RRMSE of SMP-specific total estimates depending on the sensitivity of the input data with (cv)-standardization.

In general, the MMDopt behaves better in the case of incorrect data than the optimal univariate allocations, as incorrectness can partially be balanced out by the various variables. The robustness of MMDopt.reg depends on the specific definition of the restrictions, but it is significantly poorer than the robustness of MMDopt. The sensitivity results under (cv)- and (opt)-standardization as well as under different decision-making functions are comparable with one another.

4.6.6 Algorithmic performance

In this subsection, the performance of the SSN, the SSN.reg, and the GTM algorithm is analyzed. All the numerical results are computed in the programming language R on a desktop PC with an Intel Core i7-6700 CPU at 3.40GHz × 8 and an internal memory of 32 GB.

Performance of SSN and SSN.reg

Münnich et al. (2012c) and Friedrich et al. (2015) showed that the fixed-point iteration or SSN method for the continuous allocation problem as well as the Greedy algorithm for the integer problem have huge advantages in computing time compared to the R package nloptr (cf. Ypma et al., 2017). This package provides an R interface to the open source library NLopt for nonlinear optimization (cf. Johnson, 2018). However, as pointed out in Section 4.3, the separability of the objective function is mandatory for these algorithms. Thus, these algorithms can only be applied using a weighted sum or 1-norm as decision-making function. Due to the advantages in the computing time, an alternative approach for $p \neq 1$ is suggested in Section 4.4 based on the `GTM` algorithm. The numerical performance of this approach is also analyzed.

The following results are based on the scenarios of Subsection 4.6.2 with a weighted sum setting with equal weights $w = (1/3, 1/3, 1/3)$ and (opt)-standardization. Neither the choice of the weights nor the selection of the standardization technique changes the numerical performance of the algorithms significantly. The initial point for the continuous solvers is calculated with the aid of the respective mean values of the stratum-specific sample sizes according to the three separate optimal univariate allocations. First, computing time and the number of iterations are compared in Figure 4.23 for the following three cases:

- `SSN` depending on the dimension of the allocation problem H (column 1),
- `SSN.reg` depending on the dimension of the allocation problem H (with $q_3 = 15$ constraints for regional efficiency; column 2),
- `SSN.reg` depending on the number of constraints for regional efficiency q_3 (with $H = 6272$); to three cases from `weak` (weak restrictions) up to `hard` (hard restrictions; column 3).

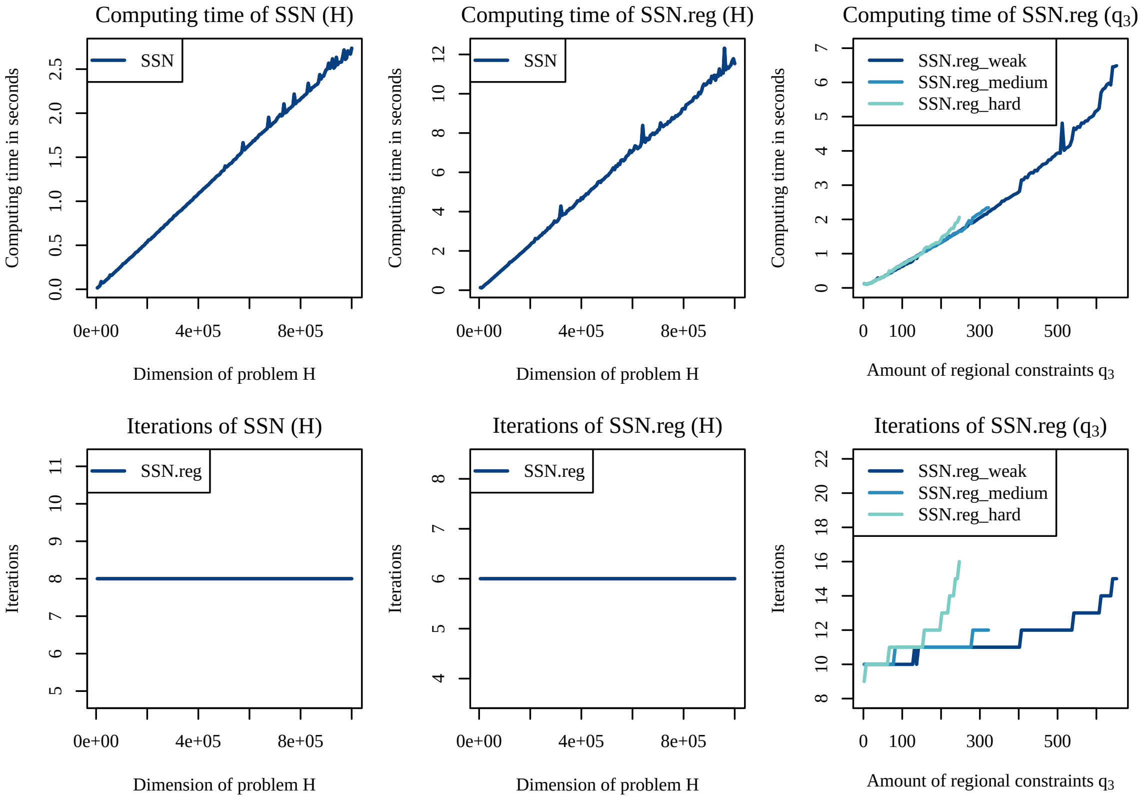


Figure 4.23: Computing time and number of iterations of `SSN` and `SSN.reg` depending on number of strata H and number of restrictions for regional efficiency q_3.

The distinction between the cases of weak, middle, and hard restrictions differs in the strictness of the predefined maximum regional errors. Thus, the case with hard restrictions results in the smallest feasible set. The computing time of the three cases and their number of iterations are plotted in the three columns of Figure 4.23. Generally, the computing time increases exponentially with an increase of the dimension of an optimization problem if standard solver are applied. Using the SSN method, a linear dependency between problem dimension and computing time is observed. The cause for this is the transformation of the system of the optimality conditions to a nonlinear system of equations (see Section 4.3). This results in a fundamental reduction of running time. In the left column, a computing time of about 2.5 seconds for 8 iterations for a dimension of $H = 1\,000\,000$ is indicated. In comparison, the standard solver nloptr has a running time of 10 seconds (requiring approximately $1\,500$ iterations) for a dimension of $H = 5\,000$. For significantly higher values of H, an extreme high computational burden using nloptr can be observed due to the exponential increase depending on H. The computing time of SSN for $H = 5\,000$ is about 0.012 seconds. In the second column, the computing time and the iterations are shown for SSN.reg. Although the running time per dimension is larger compared to SSN, the linear dependency on H also holds for this case. In the right column, the running time and number of iterations are plotted in dependency of q_3 (i.e. the number of nonlinear inequality constraints of problem (4.20)). Since the dimension of the nonlinear system of equations increases in q_3, the computational burden increases exponentially. Additionally, the number of iterations is gradually raised with an increase of q_3. If q_3 exceeds a critical value, the solver fails, as the restrictions are impossible to be fulfilled (i.e. the feasible set in empty), given that the stricter the restrictions are, the smaller the critical number.

The performance of the SSN algorithm for one specific example is shown in Table 4.7 for each iteration. By analyzing the residuals tabulated in column 2 of the table, a locally quadratic convergence rate of SSN can be observed, which is analytically proved for the SSN algorithm in Theorem 3.2.8. Within the first few iterations, the Armijo step-size rule (see Algorithm 2) reduces the step-size significantly. This accentuates the necessity of the step-size rule in order to achieve convergence.

Table 4.7: Performance of the semismooth Newton algorithm ($q_3 = 0$).

Iterations	Residual	Step-size	Lagrangian mult.	Objective
k	$\|\Phi(\lambda^k)\|$	α_k	λ^k	$f(n(\lambda^k))$
0	$8.51 \cdot 10^4$	0.0625	$3.360 \cdot 10^{-2}$	$16.57 \cdot 10^{-5}$
1	$7.49 \cdot 10^4$	0.1250	$1.372 \cdot 10^{-2}$	$11.45 \cdot 10^{-5}$
2	$5.64 \cdot 10^4$	0.2500	$6.402 \cdot 10^{-3}$	$7.49 \cdot 10^{-5}$
3	$3.16 \cdot 10^4$	0.5000	$3.864 \cdot 10^{-3}$	$5.16 \cdot 10^{-5}$
4	$8.56 \cdot 10^3$	1.0000	$3.190 \cdot 10^{-3}$	$4.03 \cdot 10^{-5}$
5	$1.25 \cdot 10^3$	1.0000	$3.262 \cdot 10^{-3}$	$3.78 \cdot 10^{-5}$
6	$2.08 \cdot 10^1$	1.0000	$3.263 \cdot 10^{-3}$	$3.72 \cdot 10^{-5}$
7	$5.89 \cdot 10^{-3}$	1.0000	$3.263 \cdot 10^{-3}$	$3.72 \cdot 10^{-5}$
8	$2.91 \cdot 10^{-10}$	1.0000	$3.263 \cdot 10^{-3}$	$3.72 \cdot 10^{-5}$

Without the control of the step-size, the algorithm diverges within the first iteration (i.e. $n_h(\lambda)$ reaches the boxes m_h or M_h for all $h = 1, \ldots, H$). For `SSN.reg` especially, a high sensibility of the convergence depending on both the parameter of the step-size rule and the initial points can be observed. If these input parameters are not chosen properly, the algorithm fails after a few iterations due to a non-solvable Newton step in iteration k. This implies that the element of the B-subdifferential of $\Phi(\lambda^k, \beta^k)$ (see (4.50)) is non-regular, which usually occurs if a high percentage of components of $n(\lambda^k, \beta^k)$ reaches the box-constraints, and many inequality constraints are not satisfied in iteration k.

Performance of `GTM`

Subsequently, the ability of the `GTM` algorithm to solve `MMDopt` problems with decision-functions for $p \neq 1$ is analyzed. The setting of the following example is equal to the setting of Subsection 4.6.3 for $p = 8$ and (opt)-standardization ($H = 6\,272$). The initial combination of weights is chosen to be $w^0 = c(0.1, 0.1, 0.8)$. The `GTM` algorithm converges after 26 iterations. Since the solution of `SSN` is repeated three times in each iteration of `GTM`, the `SSN` algorithm has to be applied 78 times, which results in a total computing time that is significantly lower than one second. Thus, a $6\,272$-dimensional restricted optimization problem with a non-separable objective function can be solved within one second, which highlights the opportunities of the `GTM` algorithm. The course of the iterations (denoted by It.) is plotted in Figure 4.24, where the color of the dots represents the value of the objective function G^p. For a better visibility, the other combinations of weights with a resolution of 0.1 are plotted transparently in the background. Nevertheless, the computation of these dots is not necessary for the application of `GTM`.

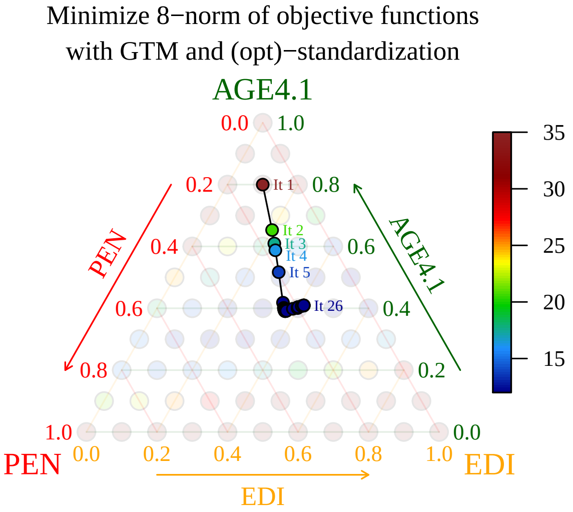


Figure 4.24: Example for the iteration of the `GTM` algorithm for $p = 8$ and (opt)-standardization.

In Table 4.8, the residuals (i.e. the norm of approximated gradient), the distances between weights (iterates) of two consecutive iterations, and the objective values G^p are tabulated for each iteration. The residual converges to a predefined tolerance close to zero. In general, the distance between the iterates decreases with increasing iterations. In more detailed examinations of the performance, the number of iterations of GTM strongly depends on both the initial point w^0. Although the number of iterations may be significantly higher than in this example, a convergence of GTM is still observed in all tested scenarios.

Table 4.8: Performance of the GTM algorithm for $p = 8$ and (opt)-standardization.

Iterations	Residual	Distance between	Objective
k	$\|s^k\|$	the iterates	$G^p(n(w^k))$
1	8.9830	0.3185	50.6469
2	4.7400	0.0935	20.7278
3	4.0665	0.0469	17.5412
4	3.7636	0.1520	16.3498
5	2.8261	0.2115	13.6143
⋮	⋮	⋮	⋮
22	0.0122	0.0140	11.9883
23	0.0321	0.0012	11.9885
24	0.0069	0.0006	11.9878
25	0.0049	0.0002	11.9875
26	0.0020		11.9875

4.6.7 Issues and limitations

In this subsection, alternative settings, occurring issues, and limitations of MMDopt and MMDopt.reg as well as the contained algorithms SSN, SSN.reg, and GTM are focused on. In this way, critical scenarios, additional capabilities, and unrealizable settings are discussed.

Application using other auxiliary variables

Up until now, MMDopt has been applied with $q_1 = 3$ auxiliary variables. Nevertheless, other choices of q_1 are also possible. In particular, MMDopt with $q_1 = 1$ is equivalent to the box-constrained optimal *univariate* allocation developed in Gabler et al. (2012) and Münnich et al. (2012c). For two auxiliary variables, the dimension of the Pareto frontier shrinks from a subset of a three-dimensional cube to a subset of a two-dimensional space. Thus, the corresponding heatmap is reduced to a straight curve as exemplarily shown in Figure 4.25. A value of $q_1 > 3$ is also permitted, but in this case the auxiliaries have to be chosen with care, as the influence of a single variable on the allocation decreases and the probability of undesired side effects as well as superimpositions increases with increasing q_1. Moreover, a graphical illustration of the solution via the Pareto frontier or the heatmap is impossible for $q_1 > 3$. If the auxiliary variables are highly correlated, the results of `MMDopt` will not be significantly different from the results of the optimal univariate allocations with box-constraints. Hence, the advantages of `MMDopt` increase with highly heterogeneous and especially with low correlated auxiliary variables.

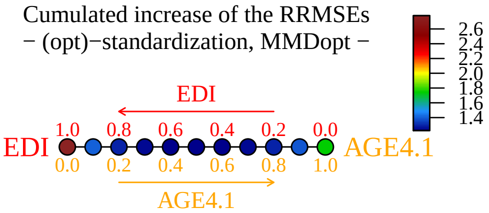


Figure 4.25: Relative cumulated increase of the RRMSEs of the population estimates under (opt)-standardization for $q_1 = 2$.

Solution as integer optimization problem

So far, we have ignored the requirement that the calculated stratum-specific sample sizes in an optimal (univariate or multivariate) allocation problem have to be in the set of non-negative integers for almost all application problems. This is due to the fact that a fraction of a person or household cannot be drawn in a sample. In general, the solution of the allocation problems with `MMDopt` as presented in the sections before is not an integer but a fractional number. In real world applications (and also in the previous application), this problem is commonly solved by a rounding strategy in the post-processing of the results. However, a rounded solution obtained this way is in general *not* an optimal solution in the set of all integral solutions. Therefore, we refer to Friedrich et al. (2018), where the `MMDopt` solved via the continuous solver `SSN` is compared with an algorithm for the computation of the globally optimal solution for integers.

As in the continuous case, the *integer* optimal multivariate allocation problem is algorithmically tractable whenever the objective function is separable and convex. The problem reduces to a single-objective optimization problem, and algorithms developed for the univariate allocation problem can be applied directly. Nevertheless, the integer solver is only applicable for the weighted sum (or $p = 1$) case with $q_3 = 0$, i.e. for the case without additional restrictions for regional efficiency.

Friedrich et al. (2015) presented three algorithms for the problem, which use the fact that the minimization of a separable and convex function is polynomially solvable for integer variables if the feasible set is a *polymatroid* (i.e. a convex polytope with strong combinatorial properties). An exhaustive discussion of the mathematical background is given in Friedrich (2016). The algorithms are based on methods referred to as *Greedy* strategies and converge to the globally optimal integer solution.

In the case of convex objective functions that are not necessarily separable ($p \neq 1$) however, the fast Greedy algorithm does not find the optimal solution. In addition, the convergence of `GTM` is also not provable using integer optimization techniques instead of `SSN`. Nevertheless, it is possible to solve these more general problems with the help of a reformulation as *linear* integer problems (Hochbaum, 1995). A reformulation of this type has been solved with the commercial software *FICO Xpress Optimization Suite* in Friedrich et al. (2015) with the result that computation times worsen significantly, i.e. it takes many hours instead of seconds. Therefore, we do not solve the integer version of the non-separable problems in this application study.

Friedrich et al. (2018) presented a detailed comparison between the rounded optimal continuous solution of the optimal multivariate allocation problem, computed by SSN, and the optimal integer solution, computed by the Greedy method. Although the differences in the solutions are measurable, they are generally extremely small so that the influence on the estimation can be ignored, especially for stratum-specific sample sizes significantly greater than zero.

Lagrange-multipliers and infeasibility of problem

Since the underlying optimization problems of MMDopt and particularly MMDopt.reg are restricted optimization problems, the convergence of the algorithms can only be proved under the assumption of a non-empty feasible set. Coincidently, these restrictions are based on geographical circumstances (e.g. regional stratification or stratum sizes) and politically or legally defined restrictions on the quality of the estimates and the total sample sizes. Hence, the assumption of a non-empty set can generally not be proved in advance. If MMDopt is applied to a problem with an empty feasible set, the SSN algorithm breaks down in iteration k when the element H^k of the B-subdifferential of $\Phi(n^k)$ is singular. In this case, the solutions in iteration k of SSN (which are the Lagrangian multipliers) reach unusually high values, which is another indicator for the non-feasibility. Consequently, the user needs to weaken some of the restrictions and restart the algorithm. In general, this process could also be automated by the software automatically adjusting the restrictions and restarting the algorithm.

4.7 Summary and discussion

The MMDopt method determines optimal allocations while considering several requirements such as conflicting variables of interest, various stratification levels, cost restrictions, restrictions for regional efficiency, and the control of sampling fractions. In order to solve such multivariate allocation problems, we proposed several scalarization and standardization techniques in Section 4.2. Whereas the scalarization reflects the decision-making function when evaluating conflicting goals, the standardization of the variances yields a rescaling of the variables fostering comparability. Furthermore, we proposed a strategy to compute the entire Pareto frontier as the set of all Pareto optimal solutions of the multi-criteria problem with the SSN algorithm using a weighted sum scalarization in Section 4.3, which was treated in a simpler case in Friedrich et al. (2018). The major benefit is the possibility of an a posteriori choice for a weighting scheme of the variables of interest. Thus, the decision-maker is able to incorporate additional information to achieve an application-specific *optimal* allocation. As a further advantage, it is not necessary to a priori assess the conflicting goals or rank the variables of interest. If the decision-maker needs to determine one specific solution, a procedure based on the GTM method has been proposed in Section 4.4. In that regard, a p-norm or a min-max approach can be utilized to determine the application-specific *optimal* allocation.

We computed solutions for instances of the MMDopt problem using the RIFOSS household dataset (see Section 2.6). Within the application study in Section 4.6, the methods and the algorithms were presented comparatively, their advantages were underlined, and recommendations for their practical use were given. In contrast to standard solvers, using the separability and convexity of the given problem yields a substantial increase in the numerical performance, which enables the calculation of the Pareto frontier in high resolution. Moreover, we observed considerable differences in (area- and stratum-specific) estimation errors and stratum-specific sample sizes when varying the weighting schemes. Therefore, we can underline the importance of the chosen scalarization, standardization, and weighting in optimal multivariate allocation.

To summarize, the MMDopt method facilitates the control of the quality of regional estimates for various variables and various stratification levels accompanied by the minimization of the (scalarized and standardized) errors of the population estimates for several auxiliary variables. The interaction of the allocation of the sample size for contradictory auxiliaries, the minimization of variances of the population estimates, the control of regional variances, and the monitoring of sampling fractions (by means of the box-constraints) conglomerate all the key points mentioned in the introduction and in the beginning of this chapter. If a high correlation between the variables of interest and the auxiliaries is given, the accuracy gain of the estimates of the auxiliary variables can be observed similarly for the variables of interest.

Chapter 5

A Generalized Calibration Method

5.1 Motivation and issues

Motivation

Calibration methods adjust design weights in order to incorporate auxiliary data into the estimation process (cf. Statistics Canada, 2003). This may result in both an increase of the accuracy of the calibration estimates compared to the HT estimates and also coherent estimates (cf. Merkouris, 2004). In this chapter, a generalized calibration method (GCAL) is developed and presented. The aim of the method is to achieve coherence between estimates gained from different sources and to facilitate flexibility in the choice of the auxiliary data on various stratification levels in order to generate accurate estimates of totals and subtotals of a variable of interest y.

The general framework of calibration techniques is found in Section 2.4, in which the main characteristics of calibration in survey statistics is stated in Definition 2.4.1. The resulting calibration estimator

$$\hat{\tau}_y^{\text{CAL}} := \sum_{k\in S} d_k g_k y_k \tag{5.1}$$

(cf. Definition 2.4.2) for the total of variable y contains design weights d_k and correction weights g_k ($k = 1, \ldots, n_{\text{s}}$), where the correction weights g_k are determined by the calibration method. The products of design and correction weights, $w_k := d_k g_k$, are called *calibration weights*. Due to the inclusion of auxiliary data, the generalized calibration method can be assigned to the group of model-assisted methods (cf. Sections 2.2 and 2.4). We have seen in Section 2.4 that the GREG estimator (2.19) for the population total can be interpreted as a calibration estimator. Thus, the GREG estimator is a special case of GCAL.

In considering applications in the context of official statistics, the justification for applying GCAL is defined within the *European Statistics Code of Practice*. It consists of 15 principles aimed at fostering the provision of high quality statistics in Europe (cf. Section 1.1 and Eurostat, 2011).

In addition to accuracy and reliability, coherence between multiple sources is one of these principles. The requirements which have to be considered in GCAL are described in the following paragraphs. Most of these requirements arise from the requirements for modern surveys, which was already stated in Section 1.1. These demands comprise simultaneous estimation of several variables of interest on different stratification levels on the one hand, and the increased amount of auxiliary data available on the other hand. Due to both the consideration of several auxiliaries and the estimation of statistics for various variables of interest, GCAL can be designated as a *multivariate* calibration method. It is a *multi-domain* method as well, since several stratification levels may be considered simultaneously.

Requirements of GCAL

Firstly, the availability of a great amount of auxiliary data accompanied by the demand for estimates on both aggregated levels and highly disaggregated stratification levels may lead to a high number of constraints in problem (2.39). Consequently, the maximal spread of the calibration weights, referred to as *Gelman Bound*

$$\mathrm{GB} := \frac{\max_{k=1,\ldots,n_\mathrm{s}} w_k}{\min_{k=1,\ldots,n_\mathrm{s}} w_k}, \tag{5.2}$$

may be high, which is undesired in many applications (cf. Gelman, 2007 and Münnich and Burgard, 2012). In addition, the standard calibration approach in problem (2.39) does not prevent from negative weights. Since there is no sensible interpretation of negative weights in official statistics, they should be omitted. Generally, the spread of the correction weights increases if the number of restrictions increases or if several restrictions on highly disaggregated stratification levels are included, since the feasible set is shrunken. In extreme cases, the feasible set can be empty.

Secondly, the accuracy of the calibration estimator may suffer from using auxiliary data from different sources, especially if incorrect register data or inaccurate estimates are included as benchmarks, e.g. gained from different surveys with various sampling or estimation methods. For example, it is possible to use both direct estimates and small area estimates simultaneously as calibration benchmarks. In general, small area estimates may exhibit a systematic bias. Since the auxiliary data are treated as known values, these inaccuracies may hand over to the calibration estimates, which cannot be avoided in classical calibration methods. Consequently, the calibration estimator may rely on incorrect data, which can result in inefficient or biased estimates.

Thirdly, official statistics in Europe often have access to a *master* sample with a comparably high sampling fraction. This master sample is usually conducted frequently for an entire country, but only in large time intervals. An example for this is the *German Census* (cf. Destatis, 2018). On the other hand, several smaller surveys are conducted within lower time intervals, possibly regionally limited, and with significantly smaller sampling fractions. In order to make use of these extra sources, official statistics aims to generate a general valid vector of calibration weights, which incorporates all sources in one vector of weights. In comparison to the results exclusively gained by the master sample, the inclusion of the extra sources may support the quality of the estimates. Moreover, this approach should yield coherent and consistent results between the master sample and the additional surveys on various stratification levels. This is important in order to achieve social acceptance of the published statistics and underlines the interest to achieve one general valid vector of calibration weights, especially in household surveys. GCAL allows for the consideration of these aspects with high flexibility. Similar approaches have been suggested in the context of GREG estimation by Renssen and Nieuwenbroek (1997), Merkouris (2004), and Merkouris (2010). Currently, a system of a master sample and affiliated surveys is developed for some of the household surveys in Germany, called the *integrated system of household surveys* (cf. Riede et al., 2013).

Finally, a method for the variance and MSE estimation is required for the calibration estimator based on GCAL to facilitate the practicability of the approach, as the techniques which are commonly applied for the GREG estimator cannot simply be taken over (see Section 5.4).

Problem formulation and literature

Several publications within the last two decades proposed extensions to the classical calibration problem (2.39) in order to meet (at least a part of) the requirements mentioned above. After introducing the mathematical problem formulation and some notations, we provide an overview of these extensions and explicitly explain the framework of GCAL.

In accordance with the notation in Section 2.1 we consider a finite population $\mathcal{U} = \{1, \ldots, N\}$ and a sample $S \subseteq \mathcal{U}$ of size $n_\text{s} \leq N$. We assume the design weights d_k to be known and strictly positive for each element $k \in S$ of the sample. The goal is to estimate the total $\tau_y = \sum_{k \in \mathcal{U}} y_k$ of variable of interest y using the calibration estimator $\hat{\tau}_y^{\text{CAL}} = \sum_{k \in S} d_k g_k y_k$ (cf. Definition 2.4.2) with correction weights g_k ($k \in S$). The correction weights are gained under consideration of the calibration benchmarks $\tau_{x_i} = \sum_{k \in S} d_k g_k x_{ik}$ regarding q auxiliary variables, whose totals are known in advance. In that regard, the vector $x_k := (x_{1k}, \ldots, x_{qk})^T \in \mathbb{R}^q$ contains the individual auxiliary values for all units $k \in S$. The basis for the development of GCAL is then given by the standard calibration approach (2.39), i.e.

$$\begin{aligned} \min_{g \in \mathbb{R}^{n_\text{s}}} \; & \sum_{k \in S} d_k D(g_k) \\ \text{s.t.} \; & \sum_{k \in S} d_k g_k x_{ik} = \tau_{x_i} \text{ for } i = 1, \ldots, q, \end{aligned} \tag{5.3}$$

where the objective function is characterized by a distance function D (see Table 2.1). In the 1980s, the approach in (5.3) was limited to the GREG-type distance function $D(g_k) = \frac{1}{2}(g_k - 1)^2$, since the calibration estimator for the population total based thereon is equivalent to the GREG estimator (cf. Theorem 2.4.3). Primarily published by Cassel et al. (1976), the GREG estimator was extensively analyzed, and several alternative expressions were given inter alia in Isaki and Fuller (1982) and Godfrey et al. (1984). A few year later, Deville and Särndal (1992) rewrote the GREG estimator such that it depends on a Lagrangian multiplier. The resulting formula exactly represents the calibration estimator $\hat{\tau}_y^{\text{CAL}}$ (cf. Definition 2.4.2) combined with the optimality conditions of the optimization problem (5.3). Thus, the equivalence was proved. Additionally, Deville and Särndal (1992) proposed further distance measures. Traditional choices for D are shown in Table 2.1 and Figure 5.1. Other distance functions were considered by Deville et al. (1993), Singh and Mohl (1996), and Stukel et al. (1996). In the following, we focus on the three distance functions presented in Table 2.1, i.e. the GREG-type, the Raking, and the ML-Raking distance function. For the applicability of GCAL under other distance functions, we refer to Subsection 5.6.7. The distance functions shown in Figure 5.1 differ in their treatment of the penalty term, which affects the functions outcome depending on how greatly the calibration weight w_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 those values with equal absolute distance between g_k and 1.0, e.g. $D(0.5) = D(1.5)$ and $D(0.1) = D(1.9)$. The Raking Ratio and the ML-Raking distance functions are based on a nonlinear dependency, where $g_k := 1 - \delta$ smaller 1.0 is penalized stronger than the appropriate $g_l := 1 + \delta$ greater 1.0 (for any $\delta > 0$). Since g_k represents a factorial deviation of w_k from d_k, a distance function fulfilling $D(g_k) = D(g_k^{-1})$ would be reasonable. Even though the Raking Ratio and ML-Raking distance function are more likely to fulfill this feature than the GREG-type objective function, the condition $D(g_k) = D(g_k^{-1})$ is not fulfilled for each of the considered distance functions. The applicability of a distance function which fulfill this feature is discussed in detail in Subsection 5.6.7.

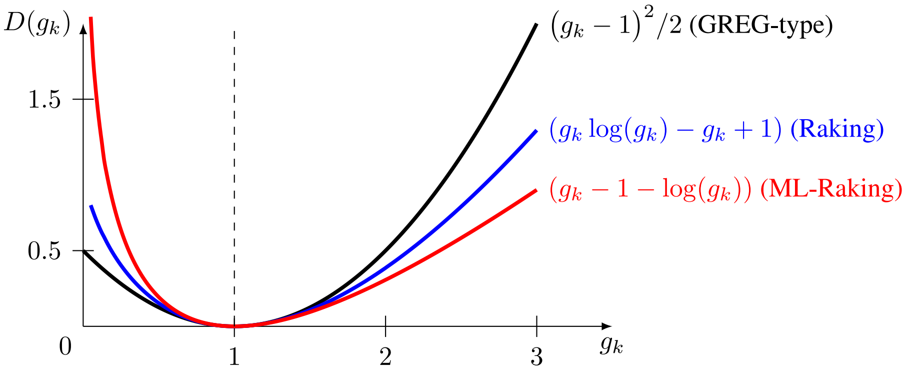


Figure 5.1: Common examples of distance functions for the calibration estimator.

In order to regulate and limit the spread of the calibration weights as well as to omit negative weights, Deville and Särndal (1992) proposed to add limits to the correction weights using the GREG-type objective function. These limits are referred to as *box-constraints* or *range restricted weights*. As described in the beginning of the section, a number of factors such as the various sources, structures (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, such that they only have to be fulfilled within specific predefined perturbations. The maximal tolerances allowed are restricted via additional box-constraints. This approach ensures the feasibility of the calibration problem even for a large number of benchmarks. Generally, an extreme large number of benchmarks can result from many auxiliary variables on different stratification levels. Since the benchmarks are obtained from known totals and different estimates gained by direct or small-area estimators (cf. Rao and Molina, 2015, and You and Rao, 2002), this relaxation may prevent coherence problems between the estimation levels. Moreover, an individual adjustment of the tolerance per benchmark can facilitate to incorporate different confidence measures for the different benchmarks. In considering a small-area estimate in a very small domain, its corresponding variance may be high compared to other benchmark estimates, Furthermore, the estimate may be biased, i.e. the confidence in it is low and the allowed tolerance for this benchmark should be higher than for a direct estimate in a larger domain.

In literature, the relaxation of benchmarks and the box-constraints for the calibration weights are often analyzed applying ridge weighting and ridge regression methods. Using these regularization techniques, the restrictions of the optimization problem are added as a penalty term to the objective function, and they deliver a close form solution depending on the penalty parameter. However, the box-constraints for the weights and deviations of the benchmarks are not necessarily fulfilled. Thus, an optimal penalty parameter has to be determined, which is often done by a trial-and-error strategy. Chambers (1996) discussed ridge weighting incorporating relaxation and box-constraints in the context of robust weighting for multipurpose establishment surveys. Rao and Singh (1997) mentioned a ridge regression method with projection of the weights to comply with the box-constraints and tolerances. The existence of solutions using ridge regression under box-constraints was discussed in Théberge (2000). Chen et al. (2002) analyzed box-constraints in a model-calibration environment by computing empirical likelihood estimators and model-calibrated empirical likelihood estimators. This approach was extended in Beaumont and Bocci (2008), and the equivalence to a ridge calibration was proved. Moreover, a calibration approach using ridge regression is presented in Montanari and Ranalli (2009). It ensures coherence using different sources of the benchmarks. In Rao and Singh (2009) a generalization of the ridge regression method was presented in order to comply with box-constraints and relaxations by using an iterative build-in tolerance specific procedure. In Guggemos and Tillé (2010), some selected benchmarks were shifted to a penalty term added to the objective function. Since box-constraints are fully neglected, a closed form was derived to apply the proposed method without an iterative solver. Nevertheless, all the methods mentioned require the determination of an optimal penalty parameter, as well as user-specified costs associated with the altitude of the deviations from the given benchmarks. The computation of both the penalty parameter and the costs is not possible in advance and is often done by examining different scenarios within the calibration process.

To overcome a predetermination of cost and penalty parameters, the relaxation and box-constraints are treated as real restrictions in GCAL (not as configurable penalty terms). This approach relies on the developments of Wagner (2013, Chapter 7) and Burgard et al. (2018). Then, the calibration problem (2.39) can be extended to the GCAL optimization problem

$$
\begin{aligned}
\min_{(g,\epsilon)\in\mathbb{R}^{n_\text{s}+q_2}} \quad & \sum_{k\in S} d_k D(g_k) + \sum_{j=1}^{q_2} \delta_j D(\epsilon_j) \\
\text{s.t.} \quad & \sum_{k\in S} d_k g_k x_{ik}^{\text{ex}} = \tau_{x_i^{\text{ex}}} \text{ for } i = 1,\dots,q_1 \\
& \sum_{k\in S} d_k g_k x_{jk}^{\text{rel}} = \epsilon_j \tau_{x_j^{\text{rel}}} \text{ for } j = 1,\dots,q_2 \\
& L_{g_k} \le g_k \le U_{g_k} \ \forall k = 1,\dots,n_\text{s} \\
& L_{\epsilon_j} \le \epsilon_j \le U_{\epsilon_j} \ \forall j = 1,\dots,q_2
\end{aligned}
\tag{5.4}
$$

with a distance function D of Table 2.1 and the number of $q = q_1 + q_2$ benchmarks. The auxiliary variables are denoted with $x^{\text{ex}}_{1k}, \dots x^{\text{ex}}_{q_1 k} \in \mathbb{R}$ $(k \in S)$ for the restrictions which should be fulfilled exactly and with $x^{\text{rel}}_{1k}, \dots x^{\text{rel}}_{q_2 k} \in \mathbb{R}$ $(k \in S)$ for the restrictions which have to be fulfilled with a certain degree of precision. The degree of precision is defined by the lower and upper bounds of $\epsilon_j \in \mathbb{R}_+$, denoted by L_{ϵ_j} and U_{ϵ_j}, respectively $(j = 1, \dots, q_2)$. Since ϵ_j states the deviation of the relaxed benchmarks and is restricted by the bounds L_{ϵ_j} and U_{ϵ_j}, the vector $\delta \in \mathbb{R}^{q_2}_+$ (which is used in the second part of the objective function) determines the magnitude of penalization within these bounds. The box-constraints for the correction weights g_k are denoted by $0 \leq L_{g_k} \leq g_k \leq U_{g_k}$ with $L_{g_k} \leq U_{g_k}$ $(k = 1, \dots, n_{\text{s}})$. It has to be remarked that the benchmark totals $\tau_{x^{\text{ex}}_i}$ $(i = 1, \dots, q_1)$ and $\tau_{x^{\text{rel}}_j}$ $(j = 1, \dots, q_2)$ may also be estimated totals instead of known totals. Due to simplifications, the common *hat*-notation is omitted.

Numerical solvers for GCAL

The inclusion of box-constraints prevents a derivation of a closed form solution as it can be given e.g. for the GREG estimator (2.19). Therefore, over the past decades, iterative methods have been developed. One common class of these algorithms is called *truncated* algorithms (TRUNC). One example is the function calib() in the R package sampling (cf. Tillé and Matei, 2016), which is generally applied only for the GREG-type distance function. In addition to TRUNC, Vanderhoeft (2001, pp. 29 f.) proposed a similar algorithm based on a projected Newton algorithm (ProjN). Both the TRUNC and ProjN algorithms are applicable in practice even for comparably large problem instances, although the computational burden of ProjN is significantly higher. However, they are not provably supposed to find the unique optimal solution of the box-constrained calibration problem (5.4). The computed solution is mostly close to the optimal solution. Nevertheless, extreme cases may yield larger differences between the optimal solution and the solution computed by TRUNC and ProjN. Their functionality is explained in Section 5.3, and the performance differences are discussed in Subsection 5.6.6. As an alternative, Wagner (2013) proposed to solve problem (5.4) using the highly efficient commercial software *IBM ILOG CPLEX Optimization Studio*[1]. Unfortunately, this software is not freely available, which is a desirable feature of the methods in this thesis.

Aside from the mentioned solvers, Münnich et al. (2012b) and Wagner (2013) developed the SSN method as a very efficient alternative to TRUNC and ProjN. The SSN algorithm provably finds the unique optimal solution of problem (5.4). Using the special structure of the problem, this method comprises a significant reduction of the dimension of the optimization problem and enables a sensitivity analysis of the benchmarks using the Lagrangian multipliers. Since SSN is applicable for very large problem instances and convergence results can be proved, it is the preferred solver for GCAL. To make SSN applicable for GCAL, the approaches of Münnich et al. (2012b) and Wagner (2013) are used and extended further here.

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

Outline

In Section 5.2, the mathematical formulation of GCAL incorporating various potential constraints is derived (depending on stratification levels, sources, etc.). Thereafter, the solution strategy using a Lagrangian approach and a SSN method as suggested in Wagner (2013) is outlined. Similar to the developments in Chapter 4, the special structure of problem (5.4) enables to equivalently rewrite the KKT-system into a lower dimensional nonlinear system of equations, where its dimension only depends on the number of constraints q_1 and q_2 and is independent of the sample size n_{s}. As discussed in Section 5.6, the algorithms based on SSN are fast enough to solve even large problem instances. The algorithmic solver is presented in Section 5.3. Since the variance and MSE estimation of classical calibration methods such as the GREG calibration estimator significantly differ from the MSE estimation under a GCAL model, an innovative MSE estimation strategy is proposed in Section 5.4 based on a rescaling bootstrap. The statistical accuracy, numerical efficiency, and practicability of GCAL and its MSE estimation method are discussed based on the household dataset of Germany (cf. Section 2.6) in a simulation study in Section 5.6. In addition, the advantages, opportunities, characteristics, and limits of the developed methods are addressed.

5.2 General calibration model

To present the solution strategy for problem (5.4), we redefine the problem using the following notations. The matrices X^{ex} is defined as design weighted auxiliary matrices for the q_1 auxiliary variables $x^{\text{ex}}_{1k}, \ldots, x^{\text{ex}}_{q_1 k} \in \mathbb{R}$ 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 $x^{\text{rel}}_{1k}, \ldots, x^{\text{rel}}_{q_2 k} \in \mathbb{R}$ for all units $k \in S$, whose benchmarks are relaxed and therefore need to be fulfilled allowing for a predefined tolerance:

$$X^{\text{ex}} = \begin{bmatrix} d_1 x^{\text{ex}}_{11} & \ldots & d_{n_{\mathrm{s}}} x^{\text{ex}}_{1n_{\mathrm{s}}} \\ \vdots & & \vdots \\ d_1 x^{\text{ex}}_{q_1 1} & \ldots & d_{n_{\mathrm{s}}} x^{\text{ex}}_{q_1 n_{\mathrm{s}}} \end{bmatrix} \in \mathbb{R}^{q_1 \times n_{\mathrm{s}}} \text{ and } X^{\text{rel}} = \begin{bmatrix} d_1 x^{\text{rel}}_{11} & \ldots & d_{n_{\mathrm{s}}} x^{\text{rel}}_{1n_{\mathrm{s}}} \\ \vdots & & \vdots \\ d_1 x^{\text{rel}}_{q_2 1} & \ldots & d_{n_{\mathrm{s}}} x^{\text{rel}}_{q_2 n_{\mathrm{s}}} \end{bmatrix} \in \mathbb{R}^{q_2 \times n_{\mathrm{s}}}.$$

The benchmark totals are denoted by $\tau_{x^{\text{ex}}_1}, \ldots, \tau_{x^{\text{ex}}_{q_1}} \in \mathbb{R}$ and $\tau_{x^{\text{rel}}_1}, \ldots, \tau_{x^{\text{rel}}_{q_2}} \in \mathbb{R}$ respectively. The approved perturbations of the relaxed variables are given by $\epsilon_1, \ldots, \epsilon_{q_2} \in \mathbb{R}_+$. Generally, the matrices X^{ex} and X^{rel} correspond to population total benchmarks. If regional (i.e. area- or stratum-specific) benchmarks are added, for example for auxiliary variable $i \in \{1, \ldots, q_1\}$, the i^{th} row of X^{ex} is extended. Assume that area-specific benchmarks are defined on stratification level r, which contains L_r areas $l_r = 1, \ldots, L_r$ (see Figure 2.1 for an example). Then the sample $S = \{1, \ldots, n_{\mathrm{s}}\}$ is divided into L_r parts S_{l_r} with $S = \bigcup_{l_r=1}^{L_r} S_{l_r}$. The i^{th} row of X^{ex} is then extended to the matrix

$$\left[\begin{array}{ccc} d_1 x^{\text{ex}}_{i1} & \ldots & d_{n_{\mathrm{s}}} x^{\text{ex}}_{in_{\mathrm{s}}} \\ \hline d_1 x^{\text{ex}}_{i1} \cdot \mathbb{1}_{(1 \in S_1)} & \ldots & d_{n_{\mathrm{s}}} x^{\text{ex}}_{in_{\mathrm{s}}} \cdot \mathbb{1}_{(n_{\mathrm{s}} \in S_1)} \\ \vdots & & \vdots \\ d_1 x^{\text{ex}}_{i1} \cdot \mathbb{1}_{(1 \in S_{L_r})} & \ldots & d_{n_{\mathrm{s}}} x^{\text{ex}}_{in_{\mathrm{s}}} \cdot \mathbb{1}_{(n_{\mathrm{s}} \in S_{L_r})} \end{array}\right] \in \mathbb{R}^{(1+L_r) \times n_{\mathrm{s}}}, \tag{5.5}$$

where each row corresponds to one area $l_r \in \{1, \dots, L_r\}$. Due to the indicator functions $\mathbb{1}_{(\cdot)}$, only these components of a row that belong to the respective area are considered. The other components are set to zero. The number of the constraints q_1 is then replaced by $q_1 \leftarrow q_1 + L_r$. 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 (e.g. state and federal states), but the totals may be relaxed on more disaggregated levels (e.g. cities and towns). This procedure is common in practical applications and will also be applied in the scenarios of the simulation study in Section 5.6.

In assembling the auxiliary matrices and benchmarks, the restriction matrix of problem (5.4) can be formulated as

$$A := \left[\begin{array}{c|ccc} X^{\text{ex}} & 0 & \dots & 0 \\ \hline & -\tau_{x_1^{\text{rel}}} & & 0 \\ X^{\text{rel}} & & \ddots & \\ & 0 & & -\tau_{x_{q_2}^{\text{rel}}} \end{array}\right] \in \mathbb{R}^{(q_1+q_2)\times(n_{\text{s}}+q_2)}, \tag{5.6}$$

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

$$b := \left(\tau_{x_1^{\text{ex}}}, \dots, \tau_{x_{q_1}^{\text{ex}}}, 0, \dots, 0\right)^T \in \mathbb{R}^{q_1+q_2}. \tag{5.7}$$

To measure the deviations of the a priori given design weights d_k from the calibration weights w_k and the perturbations ϵ_j to 1.0, one of the distance functions presented in Table 2.1 is applied, i.e. $D : \mathbb{R}_+ \to \mathbb{R}_{0_+}$ with

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, $\kappa = 1, \dots, n_{\text{s}} + q_2$ is the composed index for the respective component of the objective function of problem (5.4), i.e. indices $\kappa \leq n_{\text{s}}$ correspond to the n_{s} sampled units (index k) and indices $\kappa > n_{\text{s}}$ correspond to one of the q_2 relaxed benchmarks (index j). With this, problem (5.4) can then be equivalently rewritten as

$$\begin{aligned} \min_{z\in\mathbb{R}^{n_{\text{s}}+q_2}} \quad & P(z) := \sum_{\kappa=1}^{n_{\text{s}}+q_2} \tilde{d}_\kappa D(z_\kappa) \\ s.t. \quad & A\,z - b = 0 \\ & m \leq z \leq M \end{aligned} \tag{5.8}$$

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

Since the structure of problem (5.8) slightly resembles the structure of the `MMDopt` problem (4.20) in Chapter 4, we attempt to implement a similar solution strategy also based on a reformulation of the KKT-system. In order to do so, some properties of the objective function P of problem (5.8) are required and therefore proved in Lemma 5.2.1.

Lemma 5.2.1. Under the three distance functions $D : \mathbb{R}_+ \to \mathbb{R}_{0_+}$ of Table 2.1, the objective function P of problem (5.8) is twice continuously differentiable, strictly convex, and separable.

Proof. First, it is proved that the three functions $D : \mathbb{R}_+ \to \mathbb{R}_{0_+}$ of Table 2.1 are twice continuously differentiable and strictly convex:

1. GREG-type: The properties hold since D is a quadratic function.
2. Raking Ratio: D is twice continuously differentiable as a composition of twice continuously differentiable functions. Moreover, since $D''(z_\kappa) = \frac{1}{z_\kappa} > 0$, D is strictly convex.
3. ML-Raking: D is twice continuously differentiable as a composition of twice continuously differentiable functions. Moreover, since $D''(z_\kappa) = \frac{1}{z_\kappa^2} > 0$, D is strictly convex.

Consequently, $P(z)$ is twice continuously differentiable and strictly convex as a sum of compositions of twice continuously differentiable and strictly convex functions. Moreover, P is separable, as it can be rewritten as an independent sum over its components depending on its individual variables (cf. Remark 2.5.1). □

Due to the proof of Lemma 5.2.1, D'' is strictly positive, i.e. the inverse of the derivative D' is well-defined and given by

$$D'^{-1} : \mathbb{R} \to \mathbb{R}, \ \ u \mapsto D'^{-1}(u). \tag{5.9}$$

In Section 5.3, these properties are exploited to present an efficient numerical solver for problem (5.8). The required inverses $D'^{-1}(u)$ for the distance functions of Table 2.1 are given by the following:

1. GREG-type: $D'^{-1}(u) = u + 1$,
2. Raking Ratio: $D'^{-1}(u) = \exp(u)$, and
3. ML-Raking: $D'^{-1}(u) = \dfrac{1}{1-u}$.

5.3 Algorithmic solution

Reformulation depending on Lagrangian multipliers

The algorithmic solutions for calibration methods in survey statistics primarily differ in the kind of restrictions and the choice of the objective function. For some suitably formulated problems, a closed formula may exist. The most common example for this is the GREG estimator (2.19), which is equivalent to the calibration estimator in Definition 2.4.2 if the GREG-type distance function is applied (see Theorem 2.4.4). However, if additional restrictions such as box-constraints or relaxed benchmarks are added or other objective functions are utilized, the analysis of the optimality conditions of the calibration problem does not lead to a closed form solution. In this case, iterative solvers need to be applied. These are generally based on a Lagrangian approach as shown in Deville and Särndal (1992) and Deville et al. (1993). This strategy was picked up in Münnich et al. (2012b) and Wagner (2013, pp. 66-70) in more detail. The proposed procedure and its deviations are concisely sketched in the following paragraph for problem (5.8).

The main goal is to express the vector of correction weights $g \in \mathbb{R}^{n_s}$ and the vector of perturbations $\epsilon \in \mathbb{R}^{q_2}$ as functions $g_k(\cdot) : \mathbb{R}^{q_1+q_2} \to \mathbb{R}$ $(k = 1, \ldots, n_s)$ and $\epsilon_j(\cdot) : \mathbb{R}^{q_1+q_2} \to \mathbb{R}$ $(j = 1, \ldots, q_2)$ depending on the Lagrangian multipliers $\lambda \in \mathbb{R}^{q_1+q_2}$ of problem (5.8) belonging to the equality constraints $A\,z - b = 0$. The resulting function $z(\cdot) : \mathbb{R}^{q_1+q_2} \to \mathbb{R}^{n_s+q_2}$ with the components

$$z(\lambda) := \Big(g_1(\lambda), \ldots, g_{n_s}(\lambda), \epsilon_1(\lambda), \ldots, \epsilon_{q_2}(\lambda)\Big)^T \tag{5.10}$$

is inserted into the function of equality constraints

$$h : \mathbb{R}^{n_s+q_2} \to \mathbb{R}^{q_1+q_2}, h(z) = A\,z - b. \tag{5.11}$$

This leads to a $(q_1 + q_2)$-dimensional nonlinear system of equations $h\Big(z(\lambda)\Big) = 0$ which is solved in place of the original problem (5.8).

In analogy to the optimality conditions in Theorem 3.1.7 and the statements of Lemma 4.3.3 for the allocation problem, the same deviation can be done for the calibration problem (5.8) (cf. Wagner, 2013, Theorem 5.3.1. and Lemma 5.3.2.). Since the equality constraints of (5.8) are affine-linear and the objective function is strictly convex, the first order necessary optimality conditions are also sufficient if the Slater condition is satisfied (Theorem 3.1.7). Aside from these properties, the objective function is separable and the feasible set is convex. Thus, problem (5.8) can be equivalently reformulated as the nonlinear system of equations

$$\Psi(\lambda) = 0 \tag{5.12}$$

with

$$\Psi : \mathbb{R}^{q_1+q_2} \to \mathbb{R}^{q_1+q_2}, \;\; \lambda \mapsto A\,z(\lambda)\, - b, \tag{5.13}$$

where the function $z(\cdot) : \mathbb{R}^{q_1+q_2} \to \mathbb{R}^{n_s+q_2}$ is component-wise defined as

$$\begin{aligned} z_\kappa(\lambda) :&= \begin{cases} M_\kappa, & \text{if } -\frac{A_\kappa^T\lambda}{\tilde{d}_\kappa} \geq D'(M_\kappa) \\ D'^{-1}\left(-\frac{A_\kappa^T\lambda}{\tilde{d}_\kappa}\right), & \text{if } D'(m_\kappa) < -\frac{A_\kappa^T\lambda}{\tilde{d}_\kappa} < D'(M_\kappa) \\ m_\kappa, & \text{if } -\frac{A_\kappa^T\lambda}{\tilde{d}_\kappa} \leq D'(m_\kappa) \end{cases} \\ &= \text{Proj}_{[m_\kappa, M_\kappa]}\left(D'^{-1}\left(-\frac{A_\kappa^T\lambda}{\tilde{d}_\kappa}\right)\right) \end{aligned} \tag{5.14}$$

for $\kappa = 1, \ldots, n_s + q_2$ with the projection function $\text{Proj}_{[m_\kappa, M_\kappa]}(\cdot)$. The regularity of D is shown in (5.9). The following theorem proves the equivalence of solving (5.8) and (5.12).

Theorem 5.3.1. A vector $z^* \in \mathbb{R}^{n_s+q_2}$ is the unique solution of the optimization problem (5.8) if and only if there exist Lagrangian multipliers $\lambda^* \in \mathbb{R}^{q_1+q_2}$ such that $\Psi(\lambda^*) = 0$ defined in (5.12) is satisfied.

For the proof of Theorem 5.3.1, we refer to Münnich et al. (2012b, Theorem 3).

In applying Theorem 5.3.1, the $(q_1 + q_2)$-dimensional nonlinear system of equations in (5.12) has to be solved to achieve the optimal solution of the $(n_s + q_2)$-dimensional optimization problem in (5.8). Then, the solution $z^* \in \mathbb{R}^{n_s+q_2}$ of problem (5.8) is component-wise given by

$$z^*_\kappa = z_\kappa(\lambda^*) \tag{5.15}$$

for all $\kappa = 1, \dots, n_s + q_2$ with $z_\kappa(\cdot)$ computed by (5.14). Since $q_1 \ll n_s$ and $q_2 \ll n_s$, in general, the computational burden to solve (5.12) is supposed to be significantly lower than the computational effort needed to solve problem (5.8). Finally, the optimal solution $g^* \in \mathbb{R}^{n_s}$ and the optimal penalty parameter $\epsilon^* \in \mathbb{R}^{q_2}$ are determined by

$$g^* = (z^*_1, \dots, z^*_{n_s})^T \text{ and } \epsilon^* = (z^*_{n_s+1}, \dots, z^*_{n_s+q_2})^T. \tag{5.16}$$

Control of the spread of the weights

As mentioned in the beginning of this chapter, the Gelman Bound (5.2) (cf. Münnich and Burgard, 2012) plays a significant role in general calibration methods to prevent highly spread calibration weights. The importance of the consideration of the Gelman bound has also been observed in the context of small area estimation in business surveys by Burgard et al. (2014). For the numerical realization, Wagner (2013, Chapter 7.2) proposed adding the condition in Equation (5.2) to the general calibration problem (5.8) as an additional inequality constraint. Due to the ratio of the maximum to the minimum weight, this constraint is nonlinear, non-differentiable, and particularly non-separable, which prohibits the application of the considered solution strategy. Alternatively, the constraint can be rewritten as a set of $2n_s + 1$ linear inequality constraints. This enables the application of common nonlinear optimization solvers, such as barrier methods, augmented Lagrangian methods, and `SQP` methods (see Section 3.1). Nevertheless, the proposed reformulation to the lower dimensional nonlinear system of equations in (5.12) is not possible. Instead, the problem can be directly solved via the commercial software *CPLEX*. Since the free availability of the solver is a key point of the thesis, another strategy for the consideration of the Gelman bound is presented in the following paragraph.

The strategy in Algorithm 5, referred to as `GB_control`, is based on an adjustment of the box-constraints for the correction weights g_k $(k = 1, \dots, n_s)$. However, the Gelman bound is related to a ratio not of the g_k but of the calibration weights $w_k = d_k g_k$. As a consequence, the inclusion of `GB_control` to the `GCAL` problem (5.8) does not yield the unique optimal solution of `GCAL` with the original Gelman Bound condition (5.2) included as an inequality constraint. However, the precision of the solution is high enough for the majority of practical applications. In particular, the predefined Gelman bound is an upper bound for the resulting Gelman bound. Thus, Gelman bounds which are for instance established by law can not be breached. Since `GB_control` shrinks the feasible set of problem (5.8), the value of the objective function in the optimal solution $P(z^*)$ increases. Moreover, the risk of infeasibility increases, especially if the desired Gelman bound is smaller than the original Gelman bound computed by the design weights. Nevertheless several simulations have shown, that the use of `GB_control` does not yield to significantly higher values of the objective function in general. We refer to Figure 5.3 and Table 5.2 in Subsection 5.6.2.

Algorithm 5 Control of Gelman bound in `GCAL` (`GB_control`)

```
Input: Design weights $d \in \mathbb{R}^{n_s}$, box-constraints $L_g, U_g \in \mathbb{R}^{n_s}$, desired Gelman bound $\mathrm{GB} \in \mathbb{R}_+$
  for $k = 1, \dots n_s$
    $U_{g_k} = \min\left\{U_{g_k}, \frac{1}{d_k}\mathrm{GB}\min\{d\}\right\}$
    $L_{g_k} = \max\left\{L_{g_k}, \frac{1}{d_k \mathrm{GB}}\max\{d\}\right\}$
  end for
Return: Adjusted box-constraints $L_g, U_g \in \mathbb{R}^{n_s}$
```

In contrast to the Gelman bound strategy proposed in Wagner (2013, Chapter 7.2), the `GB_control` strategy does not necessarily yield the optimal solution. Nevertheless, this drawback is compensated by a dramatically reduced computational burden, a simple applicability, and an appropriateness of the results.

Comparison of algorithms

By introducing box-constraints to the calibration model, the function Ψ of the nonlinear system of equations in Equation (5.12) is not continuously differentiable, which prohibits us from applying the classical Newton method. However, widespread solvers for calibration problems with box-constraints make use of the reformulation to the nonlinear system of equations in (5.12), which was already mentioned in Section 5.1. Despite the non-differentiability, in a first step the classical Newton method is mostly applied to the problem (5.8) without box-constraints (i.e. the unconstrained problem), given by

$$\Psi_{\text{unconstr}}(\lambda) := A\, z_{\text{unconstr}}(\lambda)\; - b = 0 \tag{5.17}$$

with

$$z_{\text{unconstr}_\kappa}(\lambda) = D'^{-1}\left(-\frac{A_\kappa^T \lambda}{\tilde{d}_\kappa}\right) \quad \text{for all } \kappa = 1, \dots, n_s + q_2. \tag{5.18}$$

In contrast to function Ψ of (5.13), the unconstrained function Ψ_{unconstr} is continuously differentiable. Thus, a classical Newton method can be applied to solve the nonlinear system of equations $\Psi_{\text{unconstr}}(\lambda) = 0$. The actual handling of the box-constraints differs from one solver to another.

In the `TRUNC` algorithm (i.e. function `calib()` of `R` package `sampling`; cf. Tillé and Matei, 2016 and Algorithm 6), the Newton method used for solving the unconstrained problem (5.17) is successively applied several times. After each round, the components $\kappa = 1, \dots, n_{\text{s}} + q_2$ which do not fulfill the box-constraints m_κ and M_κ (i.e. these with $z_{\text{unconstr}_\kappa}(\lambda) < m_\kappa$ or $z_{\text{unconstr}_\kappa}(\lambda) > M_\kappa$) are *truncated* and frozen to the value of the respective box-constraint. Thereafter, the reduced problem without the frozen components is solved again until all boxes are fulfilled and a given tolerance is reached. Generally, TRUNC does not find the optimal solution for GCAL, since if one component fails to comply the box-constraints in one round, it is irrevocably frozen to the corresponding box-constraint and not considered any further. An example of the difference between the optimal solution and the solution via the truncated algorithm is given in Subsection 5.6.6.

Algorithm 6 Truncated algorithm for the solution of GCAL (TRUNC)

```
Input: $\Psi_{\text{unconstr}} : \mathbb{R}^{q_1+q_2} \to \mathbb{R}^{q_1+q_2}$, $\lambda_0 = 0_{\mathbb{R}^{q_1+q_2}}$ initial value,
       $k = 0$, $A^{k+1} = A$, $b^{k+1} = b$, $z^* = 1_{\mathbb{R}^{n_{\text{s}}+q_2}}$, ind $= \left\{1, \dots, (n_{\text{s}} + q_2)\right\}$
  while $\left(\|\Psi_{\text{unconstr}}(\lambda_k)\| \geq \text{tol}\right) \mid \left(\sum_{\kappa=1}^{(n_{\text{s}}+q_2)} \mathbb{1}_{\left((m_\kappa > z^*_\kappa) \vee (M_\kappa < z^*_\kappa)\right)} \neq 0\right)$
    $k = k + 1$
    solve $A^k\, z_{\text{unconstr}}(\lambda^k) - b^k = 0$ (simple computation if $D$ is GREG-type; else Newton)
    set $z^*[\text{ind}] = z_{\text{unconstr}}(\lambda^k)$
    for $\kappa \in$ ind
      if $m_\kappa > z^*_\kappa$
        $z^*_\kappa = m_\kappa$ ;  ind $\leftarrow$ ind $\setminus \{\kappa\}$
      else if $M_\kappa < z^*_\kappa$
        $z^*_\kappa = M_\kappa$ ;  ind $\leftarrow$ ind $\setminus \{\kappa\}$
      end if
    end for
    $A^{k+1} \leftarrow A[\,,\text{ind}]$
    $b^{k+1} \leftarrow b - A[\,,-\text{ind}]\, z^*[-\text{ind}]$
    $z_{\text{unconstr}}(\cdot) \leftarrow z_{\text{unconstr}}(\cdot)[\text{ind}]$
  end while
Return: Solution $z^*$
```

The algorithm ProjN proposed by Vanderhoeft (2001, pp. 29 f.) differs slightly from TRUNC. In contrast to TRUNC, the Newton method is only applied once. Within each iteration, the classical Newton step is replaced by an approximation that contains the box-constraints in form of projections. Nevertheless, even the ProjN algorithm does not guarantee convergence. In Wagner (2013, Section 5.5), the performance of TRUNC and ProjN are analyzed. In applying ProjN, several numerical instabilities such as break downs and so-called *zig-zagging* effects of the Lagrangian multipliers are observed. Moreover, the computing time of ProjN significantly exceeds the computing time of TRUNC (by approximately a factor of 100). Thus, ProjN is not further considered in the thesis. The computing time of TRUNC, in particular for a GREG-type objective function, is rather small (especially compared to ProjN). Nevertheless, it is not a suitable solver for GCAL, since the optimality of the solution cannot be verified. Hence, alternative algorithms need to be utilized.

Dealing with the non-differentiability of Ψ, Münnich et al. (2012b) proposed applying the SSN method (Algorithm 1) in a way that is similar to the allocation problem in Chapter 4, which allows convergence results to be stated. Moreover, SSN and TRUNC reveal a similar computational burden, which is shown in the study in Wagner (2013, Section 5.5). For the convergence proofs of SSN, the semismoothness of Ψ is sufficient, which is proved in the next theorem.

Theorem 5.3.2. The function Ψ defined in (5.13) is semismooth.

Proof. Following Qi and Sun (1993), the minimum and maximum function are strongly semismooth. Since $\text{Proj}_{[m_\kappa, M_\kappa]}(x) = \min\left\{M_\kappa, \max\{m_\kappa, x\}\right\}$ for $x \in \mathbb{R}$, the projection is semismooth as it is a composition of semismooth functions (Lemma 3.2.5, item 5.). For that reason, $A_l^T z(\lambda) - b_l$ $(l = 1, \ldots, q_1 + q_2)$ is semismooth. Then, since all components of Ψ are semismooth (and also Lipschitz-continuous), Ψ is semismooth due to Item 3 of Lemma 3.2.5. □

To conclude, it is proved in Theorem 5.3.2, that the SSN algorithm (see Algorithm 1) can be applied to solve the GCAL problem (5.8) of dimension $n_\text{s} + q_2$, which has been rewritten as a significantly lower dimensional nonlinear system of equations (5.12) of dimension $q_1 + q_2$. The numerical performance is discussed in Subsection 5.6.6. For reasons of numerical stability, a non-smooth version of the Armijo step-size rule is integrated (cf. Algorithm 2). As seen in Section 3.2, the convergence rates of SSN method are similar to the classical Newton method, which is exemplarily shown in the results in Subsection 5.6.6.

To conclude, the general calibration method GCAL solved via the SSN algorithm allows for the opportunity to have a timely very efficient calibration under the consideration of various constraints, such as box-constraints, relaxation of benchmarks, and the Gelman bound control via the tool called GB_control. The practicability and further results are shown in Section 5.6.

5.4 Variance estimation

In practice, a statistical method is only sensibly applicable if the quality of the estimates computed by the method is quantifiable. Thus, the development of a variance or MSE estimation technique for the calibration estimator based on GCAL is essential. As shown in Equation (2.22), a variance estimator for the GREG estimator is given by

$$\widehat{\text{Var}}(\hat{\tau}_y^{\text{GREG}}) = \sum_{k \in S} \sum_{l \in S} \left(\frac{\pi_{kl}}{\pi_k \pi_l} - 1 \right) \frac{(y_k - x_k^T \hat{\beta})(y_l - x_l^T \hat{\beta})}{\pi_{kl}} \tag{5.19}$$

(cf. Särndal et al., 1992, Chapter 6.5). Similar variance estimation methods for classical calibration techniques based on Definition 2.4.2 are given by Demnati and Rao (2004), Estevao and Särndal (2006), and D'Arrigo and Skinner (2010). All methods mentioned result in closed forms of the variance estimators, which are based on Taylor linearization strategies.

Compared to established calibration methods (potentially with box-constraints), GCAL allows more flexibility in the calibration process, i.e. the relaxation of benchmarks, the consideration of area-specific benchmarks, the inclusion of a predefined Gelman bounds, and the individually adjustable maximum perturbations and penalty parameters. Consequently, the structure of the general calibration problem changes substantially. While this has no effect on the point estimator (2.38), it does have a major impact on the variance estimation. In that regard, a significant underestimated variability of the population estimates obtained by the variance estimation methods in Deville and Särndal (1992), Estevao and Särndal (2006), and D'Arrigo and Skinner (2010) is expected, if the relaxed benchmarks are omitted in the determination of the residual variance estimator. In addition, most of the techniques are only valid for the estimation of the variance of the population estimates, but not for area-specific estimates. Furthermore, these variance estimation methods focus on the estimation of means and totals only, whereas in general, there is an interest in running regression models and obtaining correct inference on the regression parameters as well.

An alternative to the linearized variance estimators are resampling methods based on *bootstrap* strategies, which have been primarily published by Efron (1979). As also stated in Kovar et al. (1988) and Särndal et al. (1992, pp. 442 ff.), a bootstrap technique can be described by the following procedure. We start by supposing a given sample S of the population $\mathcal{U}$. The goal is to estimate the unknown parameter ϑ using the estimate $\hat{\vartheta}$ by means of the sample S. The required value to quantify the quality of the estimate is an estimation of Var$(\hat{\vartheta})$ or MSE$(\hat{\vartheta})$. A classical bootstrap procedure is given by the following:

1. Construct an artificial population $\mathcal{U}^{\text{boot}}$ that is assumed to mimic the real and unknown population $\mathcal{U}$. In general, $\mathcal{U}^{\text{boot}}$ is equal to the sample $S \subseteq \mathcal{U}$.

2. Draw a series of $R_{\text{Boot}} \in \mathbb{N}$ independent *sub-samples* S_h^r from $\mathcal{U}^{\text{boot}}$ and calculate estimates $\hat{\vartheta}^{\text{boot}}$ in the same way as $\hat{\vartheta}$ was calculated.

3. The observed distribution of $\hat{\vartheta}_1^{\text{boot}}, \ldots, \hat{\vartheta}_{R_{\text{Boot}}}^{\text{boot}}$ is considered as an estimate of the sampling distribution of $\hat{\vartheta}$. Thus, Var$(\hat{\vartheta})$ can be estimated by

$$\widehat{\text{Var}}(\hat{\vartheta})^{\text{boot}} := \frac{1}{R_{\text{Boot}} - 1} \sum_{r=1}^{R_{\text{Boot}}} \left(\hat{\vartheta}_r^{\text{boot}} - \left(\frac{1}{R_{\text{Boot}}} \sum_{\iota=1}^{R_{\text{Boot}}} \hat{\vartheta}_\iota^{\text{boot}} \right) \right)^2, \tag{5.20}$$

   which is equivalent to the empirical variance of the bootstrap estimates $\hat{\vartheta}_r^{\text{boot}}$.

Since all steps necessary to obtain the calibration weights are reproduced in every bootstrap replication $r = 1, \ldots, R_{\text{Boot}}$, a bootstrap implies a massive amount of computation time. Thus, the classical bootstrap approach appears to be unfeasible if multiple population estimates are calculated at the same time. Furthermore, due to the computational costs, users may be less likely to use a classical bootstrap for obtaining valid inference for their models. As a solution to this drawback, Dippo et al. (1984) proposed the usage of replicate weights. They stated that the implementation of a replication through replicate weights may facilitate the computation of design-based variances from a given data set by far more researchers. Many important surveys already provide replicate weights, such as the American Community Survey[2] (ACS) and the Household Finance and Consumption Survey[3]. In the German Census context, the usage of replicate weights seems to be appropriate as well, since many sources of uncertainty can be easily modeled using replications. For the calibration model GCAL, the derivation of replicate

[2] https://usa.ipums.org/usa/repwt.shtml

[3] http://www.ecb.europa.eu/pub/pdf/other/ecbsp1en.pdf

weights is rather simple as the weights correspond to the output of `GCAL` within each replication. By doing R_{Boot} replications, it is possible to obtain R_{Boot} vectors containing the replicate weights for each unit in the sample. As the way the replications are constructed may alter in every survey, the construction of the replicate weights for the StrRS design will be given an adequate attention here. The results are presented in Section 5.6.1.

In order to enable replicate weights to be used for many different estimators, Dippo et al. (1984) and Fay (1989) proposed restricting all replicate weights to non-zero values. Rao and Wu (1988) proposed a bootstrap procedure using a scaling of the data in the replications. This approach was extended by Rao et al. (1992) by scaling the weights instead of the data, thereby allowing a variance estimation for non-smooth statistics. Chipperfield and Preston (2007) further increased the efficiency of this approach by using an SRS design. The bootstrap proposed for `GCAL` is based on the *rescaling bootstrap* introduced by Preston (2009). This approach extends the former bootstraps to the case of stratified multistage sampling, which allows its application to be used in a large number of different survey designs. As stated in Burgard et al. (2018), approximately 50% of the design weights d_k $(k = 1, \dots, n_s)$ are set to values near zero, while the others are accordingly increased, such that the sum of the computed bootstrap design weights equals to the original ones. This procedure is done separately in each stratum $h = 1, \dots, H$. The replicate weights are then obtained using the appropriate rescaling bootstrap presented in Algorithm 7.

Algorithm 7 Rescaling bootstrap for the variance estimation of `GCAL` problems (`resc.boot`)

```
Input: $R_{\text{Boot}} \in \mathbb{N}$ number of bootstrap replications
  for $r \in 1, \dots, R_{\text{Boot}}$
    for $h \in 1, \dots, H$
      draw a sub-sample $S_h^r$ without replacement from $S_h$ of size $n_h^{boot} = \lfloor n_h/2 \rfloor$
      set $\lambda_h^{\text{boot}} = \sqrt{n_h^{\text{boot}} \cdot \frac{1-f_h}{n_h - n_h^{\text{boot}}}}$ and define $\delta^r := (0, \dots, 0)^T \in \mathbb{R}^{n_h}$
      for $k = 1, \dots, n_h$
        set $\delta_k^r \begin{cases} 1, & \text{if } k \in S_h^r \\ 0, & \text{else} \end{cases}$
        compute $d_h^r = \left(1 - \lambda_h^{\text{boot}} + \lambda_h^{\text{boot}} \frac{n_h}{n_h^{\text{boot}}} \delta_k^r\right) \cdot d_h$
      end for
    end for
    $g^r$ is the solution to the calibration problem using $d^r$ instead of $d$
    $w^r := \left(d_1^r g_1^r, \dots, d_{n_s}^r g_{n_s}^r\right)^T$ is the $r$-th replicate weight
  end for
Return: $r$-th replicate weights $w^r$ for all $r = 1, \dots, R_{\text{Boot}}$.
```

The variance estimates for some arbitrary totals, means, and proportions can then be obtained by the simple sum of squares method shown in (5.20). To verify the accuracy of the variance estimator $\widehat{\text{Var}}(\hat{\vartheta})^{\text{boot}}$ gained by the rescaling bootstrap, the variance estimator needs to be relatively compared with the Monte-Carlo variance of the point estimate $\hat{\vartheta}$ in a simulation study.

Thus, the relative bias of the variance estimate given by

$$\frac{\mathrm{E}\Big(\widehat{\mathrm{Var}}(\hat{\vartheta})^{\mathrm{boot}}\Big) - \mathrm{Var}(\hat{\vartheta})^{\mathrm{MC}}}{\mathrm{Var}(\hat{\vartheta})^{\mathrm{MC}}} \tag{5.21}$$

has to be evaluated. In this way, the difference of the expected value of the bootstrap variance estimates and the Monte-Carlo variances of the point estimate $\hat{\vartheta}$ computed in a Monte-Carlo simulation study with the aid of the $1\,000$ samples is compared in a relative manner.

The proposed rescaling bootstrap is applied to a `GCAL` problem in Subsection 5.6.5. The occurring optimization problems are also solved via the `SSN` algorithm. For a more detailed analysis of the rescaling bootstrap, we refer to Preston (2009) and Burgard et al. (2018).

5.5 Summary of methods

Prior to the analysis of the simulation study in Section 5.6, the presented methods and algorithms are summarized in this Section. In general, the `GCAL` method is a generalized calibration method, which offers a wide range of possibilities with regard to a flexible consideration of the restrictions and distance measures used. Moreover, it can be interpreted as a generalization of classical calibration methods like the GREG calibration (see Subsection 2.4).

In a first step, a desired distance function has to be selected, which covers the preferences of the user (see Table 2.1). Then, the restrictions have to be specified. In that regard, an individual definition of box-constraints for the correction weights is possible. By using an additional tool called `GB_control` (see Algorithm 5), it is also possible to restrict the variation of the calibration weights. In addition to the box-constraints, the calibration benchmarks have to be selected. These benchmarks may refer to any possible stratification level, i.e. the simultaneous consideration of national and regional benchmarks is enabled. Moreover, each benchmark can individually be referred to as an exact or a relaxed benchmark. This allows the simultaneous consideration of a very large number of benchmarks. Moreover, benchmarks for small regions can also be included without the risk of infeasibility problems. For each relaxed benchmark, an individual selection of the permitted tolerance can be considered as well. Thus, benchmarks for small regions or even estimated benchmarks gained from other surveys can be included in a way such that the size of the tolerance takes into account the size of the region or the estimation accuracy of the estimates.

In general, the problem can be formulated as a restricted optimization problem (5.4). Its dimension is equal to the sample size, which may exceed the number of one million in applications like the German Census (cf. Münnich et al., 2012a). Despite the high dimension, a standard solver for restricted optimization problems, which are briefly described in Section 3.1, is applicable. In order to avoid the direct solution, the original problem can be transformed into a significantly lower dimensional nonlinear system of equations by converting the optimality conditions in analogy to Münnich et al. (2012b). The resulting lower dimensional problem can then be solved using the semismooth Newton method (`SSN`) presented in Section 3.2. In that regard, a linear dependence between the computing time and the dimension of the original problem can be observed. The necessary conditions for the applicability of this strategy are the separability and convexity of the objective function as well as the convexity of the feasible set. These properties are given for all three objective functions considered. A detailed analysis of the performance of the SSN algorithm compared to the solvers TRUNC (cf. Tillé and Matei, 2016) and ProjN (cf. Vanderhoeft, 2001, pp. 29 f.), which are based on similar approaches, shows that SSN outperforms TRUNC and ProjN with regard to the computing time. In addition, it is the only algorithm which provably reaches the optimal solution (see Subsection 5.6.6).

Due to the relaxation of benchmarks and the box-constraints, common variance estimation techniques based on linearization techniques are not applicable. Instead, a rescaling bootstrap (cf. Preston, 2009) is proposed. This approach is only applicable due to the time-efficient solution strategy using the SSN algorithm, as the calibration has to be done repeatedly. Besides the variance estimation, the bootstrap enables to provide vectors of replication weights in order to allow further social-scientific studies.

Finally, it can be summarized that the methods presented show a possibility to solve a generalized calibration problem in a very time-efficient way by exploiting the specific structure of the problem. Besides, there is a linear dependence between the computation time and the dimension n_s of the origin problem, which in fact is the size of the sample. For the sake of comprehensibility, the methods and algorithms concerning GCAL are illustrated in a graphical overview in Figure 5.2. In that regard, blue boxes indicate statistical methods, orange boxes represent numerical algorithms, and green boxes show the resulting output.

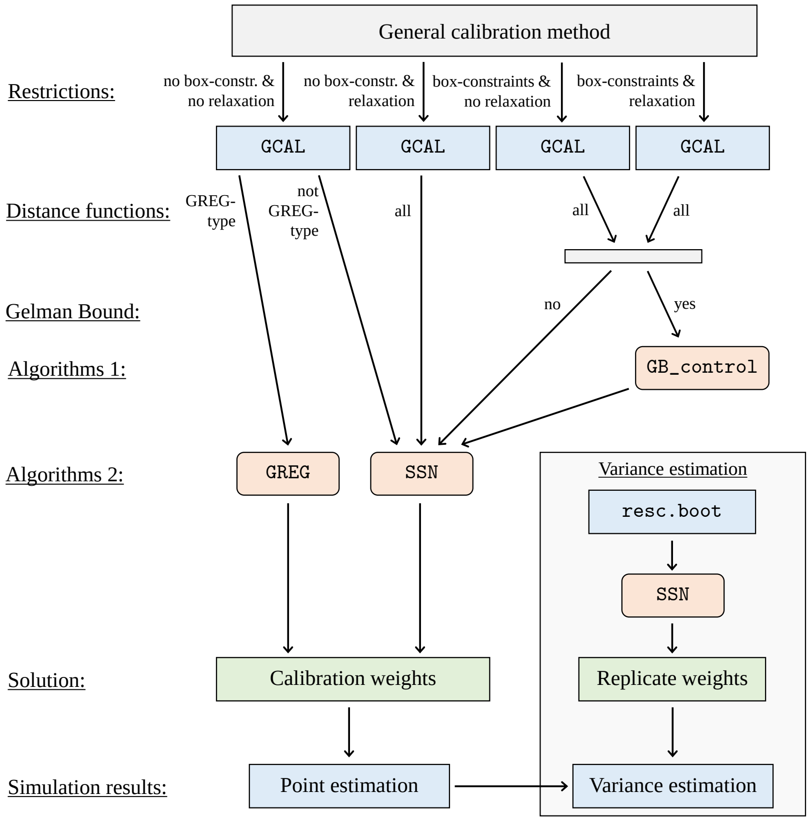


Figure 5.2: Summary of all options of the generalized calibration method GCAL.

5.6 Simulation study and results

5.6.1 Framework

As for the optimal allocation, the simulation study for GCAL is also based on the synthetic RIFOSS dataset introduced in Section 2.6 which is restricted to the federal states of Hesse, North Rhine-Westphalia, Rhineland-Palatinate, and Saarland, with a population size of $11\,121\,631$ households accommodating $30\,077\,329$ individuals. The sampling design is based on $6\,272$ strata built as cross-classifications of sampling points (SMP – 784 regional areas) and classes of household sizes (HHS – 8 classes). In contrast to Chapter 4, the overall sampling fraction is fixed to 2%, i.e. the total sample size is given by $n_\text{s} = 222\,433$ households. The stratified samples are drawn using the MMDopt method described in Chapter 4 with (cv)-standardization and auxiliary variables EDI, PEN, and AGE4.1. The allocation is computed with equal weights, lower bounds $m_h = 2$, and upper bounds equal to stratum size $M_h = N_h$ for each cross-classification stratum $h = 1, \dots, H$. This results in SMP-specific sampling fractions from 1.6% to 3.9%, where the highest sampling fractions are primarily assigned to small SMPs.

For the calibration benchmarks, the totals of considered auxiliary variables can be assumed to be known by (properly managed) registers or other surveys, namely

- ZEN (number of persons living in the household),
- EF117A, EF117B, EF117S (occupational status),
- ILO1, ILO4 (type of employment), and
- ISCEDA, ISCEDB, ISCEDD (highest graduation)

as well as additional classes of cross-classifications of age and gender, 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. For a detailed description of the variables we refer to the Tables B.1 and B.2. In practice, these benchmark totals may also be gained from other surveys with sampling designs or estimation techniques that differ completely from the regarded survey (e.g. model-based estimation methods such as small area estimation; c.f. Münnich et al., 2012a, pp. 129 f.). Then, the relaxation of these benchmarks support the handling of possible inconsistencies and biased estimates, as described in Section 5.2.

There are two major goals of applying GCAL. The first is to gain accuracy increases for (regional and overall) total estimates of some variables of interest compared to the HT estimator, which strongly depends on the correlation between auxiliary variables and variables of interest. The second goal is to ensure coherence of the estimates with the known totals of the auxiliary variables. Depending on various scenarios, the coherence should be achieved on several stratification levels and with various predefined tolerances. The six scenarios considered are described in Table 5.1. They are split into three cases, each with and without consideration of the Gelman bound. Generally, the number of benchmarks increases from the first to the last row of the table. The SMP-specific benchmarks for ZEN are included in each of the six scenarios without relaxation. In the first of the three cases, federal state-specific benchmarks for the auxiliaries are added (without relaxation). In the second case, SMP-specific relaxed benchmarks are added for the auxiliaries. The third case additionally contains benchmarks for Age×Gender classes (exact for federal states and relaxed for SMPs). If the Gelman bound is considered, the Gelman bound of the calibration weights should not exceed the original Gelman bound computed by the design weights. The overall number of benchmarks is tabulated in the last column.

Table 5.1: Various scenarios applied to `GCAL`. "✓" means the respective benchmarks are included, "–" vice versa. The values $xx\%$ refer to the maximal allowed tolerance.

	ZEN	Auxiliaries		Age×Gender		Gelman	Bench-
	SMP	Fed. state	SMP	Fed. state	SMP	Bound	marks
`BL.exact`	✓	✓	–	–	–	–	816
`BL.exact+GB`	✓	✓	–	–	–	✓	816
`SMP.rel(Aux)`	✓	✓	±15%	–	–	–	7 088
`SMP.rel(Aux)+GB`	✓	✓	±15%	–	–	✓	7 088
`SMP.rel(Aux&AxG)`	✓	✓	±15%	✓	±18%	–	13 392
`SMP.rel(Aux&AxG)+GB`	✓	✓	±15%	✓	±18%	✓	13 392

It should be noted that the calibration estimator for the population total concerning scenario `BL.exact` is equivalent to the GREG estimator for the population total with the federal state-specific totals of the auxiliaries and SMP-specific totals for ZEN used as benchmarks. To evaluate the results and analyze the strengths and weaknesses of the `GCAL`, the results of this scenarios are compared to the HT estimates. Aside from the analysis of weights and benchmarks (Subsections 5.6.2 and 5.6.3), the accuracy of point estimates on various stratification levels is compared for the auxiliaries and several other variables of interest in Subsection 5.6.4. The accuracy is measured by the Monte-Carlo RRMSE and RBIAS computed on the basis of the $R_{\text{MC}} = 1\,000$ Monte-Carlo replications. For each evaluation, the results for the three distance functions of Table 2.1 are analyzed. In addition, the variance estimation with a rescaling bootstrap (see Section 5.4) is computed with $R_{\text{Boot}} = 199$ bootstrap replications per Monte-Carlo replicate. These results are summarized in Subsection 5.6.5. Finally, the algorithmic performance of the algorithms is investigated in Subsection 5.6.6. In particular, the advantages of `SSN` compared to the truncated algorithm `TRUNC` and the GREG estimator are emphasized. The evaluations of Subsections 5.6.2, 5.6.3, and 5.6.6 are exemplarily based on the results of the sample with number 615. A randomly conducted study yields similar results for other samples.

The color selection of the headers in the `lattice`-plots are adopted from Chapter 4, blue and green are given for the auxiliary variables, and red and orange are for the variables of interest.

5.6.2 Design weights versus calibration weights

Within the calibration process, the design weights d_k are adjusted to the calibration weights $w_k = d_k g_k$, where the correction weights g_k are generated by GCAL. If the calibration has no influence, $g_k = 1$ holds for all $k = 1, \ldots, n_s$. Thus, the higher the influence of the calibration (i.e. the higher the number of benchmarks or the more restrictive the benchmarks), the higher the g_k deviates from 1.0. To illustrate this, the correction weights g_k are displayed using density plots in Figure 5.3 for the three scenarios without GB_control and the three distance functions. The light blue vertical lines in each panel highlight the position of the 5%- and the 95%-quantiles of the weights. Generally, the number of benchmarks increases from row 1 to 3. Firstly, we observe a significant increase in the variance for the scenarios with (relaxed) SMP benchmarks (rows 2 and 3), especially if Age×Gender classes are included (row 3). This effect can be seen by the lower curves at $g_k = 1$ and the bigger tails of the distributions, as displayed by the outwards shifted vertical lines of the quantiles. This is a result of the increased number of benchmarks. If some attempts were made to fulfill the SMP-specific benchmarks exactly, i.e. to omit the relaxation, the SSN algorithm for GCAL would break down due to the non-feasibility of the resulting problem.

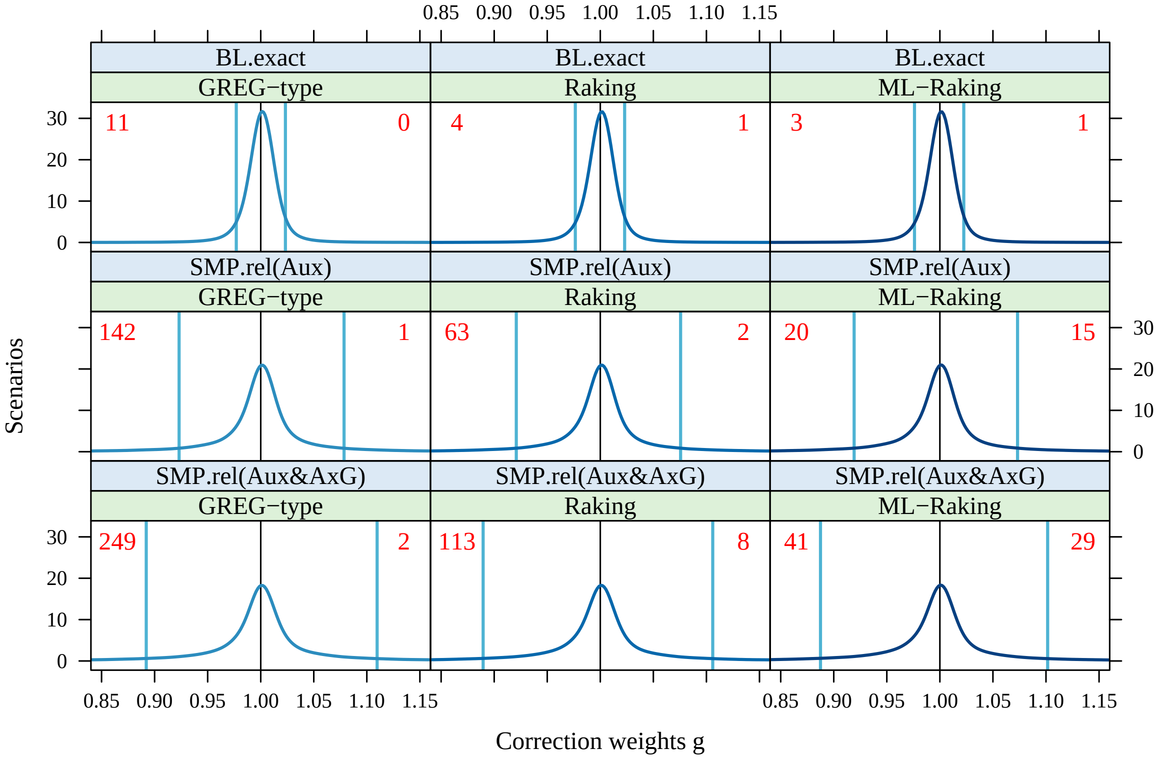


Figure 5.3: Density plots of correction weights g for scenarios without Gelman bound control.

When looking at the shapes of the density plots, a similar behavior can be observed for the three distance functions. However, in considering the red numbers at the top-right and top-left of each plot, there are significant differences between the three distance functions. The number at the top-left corresponds to the number of weights reaching the lower bound of 0.2, whereas the number at the top-right corresponds to the number of weights reaching the upper bound of 5.0. Firstly, the number of active box-constraints increases with the number of restrictions (from row 1 down to row 3). Secondly, the number of active lower bounds is clearly smaller for the Raking distance function and even smaller from the ML-Raking objective function. The opposite effect can be observed for the upper bounds. This observation is consistent with the analysis of the distance function in Section 5.1 and Figure 5.1, where the weight g_k being smaller than 1.0 are more penalized for Raking and ML-Raking compared to the GREG-type distance functions. This results in a skewed shifted distribution of the correction weights, which is also highlighted in Figure 5.4, where the correction weights g_k resulting from the three distance functions are plotted against each other (in log scale; for scenario `SMP.rel(Aux&AxG)`). The first mentioned distance function is plotted on the ordinate. In looking exemplarily at the panels including the GREG-type distance function, units with extremely low weights for the GREG-type function are assigned with higher weights for the other two functions and vice versa. Although the final choice of the distance function depends on the specific application, it may be recommended to use the ML-Raking distance function, since the calibration weights consists of the design and correction weights and a doubling of a design weight (i.e. $g_k = 2$) should approximately be penalized with the same amount as a halving of a design weight (i.e. $g_k = 0.5$). This is most likely given for the ML-Raking function and is mostly unfulfilled for the GREG-type function.

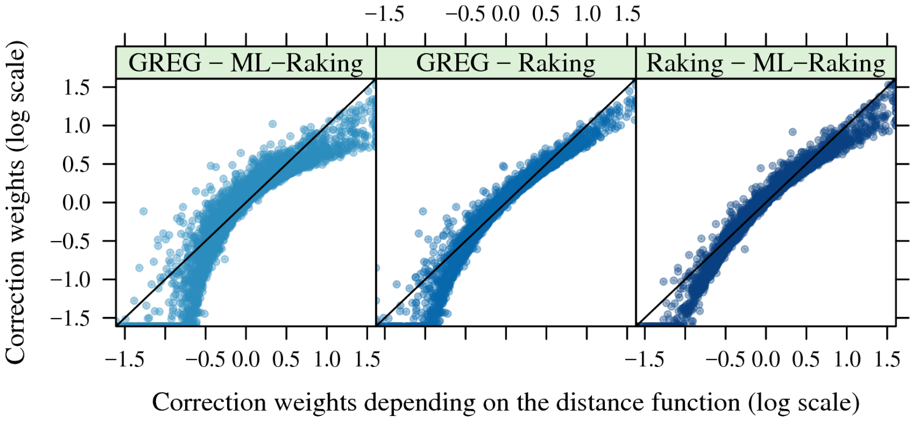


Figure 5.4: Scatterplot of correction weights under different distance function (for scenario `SMP.rel(Aux&AxG)`) - first-mentioned distance function is plotted on the ordinate.

In Table 5.2, the values of the objective functions (without penalty term for relaxed benchmarks) are shown for the three distance functions and the different scenarios. However, we only focus a column-wise comparison of the values, because it is not reasonable to compare the values of the distance functions row-wise with one another. Firstly, the more benchmarks are included or the more restrictive the benchmarks are, the higher are the values of the objective functions. This corresponds to the increasing variances of the distributions in Figure 5.3. The objective values increase by a factor of 10 to 15, if SMP-specific benchmarks are added for the auxiliaries (i.e. from scenarios `BL.exact[+GB]` to `SMP.rel(Aux)[+GB]`), which stresses the high effort to comply with regional benchmarks (despite relaxation). The additional consideration of the Age×Gender classes increases the objective values by a factor of 2 compared to the scenarios SMP.rel(Aux)[+GB]. This is caused by some SMPs with very bad HT estimates for certain Age×Gender classes. This is analyzed in more detail in Subsection 5.6.4. Secondly, the inclusion of GB_control also raises the objective value but only in a slight manner. Thus, GB_control seems to be easily includable in each application. Several other simulations reveal similar results, and it turns out that in general, GB_control only affects the few highest and smallest calibration weights. Therefore, the corresponding increase of the value of the objective function is comparably small. Nevertheless, this increase may easily result in a non-feasibility for one specific benchmark in one specific SMP, which would lead to a break down of the SSN algorithm. The sensitivity of the algorithm is examined in more detail in Subsection 5.6.7.

Table 5.2: Values of objective functions for calibration scenarios (without penalty term for relaxed benchmarks).

	GREG-type	Raking Ratio	ML-Raking
BL.exact	1 098.90	1 103.78	1 108.47
BL.exact+GB	1 457.29	1 476.72	1 497.68
SMP.rel(Aux)	15 049.09	15 005.66	15 009.58
SMP.rel(Aux)+GB	15 624.21	15 600.39	15 641.09
SMP.rel(Aux&AxG)	28 475.30	28 146.59	28 002.86
SMP.rel(Aux&AxG)+GB	29 250.17	29 000.49	28 957.23

With regard to Table 5.2, the consideration of a Gelman bound slightly increase the objective value, but the effect on the resulting Gelman bound is immense. In this application, the original Gelman bound of the design weights is $\mathrm{GB}_{\mathrm{org}} = 76.13$. This is also the Gelman bound to be reached by GCAL. We have to note, that a lower Gelman bound is almost impossible in the scenarios with relaxed benchmarks due to the number of benchmarks and relatively small SMPs. As shown in Table 5.3, the Gelman bounds after the calibration are extensively higher

Table 5.3: Gelman bounds for the calibration scenarios depending on the objective function.

	GREG-type	Raking Ratio	ML-Raking
BL.exact	388.86	388.82	388.72
BL.exact+GB	51.85	51.84	51.83
SMP.rel(Aux)	738.10	875.26	1 110.71
SMP.rel(Aux)+GB	76.13	76.13	76.13
SMP.rel(Aux&AxG)	792.72	948.31	1 222.14
SMP.rel(Aux&AxG)+GB	76.13	76.13	76.13

in the scenarios without GB_control, which is a well-known problem of calibration methods. In the scenarios with GB_control, the Gelman bounds are equal to or lower than GB_{org}. The slighter value for scenario BL.exact+GB is due to the low number of benchmarks in this scenario. Overall, the large differences of the Gelman bound and the high values occurring in the scenarios without GB_control (which are partly over $1\,000$) highlight the importance of using GB_control for GCAL. The control of Gelman bounds, which is often requested by law or at least desired in official statistics, is generally not possible in common calibration methods and therefore a unique selling point of GCAL.

5.6.3 Compliance with benchmarks

It is desirable to consider a great number of benchmarks of several auxiliary variables on various stratification levels. These benchmarks may be gained from different sources and are associated with different data quality. Infeasibility problems can easily occur if all these benchmarks have to be exactly met by the calibration procedure. Furthermore, satisfying benchmarks associated with sampling errors or biases in analogy to benchmarks of high quality is not always desirable. Therefore, GCAL permits specific benchmarks to be relaxed. Figure 5.5 highlights the functionality of the relaxation. Each boxplot contains the deviation of the totals estimated by the calibration estimator and the benchmark totals for all restrictions which are included in scenario SMP.rel(Aux&AxG) (i.e. both SMP- and federal state benchmarks for the variables mentioned in Subsection 5.6.1). The boxplots are divided into *Auxiliaries* and *Age*×*Gender* classes. The red vertical lines correspond to the maximal allowed tolerance for the relaxed benchmarks. The results with and without GB_control are almost equal, so that only three scenarios have to be distinguished.

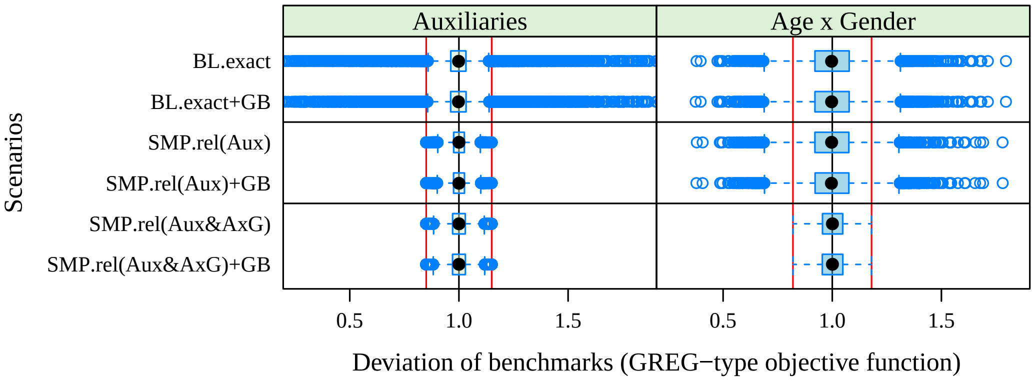


Figure 5.5: Compliance with benchmarks of SMP-specific estimates for scenarios with the GREG-type objective function.

In the scenarios without consideration of SMP benchmarks (BL.exact), there are several SMP-specific estimates which substantially differ from their benchmark totals and significantly exceed the maximal perturbations, which are used in the scenarios with SMP-specific benchmarks. In the case of BL.exact, the maximal deviations are about 100% of the correct total for the auxiliary variables and $+83\%$ and -70% for the Age×Gender classes. These extreme perturbations are unacceptable if SMP-specific estimates are subject of the survey. In official statistics for example, this is undesirable for a number of reasons. For one thing, the known totals of small regions are not able to be included in the calibration of a sample conducted in a nationwide survey, and thus the accuracy of regional estimates cannot benefit from the correlation of the auxiliaries and the variables of interest. On the other hand, it may be impossible to calibrate the sample with the aid of *well-managed* regional registers. This results in regional inconsistencies and coherence problems, and the prevention of these problems is a major task in several modern surveys. The issue is exemplarily shown in the maps of Figure 5.6 for the NUTS3- and SMP-specific estimates for variable `AGE4.1_Sex.2` (SMP left-hand side, NUTS3 right-hand side; each for scenarios `BL.exact` and `SMP.rel(Aux&AxG)`). Darker areas correspond to a higher deviation of benchmarks and estimates. Since SMP-specific benchmarks are included in scenario `SMP.rel(Aux&AxG)`, no inefficient SMP-specific estimates can be observed in this scenario. Even the NUTS3-specific estimates benefit from the inclusion of SMP-specific benchmarks. A similar behavior can be observed for other variables and is therefore not displayed. If the SMP-specific restrictions are included in `GCAL`, the accuracy of the SMP-specific estimates increases significantly for both, NUTS3- and SMP-specific estimates.

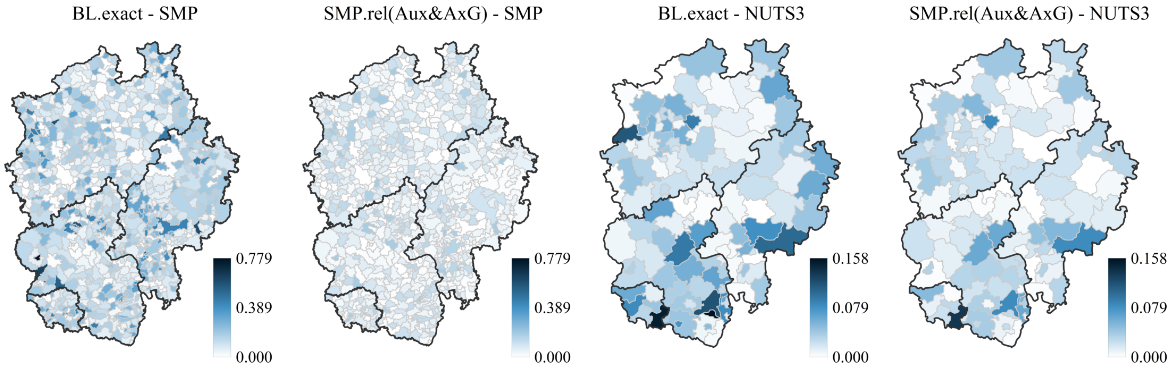


Figure 5.6: SMP- and NUTS3-specific deviations of estimates and benchmark totals for variable AGE4.1_Sex.2.

We have shown that the possibility of the relaxation of regional benchmarks in the calibration process by `GCAL` enables a range of opportunities and increases the efficiency and flexibility of users. Flexibility is particularly given by the option of choosing the height of the maximal tolerance for each variable in each region separately. Thus, the strength of the restrictions can be adapted to the framework of the regions to ensure the practicability of calibration while taking into account all restrictions. Figure 5.6 also shows that in general, the accuracy of aggregated stratification levels can benefit from relaxed benchmarks on more disaggregated stratification levels.

5.6.4 Point estimation

Aside from coherence and consistency, increasing the accuracy of estimates is a key aspect of GCAL. Within the simulation study, the efficiency of the calibration point estimates is evaluated for approximately 50 variables over all 1 000 Monte-Carlo replications. The SMP-specific RRMSEs for eight selected variables are shown in Figure 5.7 for all scenarios and for the GREG-type objective functions. The eight variables comprise four variables which are (partly) included in the calibration (green shaded header) and four variables of interest (red shaded header), which are not involved in the calibration. Each boxplot contains 784 points assigned to the 784 SMPs. Additional plots for other objective functions, other stratification levels, and more variables are shown in the Appendix B.3. In general, there are no fundamental differences observed for the other objective functions and other stratification levels. For comparability, the HT estimator is shown in the first row.

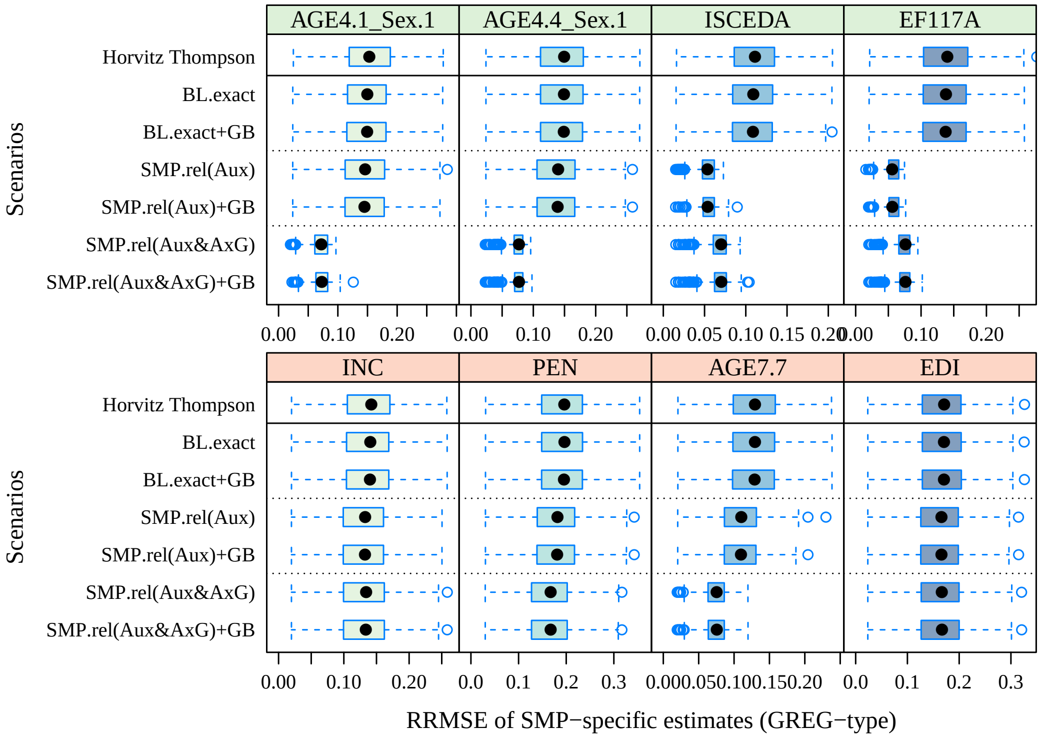


Figure 5.7: RRMSE of SMP-specific point estimates for scenarios with a GREG-type objective function.

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. Beside this, significant differences can be observed in the behavior of the point estimates of the auxiliaries (green shaded headers), which is primarily a consequence of the usage of the auxiliary variables as benchmarks. Moreover, the consideration of the Gelman bound does not yield any notable changes in the accuracy of the estimates. Some certain observations are explained in detail in the following paragraph.

With regard to the scenarios containing only federal state-specific benchmarks (see scenarios BL.exact[+GB] in rows 2 and 3), the accuracy of the SMP-specific estimates of all the auxiliary variables is similar to the HT estimates, since none of the auxiliary variables is included as benchmark on SMP-level. In case of the scenarios `SMP.rel(Aux)[+GB]` (rows 4 and 5), the accuracy of the SMP-specific estimates is significantly improved for the auxiliary variables ISCEDA and EF117A, since they are applied as relaxed benchmarks in `GCAL`. In this case, the boxplots contain no inadequate outliers due to the predefined maximal perturbations, and possibly poor estimates are controlled. Concurrently, the SMP-specific estimates of variables AGE4.1_Sex.1 and AGE4.4_Sex.1 are more or less equal to the results of the `BL.exact[+GB]` (rows 2 and 3) scenarios. Thus, the inclusion of specific benchmarks does not necessarily lead to accuracy-changes for other not included variables. In looking at the results with regard to the scenarios `SMP.rel(Aux&AxG)[+GB]` (rows 6 and 7), a significant decrease of the RRMSE for the SMP-specific estimates is also observed for the variables AGE4.1_Sex.1 and AGE4.4_Sex.1, since they are contained in `GCAL` as relaxed benchmarks. In these scenarios, the accuracy of the estimates for ISCEDA and EF117A slightly suffer due to the following reason. The feasible set is shrunken as the number of restrictions increases, i.e. the SMP-specific RRMSEs of estimates of variables which are included in `GCAL`, but which are not located at the maximal permitted tolerance may suffer due to the smaller feasible set. The RRMSEs of the NUTS3- and NUTS2-specific estimates in Figures B.7 and B.8 also benefit from the inclusion of SMP-specific benchmarks. Thus, gains in accuracy are generally also expectable for estimates on aggregated levels.

With regard to the variables of interest (red shaded header), different behaviors can also be observed. The accuracy of the SMP-specific estimates for AGE7.7 significantly increases if the Age×Gender classes are included as benchmarks (rows 6 and 7) due to the high correlation between AGE7.7 and the auxiliary variables belonging to the Age×Gender classes. For EDI and PEN, only slight efficiency increases can be observed. A reason for this is found in the simulation framework, as most of the auxiliaries are solely variables which count persons in household with specific properties and thus are not typical continuous variables (e.g. income or expenses). The specific framework for this simulation results in a correlation structure between auxiliaries and variables of interest which does not establish optimal conditions in order to allow for a significant error-decrease for EDI and PEN. Regarding this point however, it has to be considered that the main task of `GCAL` is not only to obtain accuracy increased estimates but also to have consistency between benchmarks and auxiliaries. Thus, the kindness of the calibration approach cannot be determined only by the height of the error-decrease of the point estimates of the variables of interest.

In Figure 5.8, the RBIAS is presented for the SMP-specific estimates for the same scenarios and variables as for the RRMSE in Figure 4.23. As per definition, the HT estimator is unbiased for all variables. Since the GREG estimator is model-assisted and an estimation based on `GCAL` can also be associated to this type of estimators, estimates based on `GCAL` should at least be asymptotically unbiased. Generally, unbiased estimates for most scenarios are observed, including those with relaxed benchmarks. In some SMPs, a small (mostly negative) bias occurs if `GB_control` is used, especially for the variables which are not involved in the calibration or are highly correlated with one of these. The reason for this is the significant restriction of the feasible set by `GB_control` to prevent a high Gelman bound, especially for very small and high design weights. Consequently, the calibration is slightly skewed for some variables, and this may result in a small bias. Nevertheless, the bias is relatively small and is not enough to be a

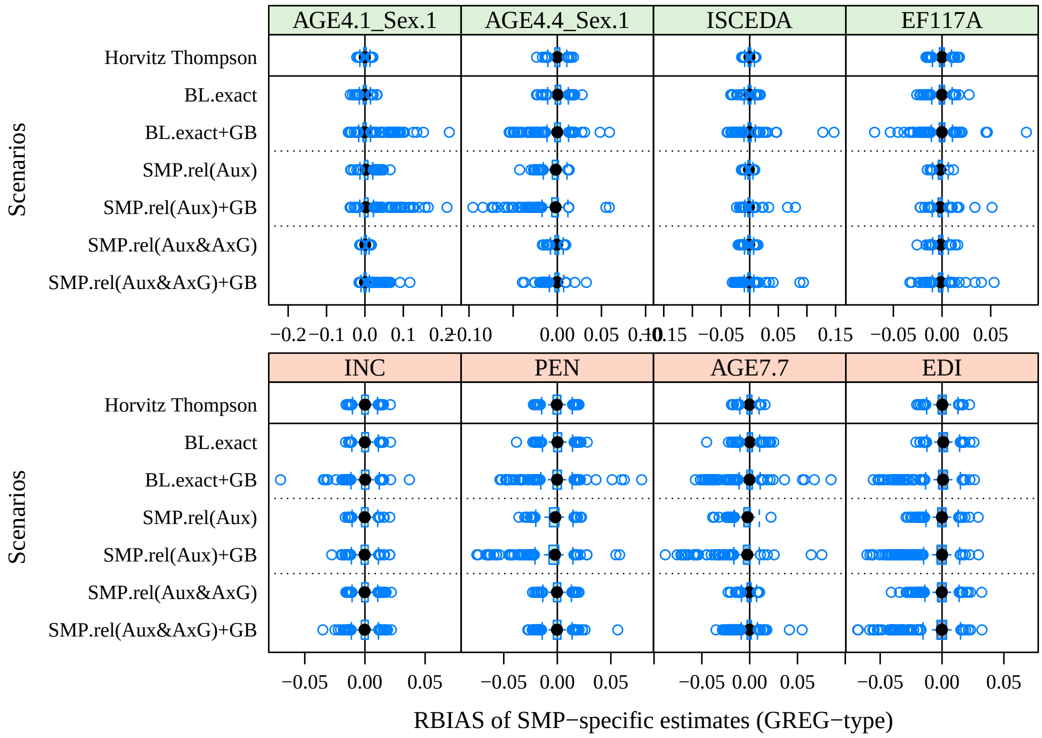


Figure 5.8: RBIAS of SMP-specific point estimates for scenarios with a GREG-type objective function.

driving force of the MSE (which is equal to the variance plus the squared bias). This can be seen in Table 5.4, where the quantiles of the ratio $\frac{\text{BIAS}^2}{\text{MSE}}$ are tabulated over all SMPs for six of the eight observed variables exemplarily shown for the scenario `SMP.rel(Aux&AxG)+GB`. The median of the share of the bias in the MSE over all SMPs is about one per mil. In a few SMPs, the share is over 10%, but these SMPs are not the ones with a high bias. A similar behavior can be observed for other variables and other distance functions (see Appendix B.3). Thus, the calibration estimator based on `GCAL` is generally unbiased for the observed variables and scenarios in this simulation framework.

Table 5.4: Quantiles (over all SMPs) of the ratio $\frac{\text{BIAS}^2}{\text{MSE}}$ for the six observed variables exemplarily shown for scenario `SMP.rel(Aux&AxG)+GB` with GREG-type objective function.

	AGE4.4_Sex.1	ISCEDA	EF117A	PEN	EDI
0%	$2.21 \cdot 10^{-8}$	$1.90 \cdot 10^{-10}$	$1.64 \cdot 10^{-9}$	$3.87 \cdot 10^{-10}$	$2.38 \cdot 10^{-9}$
25%	$2.04 \cdot 10^{-4}$	$2.90 \cdot 10^{-4}$	$2.48 \cdot 10^{-4}$	$1.39 \cdot 10^{-4}$	$1.20 \cdot 10^{-4}$
50%	$8.93 \cdot 10^{-4}$	$1.08 \cdot 10^{-3}$	$1.11 \cdot 10^{-3}$	$5.53 \cdot 10^{-4}$	$5.69 \cdot 10^{-4}$
75%	$2.72 \cdot 10^{-3}$	$3.61 \cdot 10^{-3}$	$3.17 \cdot 10^{-3}$	$1.77 \cdot 10^{-3}$	$2.12 \cdot 10^{-3}$
100%	$2.58 \cdot 10^{-1}$	$8.02 \cdot 10^{-1}$	$5.77 \cdot 10^{-1}$	$1.96 \cdot 10^{-1}$	$1.20 \cdot 10^{-1}$

5.6.5 Variance estimation

As mentioned in Section 5.4, a rescaling bootstrap is applied to receive a robust variance estimation of estimates gained by GCAL. Aside from the number of $R_{\text{MC}} = 1\,000$ Monte-Carlo replications, the number of bootstrap replicates is set to $R_{\text{Boot}} = 199$, which turns out to be a reasonable choice regarding the trade-off between computational burden and accuracy of the computed estimates. However, we have to remark that a generally valuable choice of R_{Boot} highly depends on the application and is not further investigated in this thesis. In the following evaluations, the six scenarios presented in Subsection 5.6.1 are considered in analogy to the presentation of the point estimates in Subsection 5.6.4. The first two scenarios omit relaxed benchmarks. With the exception of the box-constraints, the first scenario with GRGE-type distance function is equal to the GREG estimator. This statement is only valid for the population estimate, since area-specific calibration estimates gained by GCAL differ from the classical area-specific GREG estimate in general. The reason is a different foundation of the assisting model. In addition to the scenarios, a distinction is made between the three distance functions of Table 2.1 and the results for variables of interest and auxiliaries, which are partly used as relaxed benchmarks in the calibration model GCAL. The variance estimates are analyzed for estimates on various stratification levels, which is a special ability of GCAL.

In Figure 5.9, the relative bias of the variance estimates computed by the rescaling bootstrap of SMP-specific estimates are shown using the GREG-type distance function, i.e.

$$\frac{\mathrm{E}\left(\widehat{\mathrm{Var}}(\hat{\tau})^{\text{boot}}\right) - \mathrm{Var}(\hat{\tau})^{\text{MC}}}{\mathrm{Var}(\hat{\tau})^{\text{MC}}}. \tag{5.22}$$

In doing so, the difference of the expected value of the $1\,000$ bootstrap variances and the Monte-Carlo variances computed with the aid of the $1\,000$ samples is relatively compared. In general, absolute deviations of Equation (5.21) under 15% for SMP-specific estimates are considered as sufficiently accurate (marked with red dashed vertical lines in the figures). In Figure 5.9, the RBIAS of the variance estimates in plotted for the GREG-type distance function. To evaluate the robustness of the rescaling bootstrap concerning different distance functions, the results for the ML-Raking distance function are plotted in Figure 5.10. In the upper panels, four auxiliary variables are plotted. Whereas the variables ISCEDA and EF117A are considered as relaxed benchmarks on SMP-level in the scenarios in rows 3 to 6, the variables AGE4.1_Sex.1 and AGE4.4_Sex.1 are considered only in the last two scenarios. In the lower panels, four variables of interest are plotted, which have not been involved in the calibration process. Since the dots of most of the SMP-specific estimates are located within the two red dashed vertical lines, the bootstrap method provides accurate variance estimates in general. Some slight underestimations can be observed in particular for the scenarios in rows 5 and 6 where a high number of benchmarks is included. Nevertheless, they are only observed for those variables, which are included in GCAL as auxiliaries (green shaded header). The variance estimates for the observed variables of interest (red shaded header) are almost unbiased, which is important for the practical application. Although the majority of the variance estimates of the SMP-specific estimates is relatively biased with values smaller than 0.15, there are some outliers that are not negligible. Further analyses have shown that these outliers mostly correspond to small SMPs with a small sample size. The results for the other stratification levels follow similar patterns, whereas the relative bias for the larger NUTS2 and NUTS3 regions is absolutely smaller than the relative bias of the SMP-specific variance estimates (see Appendix B.3).

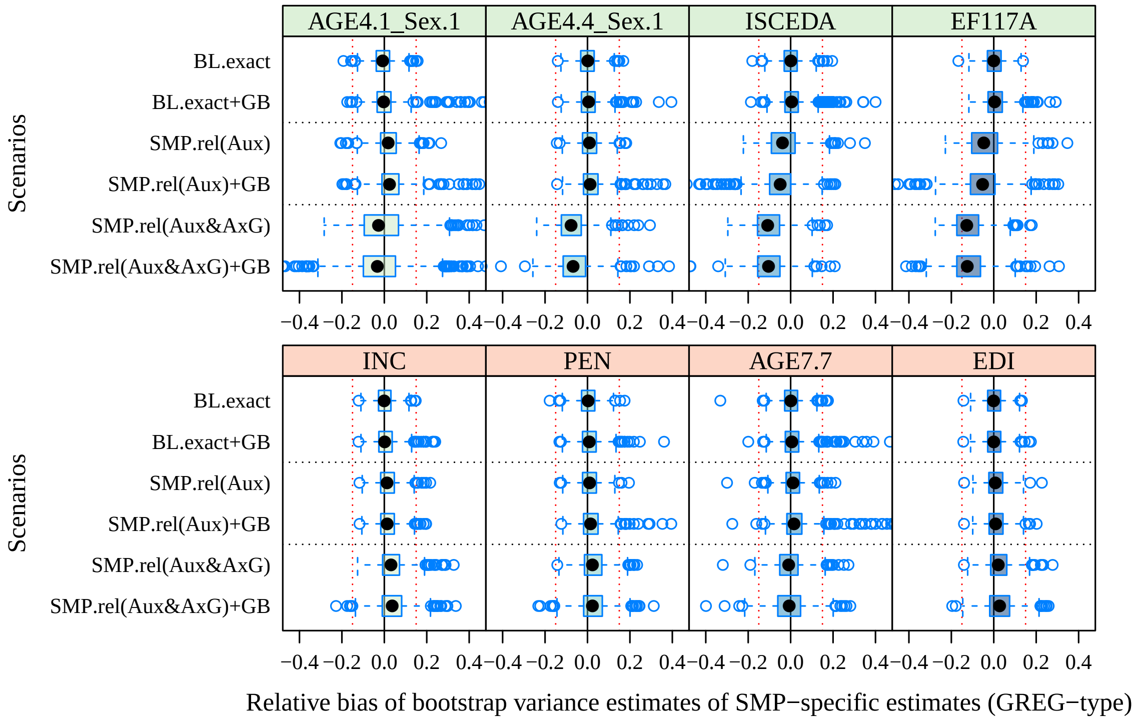


Figure 5.9: Relative bias of SMP-specific variance estimates computed by the rescaling bootstrap for scenarios with a GREG-type objective function.

In Figure 5.10, the bootstrap results are shown using the ML-Raking distance function. The results closely resemble those presented in Figure 5.9 for the GREG-type distance function. Similar results are also observable for the Raking distance function as well (see Appendix B.3). The high efficiency of the rescaling bootstrap for all considered distance functions buttresses the stability and flexibility of the rescaling bootstrap as a variance estimation technique for GCAL. This observation is in contrast to certain residual variance estimators, whose structure would clearly differ if another distance function was chosen.

Besides the use of the rescaling bootstrap as an appropriate technique for the variance estimation of GCAL, it does have further advantages compared to both a classical bootstrap and a residual variance estimator. As the rescaling bootstrap generates R_{Boot} vectors of replicate weights for GCAL, these can be utilized in further ways after the calibration. On the one hand, they can be provided to third party users as additional material to the vector of calibration weights. This allows the users to apply a simple and robust variance estimation for their generated point estimates. We are able to assume robustness and stability in these cases, since good results have been shown in Figures 5.9 and 5.10 for variables of interest, which are completely uninvolved in the calibration process. The strategy of providing replicate weights to third party users is not possible for a classical bootstrap, as it would mean drawing R_{Boot} subsamples out of the original sample. Thus, users would have to apply sampling techniques on their own. One major disadvantage of resampling strategies is the computational burden. In the simulation study that has been performed for this thesis, $R_{\text{Boot}} = 199$ calibrations have been executed for each of the $R_{\text{MC}} = 1\,000$ samples. In total, the computational effort contains solving GCAL for 199 000 times.

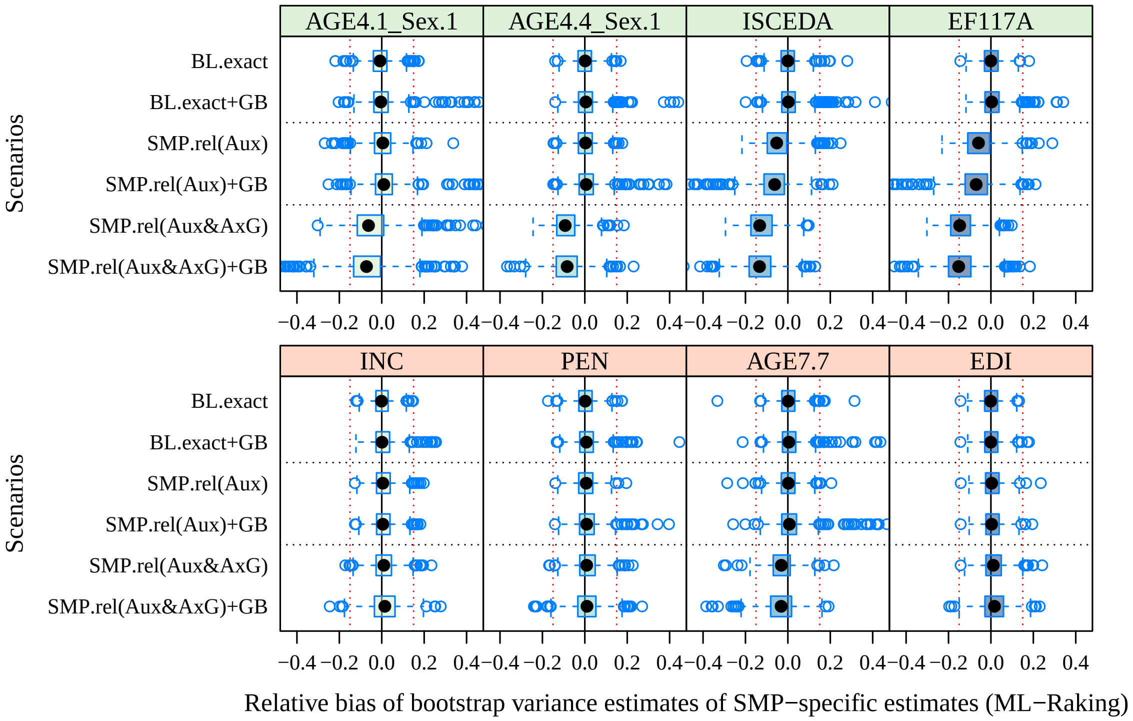


Figure 5.10: Relative bias of SMP-specific variance estimates computed by the rescaling bootstrap for scenarios with a ML-Raking objective function.

5.6.6 Algorithmic performance

As in Chapter 4, the numerical results are computed in R on a desktop PC with an Intel Core i7-6700 CPU at 3.40GHz × 8 and an internal memory of 32 GB. To achieve comparability, the algorithms are solely implemented in R. However, it has to be remarked that the computing time can be significantly reduced if C++ implementations are used. Firstly, the performance of SSN (see Algorithm 1) for GCAL is analyzed. Thereafter, the results of GCAL computed by the SSN algorithm are compared to the results computed by the common solver TRUNC (see Algorithm 6).

Performance of SSN for GCAL

As observed in Münnich et al. (2012b) and Wagner (2013), SSN for GCAL has huge advantages in computing time compared to standard solvers for nonlinear optimization. The fundamental reason for this is the avoidance of directly solving a highly dimensional optimization problem (that could have over one million variables) with thousands of equality constraints and box-constraints. A substituted lower dimensional problem needs to be solved instead, which can be derived exploiting the special structure of the problem such as the separability of the objective function. This allows us to solve even large problems within seconds. In Figure 5.11, the computing time of SSN and the number of iterations are plotted depending on the dimension of the problem (i.e. depending on the sample size; see left column), and the number of restrictions (i.e. depending on the total number of benchmarks used; see right column). The scenario considered is SMP.rel(Aux&AxG) with the GREG-type objective function. With regard to the left column, the number of restrictions is fixed to 784, which is a rather low number compared to the scenarios considered in Table 5.1. For the evaluations in the right column, the sample size is set to $n_s = 222\,433$.

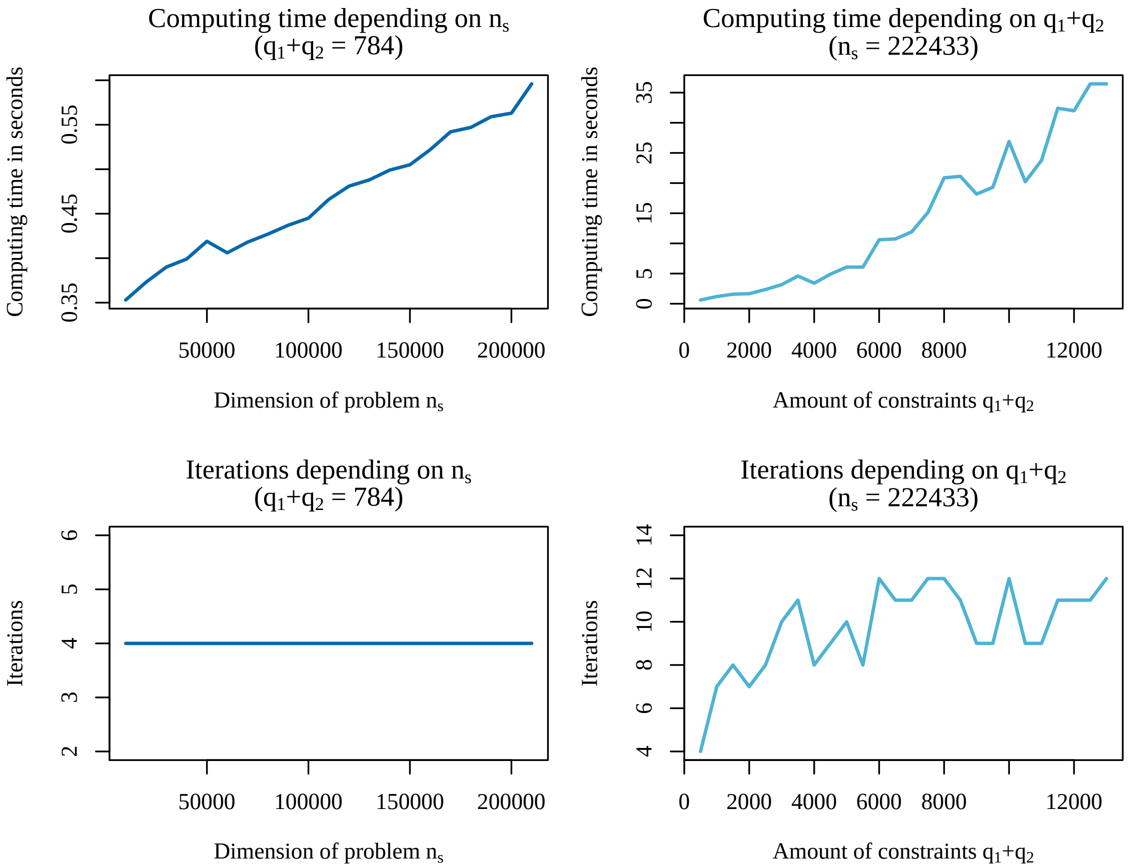


Figure 5.11: Computing time and number of iterations of SSN for GCAL depending on sample size n_s and number of restrictions (exact and relaxed) for scenario SMP.rel(Aux&AxG) and GREG-type objective function.

In looking at the upper left plot, the running time is linearly dependent on the problem dimension n_s. In the example with a low number of restrictions it is in the range within one second, even for a problem dimension of about $n_s = 200\,000$. This speed is unreachable using standard solvers, which would solve the optimization problem directly without the reformulation as lower dimensional nonlinear system of equations. The lower left plot shows that the number of iterations stays constant if n_s increases and the number of restrictions is fixed. Thus, the linear increase of the computing time can be attributed to the increased computational burden within each iteration. In the right column, the running time and number of iterations are plotted depending on $q_1 + q_2$. Since the dimension of the nonlinear function $\Psi(\lambda)$ (cf. (5.13)) is $q_1 + q_2$, increasing $q_1 + q_2$ results in solving a higher dimensional nonlinear system of equations with SSN. Thus, the running time exponentially increases in $q_1 + q_2$. The slight variation in the plots occurs due to the different strictnesses of the restrictions which are progressively added. An increased number of constraints also yields an increase in the number of iterations. Aside from the number of restrictions, the strictness of the restrictions also has a significant effect on the computing time. Overall, we have shown that GCAL is solvable in a short time for highly dimensional surveys, such as the German Census.

The performance of the SSN for GCAL for the example considered in the simulation study before ($n_s = 222\,433$; $q_1 + q_2 = 13\,392$, scenario SMP.rel(Aux&AxG) and GREG-type objective function) is shown in Table 5.5. In looking at the residuals in column 2, a locally quadratic convergence rate of SSN can be observed (cf. Qi and Sun, 1993). The convergence is slowed down in some iterations where the Armijo step-size rule reduces the normal step-size of 1.0. Without applying the step-size rule, the algorithm would not converge, i.e. the linear system of equation which has to be solved in each Newton step would not be solvable in iteration 2. Thus, the inclusion of the Armijo step-size rule is strongly necessary in order to achieve convergence. In addition, the convergence rate of the algorithm is quite sensible to the choice of the parameters of the step-size rule. Nevertheless, it is not very hard to find appropriate parameters for one example, but it is difficult to find a general set of parameters which is appropriate for all possible application. This is one of the most challenging tasks in developing a universally applicable software tool.

Table 5.5: Performance of the semismooth Newton algorithm for GCAL ($n_s = 222433$; $q_1 + q_2 = 13392$).

Iterations	Residual	Step-size	Objective	Lower	Upper
k	$\|\Psi(\lambda^k)\|$	α_k	$P(z(\lambda^k))$	bound	bound
0	$1.06 \cdot 10^{10}$	1.0000	0	0	0
1	$4.95 \cdot 10^{8}$	1.0000	$565.53 \cdot 10^{2}$	214	2
2	$1.04 \cdot 10^{7}$	0.2500	$613.49 \cdot 10^{2}$	261	1
3	$7.15 \cdot 10^{6}$	0.1250	$615.56 \cdot 10^{2}$	262	1
4	$5.81 \cdot 10^{6}$	0.0625	$614.11 \cdot 10^{2}$	260	1
5	$3.99 \cdot 10^{6}$	0.5000	$614.18 \cdot 10^{2}$	260	1
6	$1.27 \cdot 10^{6}$	0.0625	$614.14 \cdot 10^{2}$	261	1
7	$1.12 \cdot 10^{6}$	0.0312	$614.15 \cdot 10^{2}$	261	1
8	$9.98 \cdot 10^{5}$	1.0000	$614.08 \cdot 10^{2}$	261	1
9	$2.20 \cdot 10^{4}$	1.0000	$614.72 \cdot 10^{2}$	263	1
10	$6.60 \cdot 10^{-14}$	1.0000	$614.80 \cdot 10^{2}$	263	1

In column 4, the objective values are tabulated per iteration. The greatest changes are observed within the first iterations, i.e. the most significant changes to the correction weights are made within these iterations. From iteration 4 to 10, the alteration in the objective is only about one per mil of the value. This observation is consistent with columns 5 and 6 where the number of weights g_k that has reached the lower or upper bound are shown. As mentioned in Subsection 5.6.2, the number of weights that are equal to the lower bound is much higher, which is caused by the strongly skewed distribution of the penalties for the GREG-type objective function. The penalty for a weight $g_k = 0.2$ is $\frac{1}{2}(0.2-1)^2 = 0.32$, whereas the penalty for a weight $g_k = 5$ is $\frac{1}{2}(5-1)^2 = 8$. In examining the lower bound and iterations 3 and 4, the number of weights with $g_k = 0.2$ decreases from 262 to 260. Thus, the weights which reach the bound within one iteration may differ from those reaching it in the new iteration. In that regard, there is no monotonic behavior of the weights over the iterations. This will also be discussed in the next paragraph, where SSN is compared to TRUNC.

Comparison of SSN and TRUNC

In this paragraph, solving GCAL with the common truncated algorithm TRUNC (i.e. function calib in R package sampling; cf. Tillé and Matei, 2016) is compared to the solution with the SSN algorithm. SSN provably reaches the unique optimal solution λ^* of $\Psi(\lambda) = 0$ if the initial value is *good enough*, Ψ is semismooth, and $H \in \partial_B \Psi(\lambda^*)$ is regular (cf. Theorem 3.2.7). By contrast, it is not guaranteed that the solution of TRUNC is the optimal solution because TRUNC fixes and cuts off weights that reached the bounds such that a reduced problem is solved in the next iteration. Consequently, weights which have reached the bounds once are no longer considered for any adjustments in the iterations thereafter. This effect is exemplarily shown in Figure 5.12 for the scenario SMP.rel(Aux&AxG) and the GREG-type objective function. Solving this problem with TRUNC and SSN results in very similar but unequal solutions. The

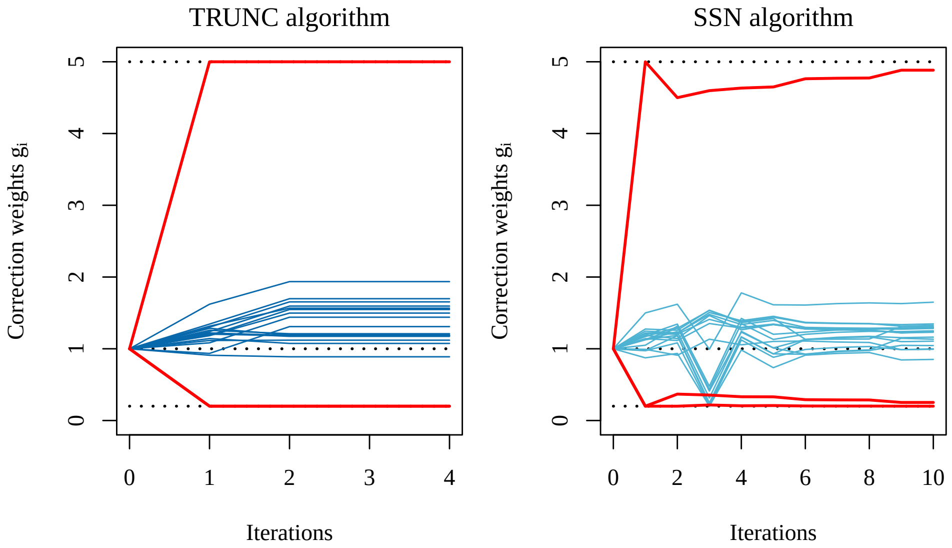


Figure 5.12: Comparison of performance of SSN and TRUNC for GCAL: plot of some correction weights g_k, which significantly differ from TRUNC to SSN ($n_s = 222433$; $q_1 + q_2 = 13392$).

majority of correction weights g_k show no discernable differences between TRUNC and SSN, but approximately 30 weights differ in more than 10% between SSN and TRUNC. The modification of these weights over the iterations is plotted in blue lines in Figure 5.12 (left indicates TRUNC, and right indicates SSN). Additionally, there are three weights (plotted in red) which are located on the bounds for TRUNC but not for SSN. These weights determine the different results of the algorithms. Using TRUNC they reach the bounds in iteration 1 and stay there for the other iterations, because they are not further considered after iteration 1. Using SSN, they also reach the bounds in iteration 1, but they leave the bounds in the next iterations, which results in different solutions. This yields different solutions for algorithms TRUNC and SSN with the solution of SSN being provably optimal.

The results of TRUNC and SSN and their performance are compared in Table 5.6. In looking at the objective values in column 2, the objective of SSN is slightly smaller than the objective of TRUNC, and at the same time, the accuracy of the compliance of the benchmarks is even higher for SSN (see column 1). Thus, the solution of SSN (which is optimal) is better than the solution of TRUNC. In this example, the differences are quite small and may not be relevant for the user, but the differences can be significant under other circumstances, especially if the box-constraints are strictly chosen. With regard to the last column, the computing time of SSN is 30 seconds, whereas TRUNC converges within 15 seconds. Even the computing time is application-specific and can also be the other way around, but the magnitude of the computational burden is similar for TRUNC and SSN.

Table 5.6: Comparison of performance of SSN and TRUNC for GCAL ($n_s = 222433$; $q_1 + q_2 = 13392$).

	Residual	Objective	Lower	Upper	Iterations	Time
	$\|\Psi(\lambda^k)\|$	$P(z(\lambda^k))$	bound	bound		in sec.
TRUNC	$1.55 \cdot 10^{-12}$	$615.40 \cdot 10^2$	265	2	5	15.77
SSN	$6.60 \cdot 10^{-14}$	$614.80 \cdot 10^2$	263	1	11	30.09

To summarize, even though the computational burden of the two algorithms is comparable, SSN provably computes the optimal solution of GCAL in contrast to TRUNC. The differences are quite small in the majority of applications tested, but they may also be higher in extreme cases. On the other hand, SSN (and especially the step-size rule) is more sensitive with regard to convergence issues. Nevertheless, the advantages of SSN over TRUNC outweigh the drawbacks and therefore an application of SSN can be recommended.

5.6.7 Sensitivity, issues, and limitations

Other distance functions

While choosing an appropriate distance function for GCAL, it is generally possible to use different functions for each weight g_k and deviation ϵ_j of the GCAL problem (5.8) since the distance function is only applied component-wise in the deviation of function Ψ in (5.13). Nevertheless, it is generally not sensible in practice to vary the distance functions from one weight to another. However, it might be useful to apply one distance function for the weights and another one for penalizing the epsilons.

As mentioned in Section 5.1, it might also be sensible to penalize a weight of $g_k = 0.5$ with the same penalty as $g_k = 2$, because doubling the weight may be handled equal to halving the weight. This feature is least given for the GREG-type distance function, more likely given for ML-Raking, but not fully given for any of the distance functions listed in Table 2.1. The distance function which fully fulfills this feature is

$$D(g_k) := \log(g_k)^2, \tag{5.23}$$

which is referred to as the *Balanced* distance function from this point onward. The comparison of the Balanced distance function to the other distance functions considering its derivatives and its values for $g_k = 0.5$ and $g_k = 2$ are shown in Table 5.7 and Figure 5.13. However, the application of the Balanced distance functions to GCAL is not possible in general because the derivative

$$D'(g_k) = 2\log(g_k)\frac{1}{g_k} \tag{5.24}$$

of the Balanced distance function is not monotonically increasing on $(0, \infty)$. The gradient of D' is zero in Euler's number $e \approx 2.71828$. Thus, the inverse D'^{-1} does not exist, the definition of (5.9) does not hold, and therefore the function Ψ defined in (5.13) and (5.14) cannot be built. Since D' is monotonically increasing on the interval $(0, e)$, D'^{-1} is well-defined if $U_{g_k} \leq e$ and $U_{\epsilon_j} \leq e$ hold for the upper bounds for all $k = 1, \ldots, n_\mathrm{s}$ and $j = 1, \ldots, q_2$ of the GCAL problem (5.8). Nevertheless, the derivation of the inverse of D' is not trivial in this case, since an equation consisting of a product of g_k and its logarithm $\log(g_k)$ has to be solved (reformulation of Equation (5.24)). This can be done by applying the Lambert W function (cf. Corless et al., 1996), also called omega function, which defines a relation for the inverse of a function $f(x) = xe^x$ for a real number $x \in \mathbb{R}$. For the Lambert W function, the following equation holds:

$$W(x)e^{W(x)} = x. \tag{5.25}$$

For more details, we refer to Corless et al. (1996). To summarize, the application of the Balanced distance function to GCAL is possible if $U_{g_k} \leq e$ and $U_{\epsilon_j} \leq e$ hold for all $k = 1, \ldots, n_\mathrm{s}$ and $j = 1, \ldots, q_2$. If the Balanced distance function is applied, the computational burden of SSN increases and the stability of the algorithm decreases. This occurs due to the behavior of the Balanced distance functions for the weights g_k being clearly smaller than 1.0 (as the gradient is very steep; see Figure 5.13). This would lead to ill-conditioned linear problems in the Newton step within each iteration of SSN, which may decelerate the convergence rate and may

Table 5.7: Analysis of distance functions for the calibration estimator.

	$D(g_k)$	$D'(g_k)$	$D(0.5)$	$D(2)$
GREG-type	$\frac{1}{2}(g_k-1)^2$	g_k-1	0.125	0.500
Raking	$g_k\log(g_k)-g_k+1$	$\log(g_k)$	0.153	0.386
ML-Raking	$g_k-1-\log(g_k)$	$1-\frac{1}{g_k}$	0.193	0.307
Balanced	$\log(g_k)^2$	$2\log(g_k)\frac{1}{g_k}$	0.480	0.480

raise the chances of instabilities. Hence, we have theoretically shown that the implementation of the Balanced distance function in GCAL is possible. However, we did not focus on this in the simulation study due to the issues concerning the numerical implementation.

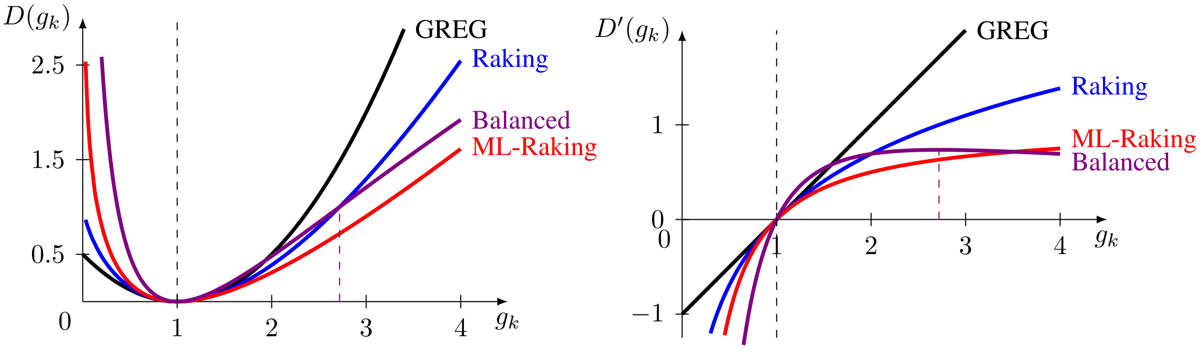


Figure 5.13: Analysis of distance functions for the calibration estimator (left: distance function D; right: derivative D' of distance function.

Sensitivity analysis via the Lagrangian multipliers

The functionality of SSN is discussed in detail in Section 3.2. The Lagrangian multipliers λ of the original optimization problem are iteratively updated, and they converge to a unique optimal solution λ^* of the corresponding nonlinear system of equations. Afterwards, the optimal solution of the optimization problem can be computed with λ^*. Each Lagrangian multiplier corresponds to one equality restriction of GCAL (i.e. one benchmark). Thus, the modifications of the multipliers along the iterations can be utilized to analyze the effort raised by the strict compliance with the restrictions (i.e. the level of difficulty to fulfill the restrictions). This is plotted in Figure 5.14, where values of 30 randomly selected Lagrangian multipliers are plotted for each iteration in comparison to the corresponding height of the deviation of the respective benchmark. Dark blue lines correspond to restrictions that have perturbations equal to the box-constraints. The light blue lines represent restrictions with a deviation smaller than the maximal defined tolerance.

In general, high values of λ_k corresponds to high perturbations of the benchmarks. This observation can be utilized to establish a sensitivity analysis. Firstly, if the algorithm SSN diverge in some settings, the analysis of the Lagrangian multipliers can contribute to assessing those benchmarks, which have a relaxation set too strict. Thus, a restart with weaker assumptions may yield to a convergence of the algorithm. Secondly, if the algorithm SSN converges but the variance of the resulting calibration weights w_k is too high (i.e. the tails of the density plots in Figure 5.3 are large), an observation of the Lagrangian multipliers may yield recommendations to slightly adjust the maximum perturbations in order to achieve calibration weights with a smaller variance.

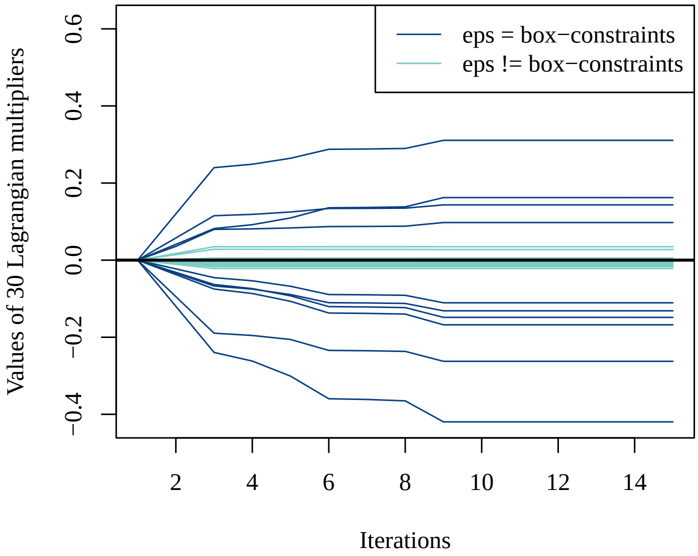


Figure 5.14: Behavior of selected Lagrangian multipliers over the iterations of SSN for GCAL (scenario SMP.rel(Aux&AxG)+GB).

Infeasibility and its prevention

If the number of restrictions is high, the risk of infeasibility issues in the GCAL problem (5.8) should not be neglected. In that regard, it is sufficient for an infeasible problem if at least two restrictions exclude each other. Then, the feasible set is empty, and the problem is unsolvable. Indeed, this problem has been observed in the considered simulation study. A very few samples have produced such infeasibilities. They have been replaced in order to prevent distorted simulation results. The risk of an empty feasible set increases when strict box-constraints for weights and/or relaxations are required. Unfortunately, the detection of such an unsolvable case is a priori almost impossible. In particular, the a priori identification of the causal restrictions is not practicable. A possible remedy is to weaken the box-constraints for *all* the relaxed benchmarks, i.e. to weaken L_{ϵ_j} and U_{ϵ_j} for all j. To ensure the compliance with the original strict benchmarks for most of the restrictions, the penalty parameters δ_j ($j = 1, \dots, q_2$) in the objective function of (5.4) may be raised. This yields an elevated penalty for a deviation of the benchmarks without shrinking the feasible set. Adjustment options of δ_j are therefore an opportunity to manage the strictness of the penalization for perturbed restrictions.

An example of the described issue is shown in Figure 5.15 for the same case of row 6 in Figure 5.5 (scenario `SMP.rel(Aux&AxG)+GB`, GREG-type objective function). The values of δ_j are varied around the standard case $\delta_j = 1\,300$, which has been computed depending on the number of units in the sample and the number of relaxed benchmarks. The six scenarios in Figure 5.5 are based on values of the factors 0.5, 1.0, 5.0, 10, 100, and $1\,000$ of the original δ_j (i.e. the second line corresponds with the original setting). It is observed that the deviations decreases with an increasing δ_j, since the influence of the corresponding penalty term in the objective function increases. Thus, it is possible to almost comply with the original box-constraints if these are weaken and the δ_j is increased. In particular, if `GCAL` is applied to a setting with an empty feasible set, the feasible set can be widened by weakening the box-constraints for the perturbations ϵ_j ($j = 1, \ldots, q_2$). To still enforce the compliance with the original box-constraints, the penalty parameters δ_j must be pushed up. In consequence, the algorithm converges (i.e. the feasible set is not empty at all) but the original box-constraints are still not exceeded for most of the restrictions, despite for those restrictions which have led to the infeasibility.

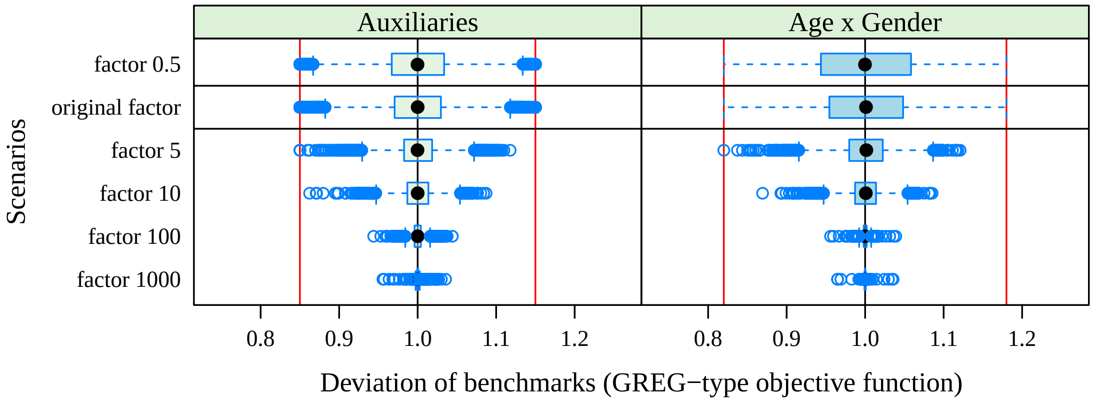


Figure 5.15: Variation of the penalty parameter for the relaxation to ensure feasibility of `GCAL` (scenario `SMP.rel(Aux&AxG)+GB`, $n_s = 222433$; $q_1 + q_2 = 13392$).

Overall, the described strategy is useful for preventing infeasibility issues. Nevertheless, the adjustment of δ_j affects the condition of the matrix, which has to be solved in each Newton step within `SSN`. Therefore, the strategy needs to be handled with care, i.e. the stability of `SSN` is only guaranteed for a limited adjustment of δ_j. Beside the issues concerning the numerical stability, the variance estimation via the rescaling bootstrap may also be influenced by an adjustment of δ_j. Some investigations have shown, that higher values of δ_j tend to enhance the slight underestimations observed for some auxiliary variables in Subsection 5.6.5.

5.7 Summary and discussion

The generalized calibration method GCAL determines calibration weights for calibration estimators while considering several requirements, such as a large number of benchmarks on various stratification levels, which are possbily obtained from different sources, the use of box-constraints, and the control of the Gelman bound. The general framework of GCAL was presented in Section 5.2, and the numerical solution strategy of such generalized calibration methods was proposed Section 5.3 based on the approaches of Münnich et al. (2012b) and Wagner (2013). To quantify the quality of estimates, a strategy for the variance estimation based on a rescaling bootstrap in proposed in Section 5.4, which we also dealt with in Burgard et al. (2018).

The relevance of GCAL for European official statistics can be exemplarily substantiated by two fundamental surveys in Germany, namely the *German Census* (cf. Münnich et al., 2012a) and the *German System of Household Surveys* (cf. Riede et al., 2013). The German Census in 2011 has been a stratified random sample of all addresses in Germany with an overall sampling fraction of approximately 9%. The next German Census will be conducted in 2021 using a similar methodology. Its main goal is to deliver a statistic regarding the total of all inhabitants of Germany, the 16 federal states, and certain smaller regions. Since auxiliary data is available from registers and other surveys, calibration may lead to a significant improvement of the quality of the survey estimates. The diversity of the auxiliary variables (in relation to the various sources, stratification levels, quality, and relevance) necessitates a flexible generalized calibration method. As we have presented in the simulation study in Section 5.6, GCAL is able to fulfill these multiple requirements. Next to the German Census, the German System of Household Surveys may also benefit from GCAL. By 2021, the System of Household Surveys will be conducted as a renewed integrated system consisting of several surveys, such as the Micro-Census, the Labor Force Survey (LFS), the European Union Statistics on Income and Living Conditions (EU-SILC), and other surveys (cf. Riede et al., 2013). The integration of several surveys requires special attention with regard to the coherence and consistency of the results. Due to high sample sizes, it is of great importance to solve GCAL even for large problem instances in appropriate time, especially if an MSE estimation using resampling methods need to be applied. This is possible due to the solution strategy via the SSN algorithm, in particular due to the linear dependency between computing time and total sample size.

Chapter 6

Conclusion and Outlook

The main developments of the thesis are concluded in this chapter. Advantages as well as drawbacks are stated and elucidated in the context of survey statistics in modern societies. Moreover, further outstanding and unprocessed research potentials are shortly discussed in an outlook. Finally, further potential scopes of application beyond survey statistics are explored.

6.1 Conclusion

As mentioned in the introduction, the complexity of statistical models has steadily increased due to new challenges and requirements. This is often accompanied by an enhanced analytical and computational expense. In order to consider these issues, to maintain practicability, and to improve the computational performance, this thesis addressed two main research topics concerning the generation of multivariate and multi-domain statistics, namely an optimal multivariate and multi-domain allocation method (`MMDopt`) as well as a generalized calibration method (`GCAL`). Aside from the statistical modeling, innovative numerical optimization strategies have been developed, and existing approaches have been expanded further.

One of the principal objectives of this work is to allow more flexibility with regard to the inclusion of known auxiliary data. Since auxiliary data is a strong instrument for increasing the quality and ensuring the consistency of estimates, its usage is required in almost all applications. In order to handle the respective data with regard to its sources, structures, and quality, several extensions of established methods have been developed. To consider multivariate auxiliary data, the `MMDopt` allocation method deals with possibly conflicting objectives on various stratification levels based on results of the theoretical analysis of multi-criteria optimization. The calibration method `GCAL` enables the use of a huge amount of auxiliary data and satisfies the quality of the data with optionally includable individual relaxations. A further key objective is the compliance with restrictions for various stratification levels, possibly based on legal regulations. `MMDopt` allows for the compliance with both quality restrictions for regional estimates as well as lower and upper bounds for the stratum-specific sample sizes. Due to the possibility of the relaxation of benchmarks in `GCAL`, even benchmarks for small regions may be regarded, which would otherwise lead to non-feasible problems. Moreover, additional assumptions have been considered, e.g. the opportunity for the control of the Gelman bound in `GCAL`.

In addition to the statistical properties, the numerical implementation is a key objective of the thesis. If possible, the application of the developed methods should be realizable with an ordinary desktop PC in an appropriate time, which allows for a wide application range, such as applications in national statistical offices, in independent institutes, in the industry, and also by individual private users. The numerical solution strategies are mainly based on developments of Münnich et al. (2012b), Gabler et al. (2012), and Münnich et al. (2012c). In that regard, the strategies take advantage of a specific transformation of the optimality conditions. This results in significantly lower dimensional problems, which can be solved efficiently by applying the semismooth Newton method for `MMDopt` and `GCAL`. To emphasize the performance, problems of several ten-thousand variables can be solved within one second. Moreover, the computing time is only linearly dependent on the dimension of the original optimization problem, with a result that the exponential increase of the running time depending on the problem size is omitted. Beside the semismooth Newton method, a projected inexact quasi-subgradient method is developed to solve a specific class of `MMDopt` problems, that do not fulfill the assumptions to apply the semismooth Newton method.

Finally, the statistical flexibility and versatility as well as the efficient solution strategy allows a straightforward application of `GCAL` and `MMDopt`. Official statistics may especially obtain significant benefits using both methods, since the requirements for modern surveys have increased (cf. Münnich et al., 2012a and Riede et al., 2013). Firstly, efficient estimates may be gained for the population level as well as for certain regional levels. In that regard, potential negative outliers can be omitted by both including restrictions for regional efficiency in `MMDopt` or applying relaxed benchmarks in `GCAL`. Secondly, coherence and consistency between various sources such as surveys, censuses, and registers may be guaranteed without neglecting disruptive factors, for instance inaccurate survey data and incorrect register data. Thirdly, surveys with several possibly conflicting goals can be conducted in order to achieve accurate estimates for all variables of interest, while utilizing various well-established decision-making functions.

Conveniently, both methods are developed to gain efficient estimates under similar circumstances, i.e. under a huge amount of auxiliary data, with regard to several conflicting objectives, and under consideration of different stratification levels. With regard to the survey process, the `MMDopt` method aims at improving the selection process by an optimal allocation of the total sample size to the strata. By contrast, `GCAL` is applied to generate accurate and coherent estimates in the estimation process. Thus, it may be sensible to apply both methods subsequently in one survey in order to exploit all the benefit from both methods.

6.2 Outlook

In this work, we focused on the theoretical foundation of the methods on the one hand, and on the analysis of their functionality and the resulting statistical quality on the other hand. The `MMDopt` allocation method is based on a stratified population. The assumption of predefined and fixed strata has been compulsory in this thesis. Indeed, this assumption is sensible, since all the applications considered and the majority of possible applications in official statistics are based on predefined strata. Often times, these are determined by natural or administrative circumstances, such as a regional structuring of countries or content-related classifications of the population. These predefinitions have an impact on the resulting estimates. Thus, a smart adjustment of the definition of the strata may improve the results of a survey. This is known as *optimal stratification* in the literature and has been primarily suggested by Dalenius and Hodges (1959) and Wright (1983). The inclusion of a strategy of optimal stratification can be identified as an opportunity for improvement of the MMDopt method.

Although this thesis focuses on applications of household surveys, MMDopt and GCAL are also applicable in further areas of survey statistics such as biogeographic applications like agricultural or biological surveys. Examples can include the vegetable survey[1] or other animal counting surveys. Moreover, both developed methods may be usable with regard to the processing of a huge amount data gained by geographical localization using satellite data, which will be a raising research field within in the next years. In addition, several applications beyond the horizon of survey statistics are possible. One example is the field of developing automated driving and driver assistance systems. In the course of this type of development, a huge amount of data is generated within extremely short time periods by several sensors, which needed to be processed and analyzed. MMDopt may yield support to choose the important parts of the data. Another potential field of application is the wide field of economics. Both developed methods may assist in data processing and analyzing if a huge amount of data is available. The allocation method MMDopt may provide support in how to choose and sample the relevant data from the overall volume of data. Moreover, GCAL may support the use auxiliary data and paradata (information about the process by which the data were collected) to improve process structures and services offered.

To conclude, both developed methods MMDopt and GCAL contribute to improve the quality of statistics gained by means of a survey. The improvement of the accuracy of the estimates and the obtaining of consistent results are accompanied by a high flexibility with regard to the input data and the determination of restrictions. In addition to these statistical advantages, highly efficient numerical solvers allow users to apply the outlined methods in an appropriate time with a common IT infrastructure.

[1] https://www.destatis.de/DE/Publikationen/Thematisch/LandForstwirtschaft/ObstGemueseGartenbau/Gemueseerhebung2030313177004.pdf;jsessionid=8E086CD7915F866FEDAB4E7D93EE3E48.InternetLive2?__blob=publicationFile

Appendix A

Quality Measurement in Survey Statistics

In order to analyze the functionality of the developed statistical methods, applications and simulation studies are generally conducted to measure the quality of the estimates obtained from surveys by means of the developed methods. These studies are mostly based on datasets, which may be either synthetically generated or based on real observations. Thus, the true values ϑ_d of a (sub-)population $\mathcal{U}_d \subseteq \mathcal{U}$, which are supposed to be estimated by means of the survey, are known and can be compared with the estimated values $\hat{\vartheta}_d$. With respect to the quality of the *point estimates*, two types of quality measures basically have to be considered. On the one hand, the estimates should not systematically under- or overestimate the true values (*bias*, cf. Equation (2.11)) and, on the other hand, they should not exhibit a large variation (*variance*, cf. Equation (2.9)). Both the bias and the variance of the estimate $\hat{\theta}_d$ are derived using the expected value of $\hat{\theta}_d$, as defined in Equation (2.8). As mentioned in Chapter 2, the efficiency of the estimate $\hat{\theta}_d$ is generally measured by the mean squared error (*MSE*, cf. Equation (2.12)), which is defined as the sum of the variance and the squared bias

$$\text{MSE}(\hat{\vartheta}_d) := \text{Var}(\hat{\vartheta}_d) + \text{Bias}(\hat{\vartheta}_d)^2. \tag{A.1}$$

The evaluation of the MSE may follow one of two main approaches, which are described in the following two paragraphs.

If an estimator is applied which is design-unbiased and for which the variance can be determined with the aid of a closed formula (e.g. for the HT estimator) the MSE can be computed straight forward via Equation A.1 without conducting repeated Monte-Carlo strategies. This is the case for the analysis of the `MMDopt` method in Chapter 4, since `MMDopt` only aims at optimizing the allocation of the total sample size to the strata in a stratified sampling design. Thus, the developments are only referred to the design-stage and in particular to the computation of the design weights in StrRS. As the variance of the HT estimator in StrRS can be computed by (2.36) and its bias is supposed to be zero per definition (cf. Särndal et al., 1992, Section 2.8), the MSE of the estimate is equal to its variance. As a result of this, the performance of `MMDopt` is analyzed in an application study (without Monte-Carlo simulations) in Section 4.6. However, is needs to be noted that this strategy does not allow for the analysis of the *spread* of the estimates concerning repeatedly drawn samples. Nevertheless, this primarily depends on the definitions of the HT estimator and the StrRS design, whose analyses are not a subject of the thesis.

If the object of investigation is related to the structure of the estimator, it would not be possible to determine the variance, the bias, and the MSE using the aid of a closed formula. This is the case for the `GCAL` method proposed in Chapter 5, as it optimizes the adjustment of the calibration weights for the calibration estimator (2.38). Hence, `GCAL` has influences on the estimation. As a consequence, the quality of `GCAL` can only be observed by conducting a Monte-Carlo simulation study, in which R_{MC} estimates related to R_{MC} independently drawn samples are computed. As a result of the simulation study, the MC-bias of $\hat{\vartheta}_d$ can be computed by

$$\text{BIAS}_{\text{MC}}(\hat{\vartheta}_d) := \frac{1}{R_{\text{MC}}} \sum_{r=1}^{R_{\text{MC}}} \left(\hat{\vartheta}_{d,r} - \vartheta_d\right), \tag{A.2}$$

where $\hat{\vartheta}_{d,r}$ is the resulting estimate for ϑ_d in replication $r = 1, \dots, R_{\text{MC}}$. The MC-bias BIAS_{MC} is an approximation of the analytical bias (2.11). Due to the law of large numbers, BIAS_{MC} converges to the bias for $R_{\text{MC}} \to \infty$. The speed of convergence strongly depends on the considered method or estimator. As a result, the MC-bias computed by Equation (A.2) will generally not be equal to zero in cases where the unbiasedness of the estimator is given per definition. Analogously, the MC-MSE is an approximation of the MSE (A.1) and is given by

$$\text{MSE}_{\text{MC}}(\hat{\vartheta}_d) := \frac{1}{R_{\text{MC}}} \sum_{r=1}^{R_{\text{MC}}} \left(\hat{\vartheta}_{d,r} - \vartheta_d\right)^2. \tag{A.3}$$

Since the variance, the MSE, and the bias are each dependent on the scale of the respective variable, they are often defined in relation to the true value ϑ_d. The corresponding relative measures are introduced in the following for the Monte-Carlo case, whereas the formulas for the analytical case can be defined in analogy. The relative bias and the relative MSE are given by

$$\text{RBIAS}_{\text{MC}}(\hat{\vartheta}_d) := \frac{1}{R_{\text{MC}}} \sum_{r=1}^{R_{\text{MC}}} \frac{\left(\hat{\vartheta}_{d,r} - \vartheta_d\right)}{\vartheta_d} \tag{A.4}$$

and

$$\text{relMSE}_{\text{MC}}(\hat{\vartheta}_d) := \frac{1}{R_{\text{MC}}} \sum_{r=1}^{R_{\text{MC}}} \frac{\left(\hat{\vartheta}_{d,r} - \vartheta_d\right)^2}{\vartheta_d^2}, \tag{A.5}$$

respectively. In addition, the relative root MSE (RRMSE) is often considered, given by

$$\text{RRMSE}_{\text{MC}}(\hat{\vartheta}_d) := \sqrt{\frac{1}{R_{\text{MC}}} \sum_{r=1}^{R_{\text{MC}}} \frac{\left(\hat{\vartheta}_{d,r} - \vartheta_d\right)^2}{\vartheta_d^2}}. \tag{A.6}$$

Generally, the use of the relative values enables a comparison between different variables. Nevertheless, an issue arises, if the denominator of (A.4), (A.5), and (A.6) (i.e. the known value) is close to zero. In this case, even an extremely small absolute bias and MSE can result in immense relative values. Under these circumstances, it may be preferable to analyze the absolute bias and MSE, instead. Basically, MSE, relMSE, and RRMSE can take non-negative values. In this way, a value close to zero suggests an efficient estimate and is therefore most desirable. By contrast, the BIAS and RBIAS can take positive and negative values. Moreover, a value of zero corresponds to an unbiased estimate. For unbiased estimators and a sufficiently large number of replications, the MC-RRMSE will be very close to the relative standard error. This is a criterion often used in official statistics, since in general, only results with a relative standard error less than a predefined maximum are allowed to be published.

Up to now, the proposed quality measures are based on a known population since the true values ϑ_d are utilized to compute the respective measures. The true values are unknown in practice, as their estimation is the aim of the survey. In this case, the variance or the MSE of the estimate need to be estimated either directly using computed linearization techniques or indirect using resampling methods (see Section 5.4 for more details). Thus, a variance or MSE estimator is determined to quantify the quality of the point estimates, in particular its variance or MSE, even if the real values ϑ_d are unknown. Therefore, a precise statement on the quality of the point estimates in practical applications depends on the precision of the variance or MSE estimates. In order to quantify this precision, the variance estimates are compared with the real variances or MSEs in a simulation study.

Appendix B

Additional Material of Application and Simulation Studies

B.1 Table of variables within the RIFOSS dataset

A list of all applied variables of the RIFOSS dataset in the original form on person level is given in Table B.1 with a brief description based on the description published in the data manual of the German Microcensus 2008[1]. Thereafter, the variables are suitably transformed to mostly (quasi-) continuous variables in the household structure. The resulting variables are tabulated in Table B.2 with a short explanation of the process of their generation and its type. The expression *continuous* (in italics) is referred to variables which can be characterized as *quasi-continuous*, i.e. discrete variables with many different values.

Table B.1: List of variables within the RIFOSS dataset on person level (original data of the dataset).

name	type	description
EF44	*continuous*	age of the respective person
inc_unemp5	continuous	unemployment benefits
hpw1	continuous	working hours
inc_w5	continuous	total income
inc_pen5	continuous	total retirement pension
AGS_neu	character	municipality key
HID	character	household ID
ADR	character	address ID (sequential number)
EF3_mod2	character	number of selection area
SMP	character	sampling point
EF29	categorical	type of acquisition
EF46	categorical	gender
EF49	categorical	family status
EF117	categorical	occupational status

[1] http://www.forschungsdatenzentrum.de/bestand/mikrozensus/suf/2008/fdz_mz_suf_2008_schluesselverzeichnis.pdf

EF310	categorical	highest school graduation
EF312	categorical	highest professional school graduation
EF540	categorical	ISCED level (education)

Table B.2: List of variables within the RIFOSS dataset on household level (input variables for `MMDopt` and `GCAL`; calculation basis in brackets behind description).

name	type	description
ZEN	*continuous*	Number of persons living in household
SEX1 / SEX2	*continuous*	Number of male / female persons living in household (EF46)
AGE_INC	*continuous*	age of main income earner of household (EF44, inc_w5)
HAGE	*continuous*	age of oldest person of household (EF44)
MAGE	*continuous*	mean age all persons of household (EF44)
medianAGE	*continuous*	median age all persons of household (EF44)
AGE4.1 - AGE4.4	*continuous*	number of persons in household of age 0 to 19, 20 to 39, 40 to 59, and ≥ 60 (EF44)
AGE7.1 - AGE7.7	*continuous*	number of persons in household of age 0 to 14, 15 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and ≥ 65 (EF44)
AGE4.1_Sex.1 - AGE4.4_Sex.2	*continuous*	male / female number of persons in household of age 0 to 19, 20 to 39, 40 to 59, and ≥ 60 (EF44)
AGE7.1_Sex.1 - AGE7.7_Sex.2	*continuous*	male / female number of persons in household of age 0 to 14, 15 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and ≥ 65 (EF44)
ALG	continuous	total unemployment benefits of household (sum of inc_unemp5)
ASTD	continuous	total working hours of household (sum of hpw1)
INC	continuous	total income of household (sum of inc_w5)
INC.PP	continuous	mean income per person of household (inc_w5, ZEN)
EDI	continuous	equivalized disposable household income (inc_w5, ZEN, EF44)
PEN	continuous	total retirement pension (inc_pen5)
ABSCHL1 - ABSCHL4	*continuous*	number of persons in household with prof. school graduation (EF312 $\in \{1,2,3,12\}$, $= 4$, $\in \{5,6,7,8,11\}$, and $\in \{9,10\}$)
SCHUL1 - SCHUL3	*continuous*	number of persons in household with school graduation (EF310 $\in \{1,6,7\}$, $\in \{2,3\}$, and $\in \{4,5\}$)
EF117A, EF117B, EF117S	*continuous*	number of persons in household with occupational status (EF117 $\in \{1,2,3\}$, $\in \{4,5,6,9\}$, and $\in \{7,8,10,11,12,13\}$)
FAM1 - FAM4	*continuous*	number of persons in household with family status (EF49 $= 1$, $= 2$, $= 3$, and $= 4$)
ISCEDA - ISCEDD	*continuous*	number of persons in household with ISCED level (EF540 < 4, $\in \{4,5\}$, $= 6$, and ≥ 7)
ILO1 - ILO4	*continuous*	type of acquisition (EF29$= 1$, $= 2$, $= 3$, and $= 4$)
AGS	character	municipality key (AGS_neu)
HID	character	household ID (HID)
ADR	character	address ID (ADR)
FS	character	name of federal state (AGS_neu)
NUTS2	character	name of NUTS2 region (AGS_neu)
NUTS3	character	name of NUTS3 region (AGS_neu)
HHS	character	name of class of household sizes (ZEN)
SMP	character	name of sampling point (SMP)
strata	character	name of cross-classification strata (SMP, HHS)

B.2 Optimal multivariate and multi-domain allocation

In this section, additional figures are shown for the evaluation of the application study for the allocation method `MMDopt` in Section 4.6. The plotted scenarios resemble the figures of Section 4.6, whereas the results of estimates for other stratification levels are shown here.

Outline of the following figures:

- Figure B.1: RRMSE for ten combinations of weights.
- Figure B.2: Relative increase of the RRMSE (heatmaps).
- Figure B.3: RRMSE for selected allocation strategies.
- Figure B.4: Relative change of RRMSE depending on the decision-making function.
- Figure B.5: RRMSE with restrictions for regional efficiency.

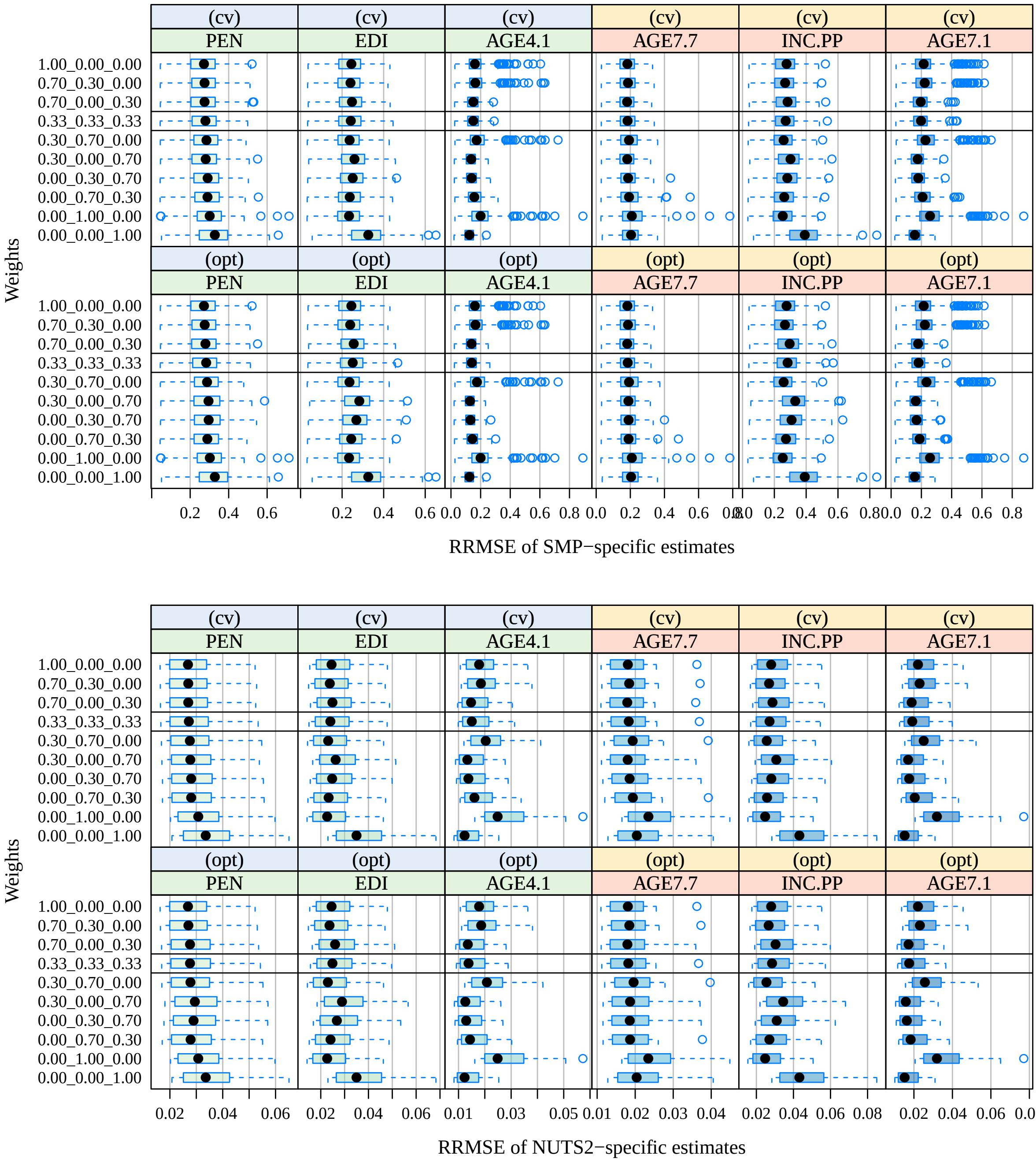
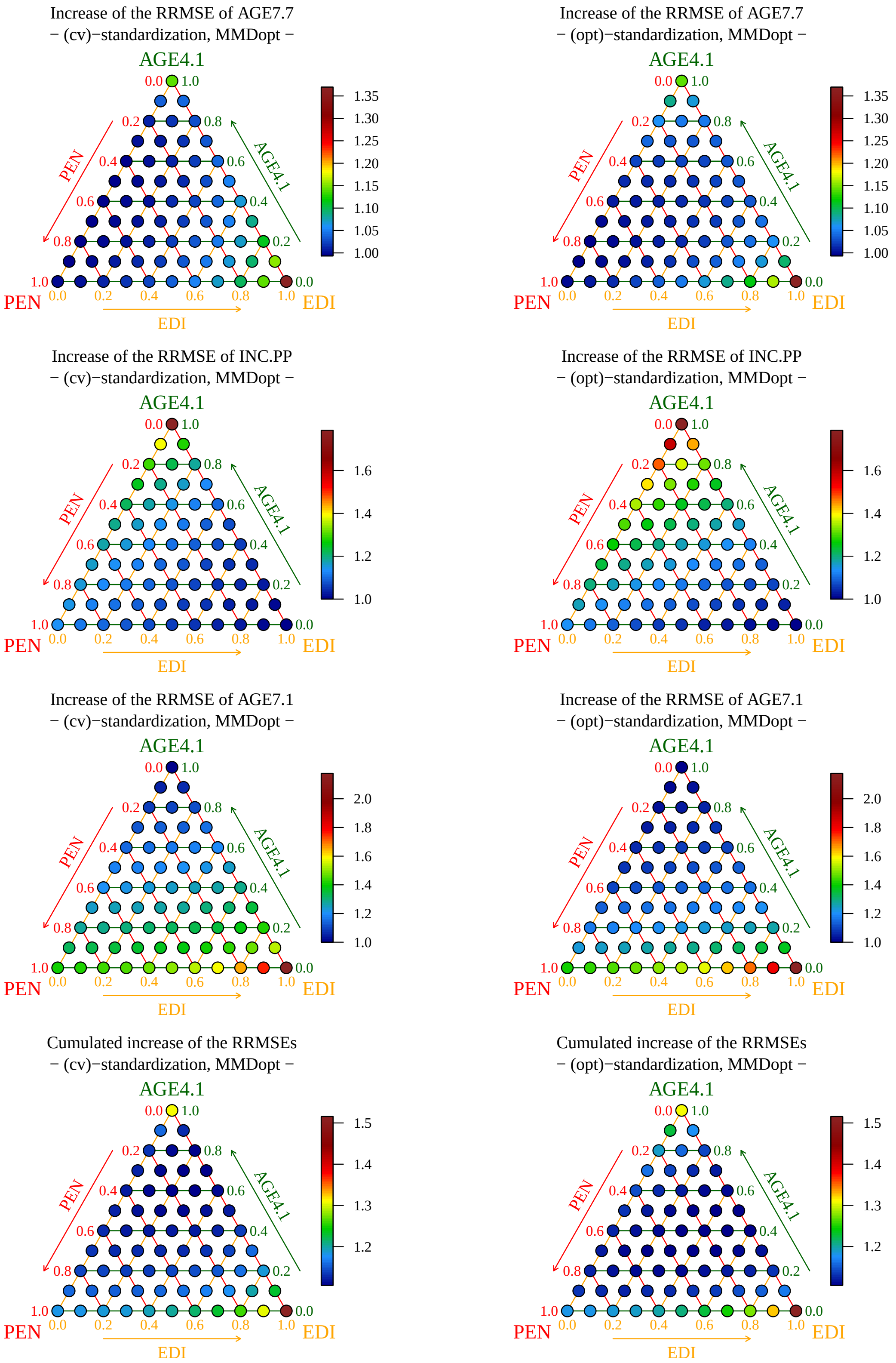


Figure B.1: RRMSE of SMP- and NUTS2-specific total estimates for ten combinations of weights with (cv)- and (opt)-standardization.



Increase of the RRMSE of AGE7.7
− (cv)−standardization, MMDopt −

Increase of the RRMSE of AGE7.7
− (opt)−standardization, MMDopt −

Increase of the RRMSE of INC.PP
− (cv)−standardization, MMDopt −

Increase of the RRMSE of INC.PP
− (opt)−standardization, MMDopt −

Increase of the RRMSE of AGE7.1
− (cv)−standardization, MMDopt −

Increase of the RRMSE of AGE7.1
− (opt)−standardization, MMDopt −

Cumulated increase of the RRMSEs
− (cv)−standardization, MMDopt −

Cumulated increase of the RRMSEs
− (opt)−standardization, MMDopt −

Figure B.2: Relative increase of the RRMSE of the population total estimates under (cv)- and (opt)-standardization for 66 combinations of weights for each variable of interest.

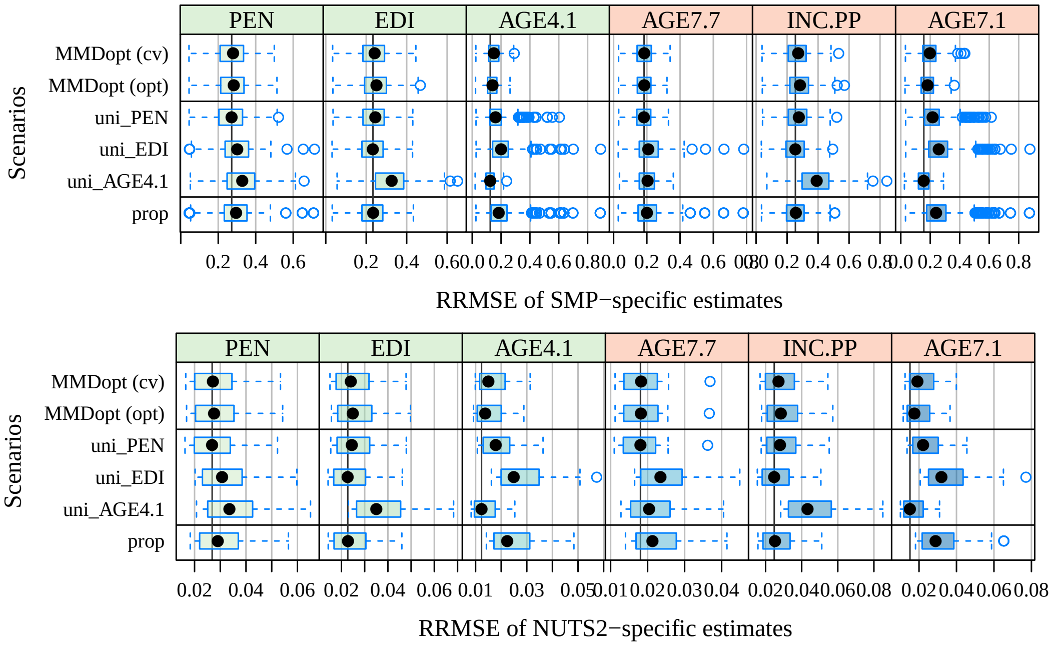


Figure B.3: RRMSE of SMP- and NUTS2- specific total estimates for selected allocation strategies.

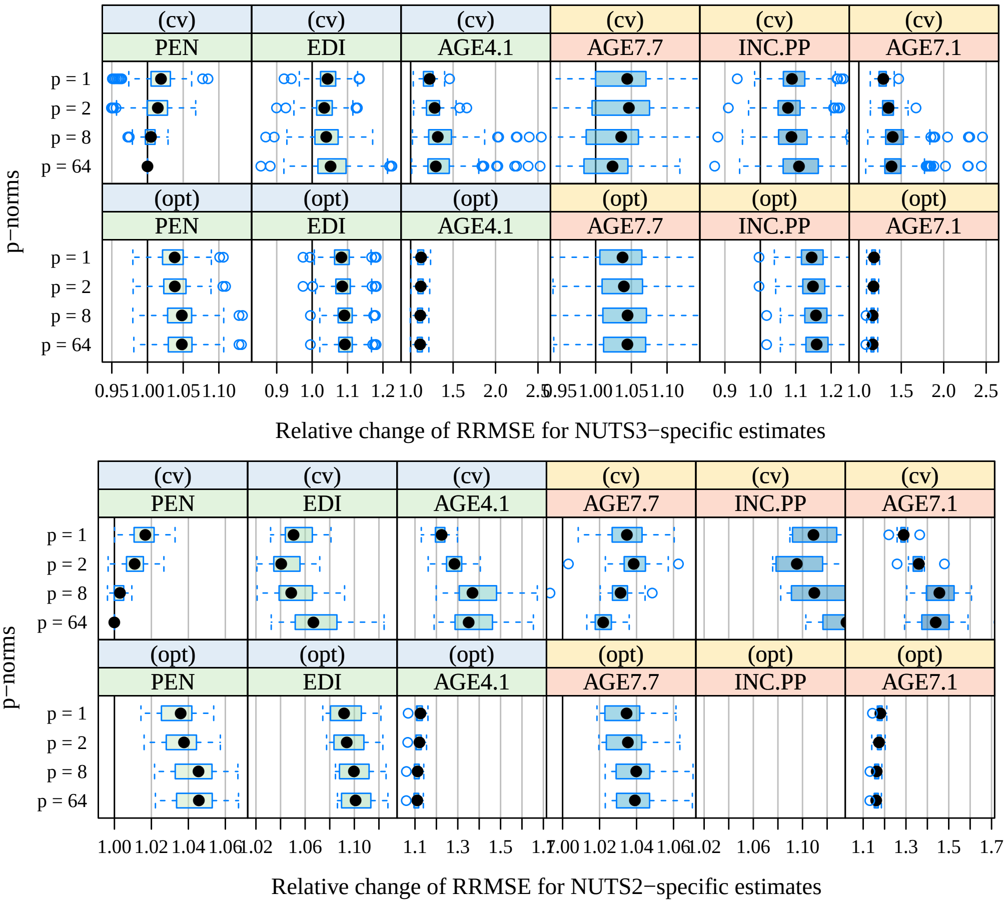


Figure B.4: Relative change of RRMSE of NUTS3- and NUTS2-specific estimates depending on the decision-making function.

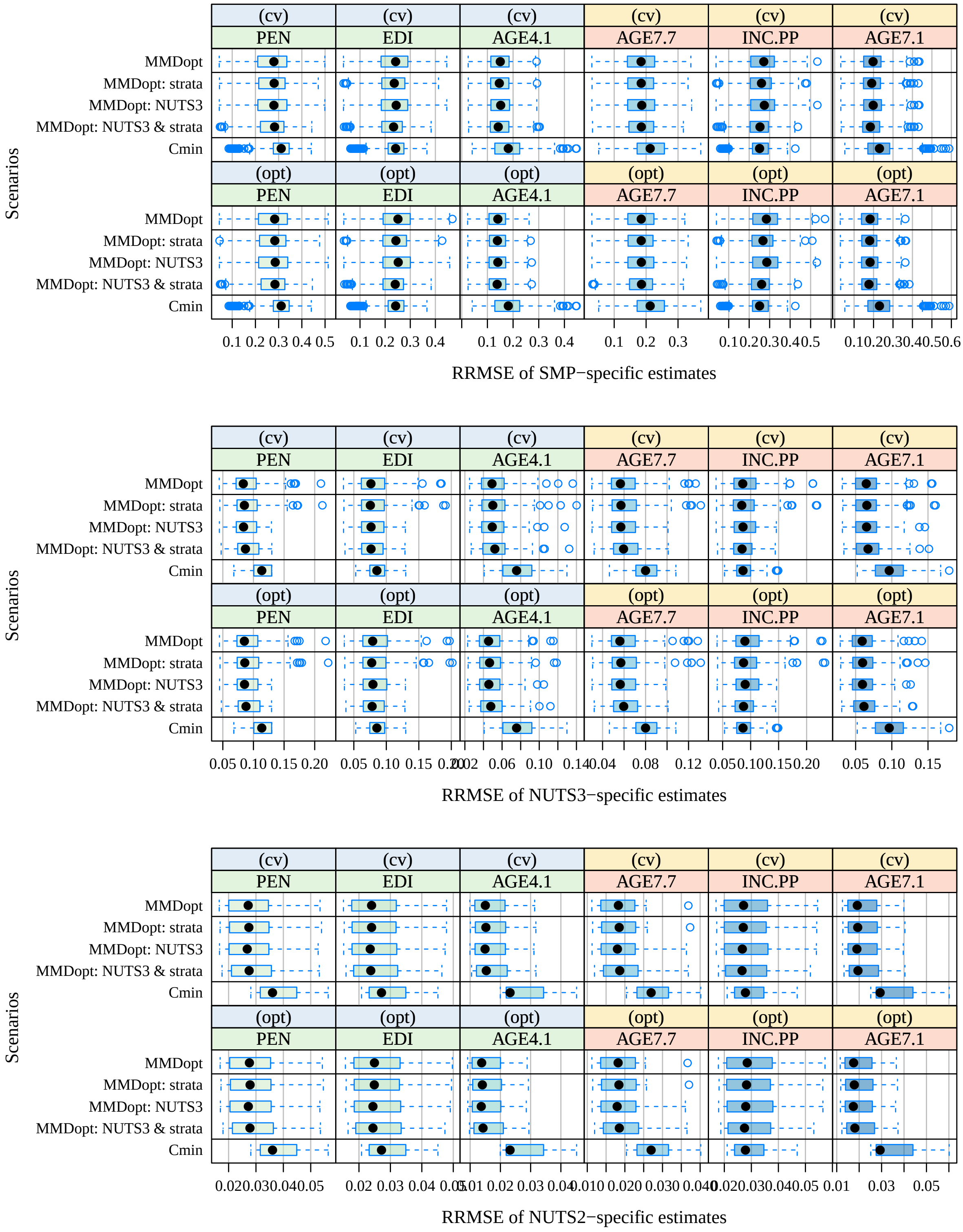


Figure B.5: RRMSE of SMP-, NUTS3, and NUTS2-specific estimates with restrictions for regional efficiency.

B.3 A generalized calibration method

In this section, additional figures are shown for the evaluation of the simulation study in Section 5.6 concerning `GCAL`. The figures show the RRMSE and RBIAS of the point estimates as well as the RBIAS of the variance estimates computed by the rescaling bootstrap. In that regard, it is distinguished between the three distance functions (see Table 2.1), the six scenarios tabulated in Table 5.1, and variables of interest and auxiliaries (visible by the shaded headers of the panels).

Outline of the following figures:

- Figure B.6: RRMSE of SMP-specific point estimates for the three distance functions.
- Figure B.7: RRMSE of NUTS3-specific point estimates for the three distance functions.
- Figure B.8: RRMSE of NUTS2-specific point estimates for the three distance functions.

- Figure B.9: RBIAS of SMP-specific point estimates for the three distance functions.
- Figure B.10: RBIAS of NUTS3-specific point estimates for the three distance functions.
- Figure B.11: RBIAS of NUTS2-specific point estimates for the three distance functions.

- Figure B.12: RBIAS of SMP-specific variance estimates for the three distance functions.
- Figure B.13: RBIAS of NUTS3-specific variance estimates for the three distance functions.
- Figure B.14: RBIAS of NUTS2-specific variance estimates for the three distance functions.

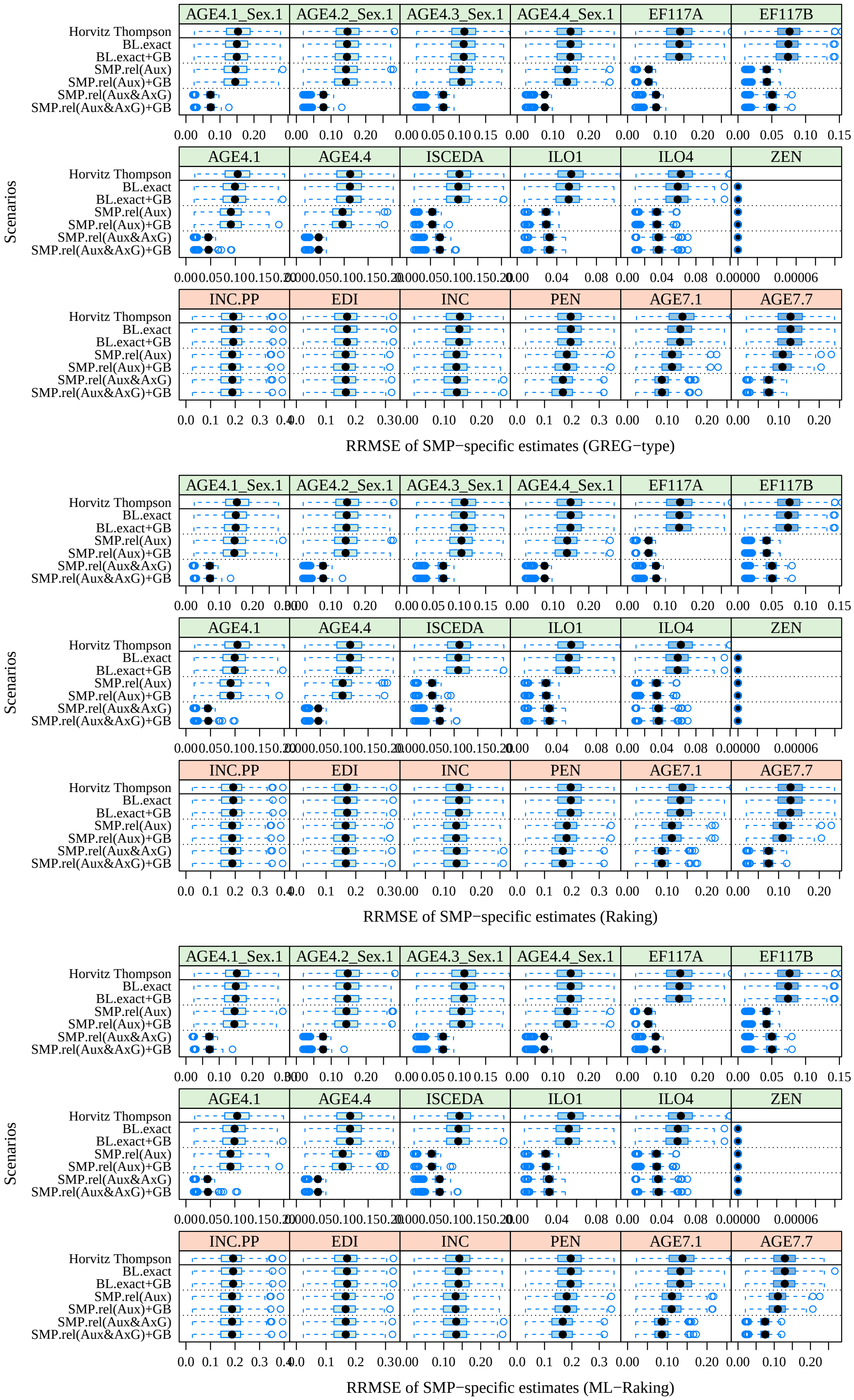


Figure B.6: RRMSE of SMP-specific point estimates.

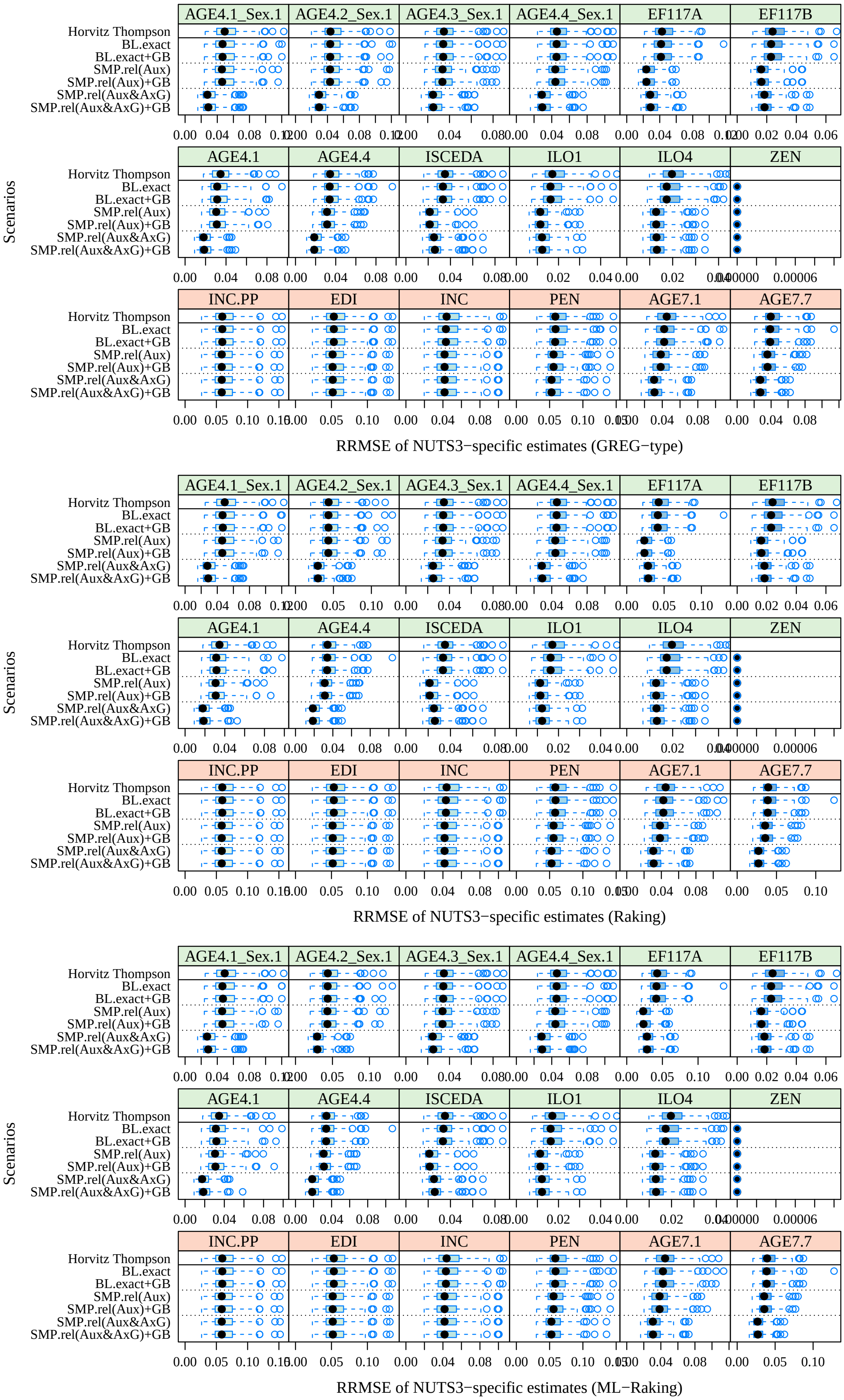


Figure B.7: RRMSE of NUTS3-specific point estimates.

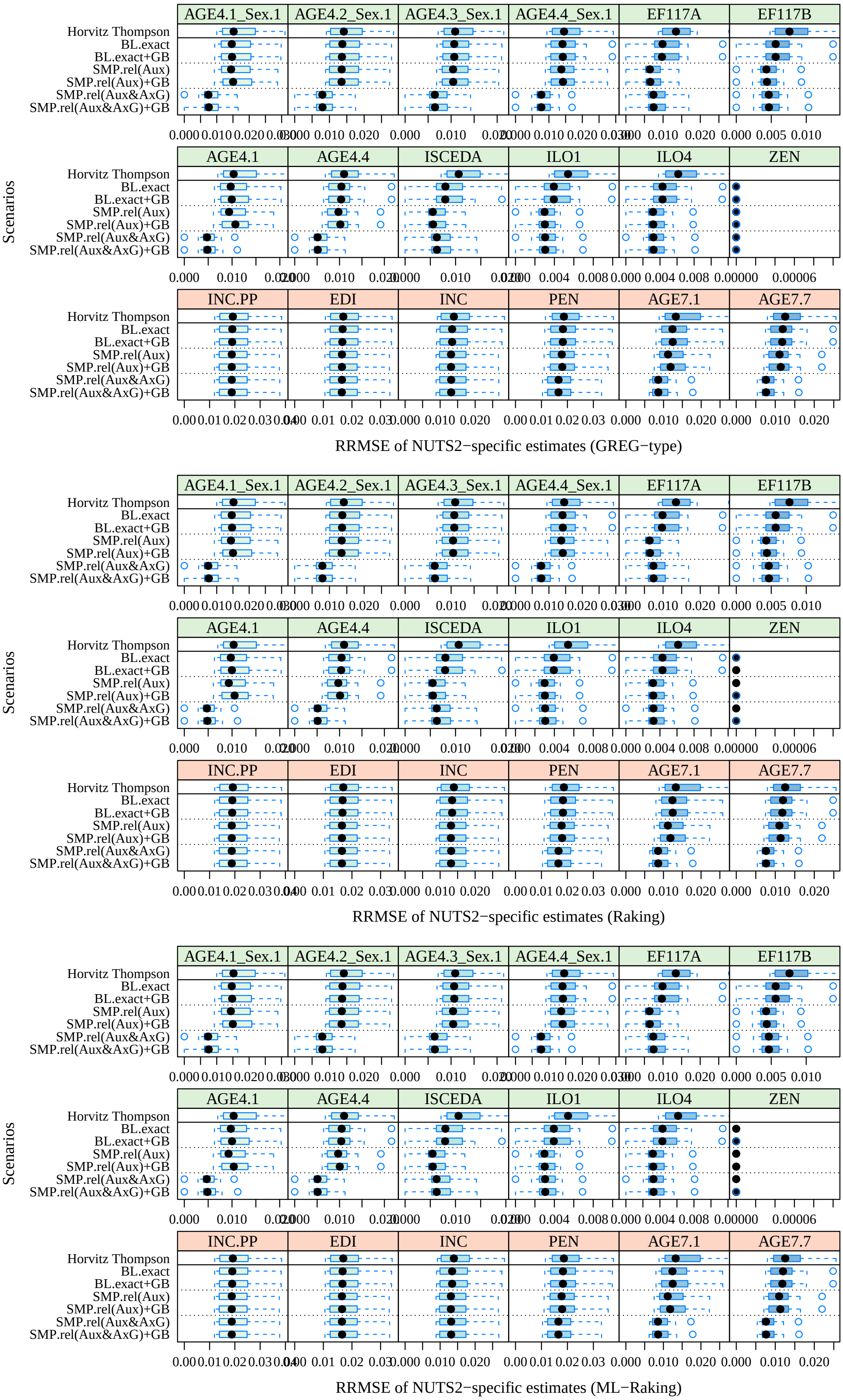


Figure B.8: RRMSE of NUTS2-specific point estimates.

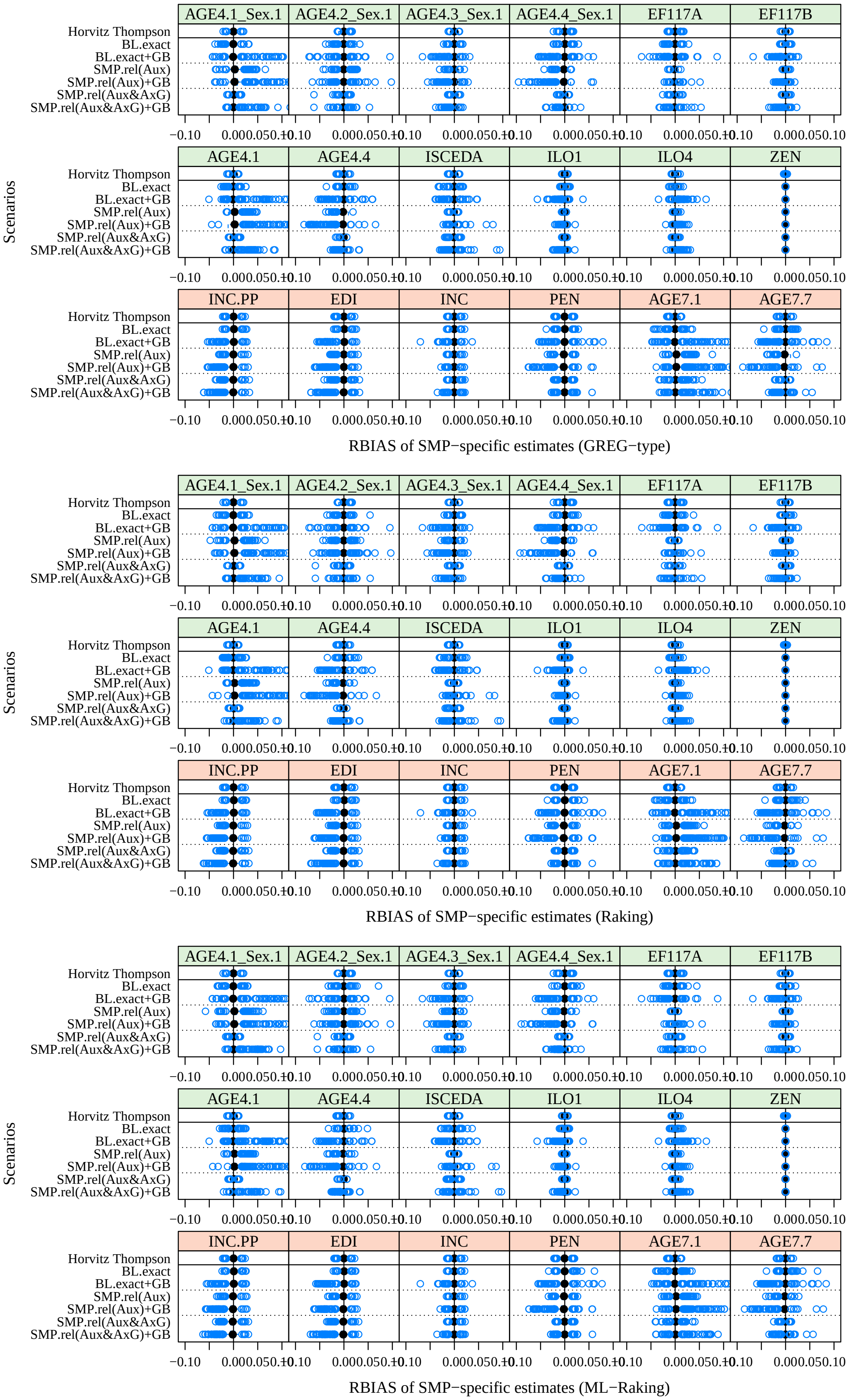


Figure B.9: Relative bias of SMP-specific point estimates.

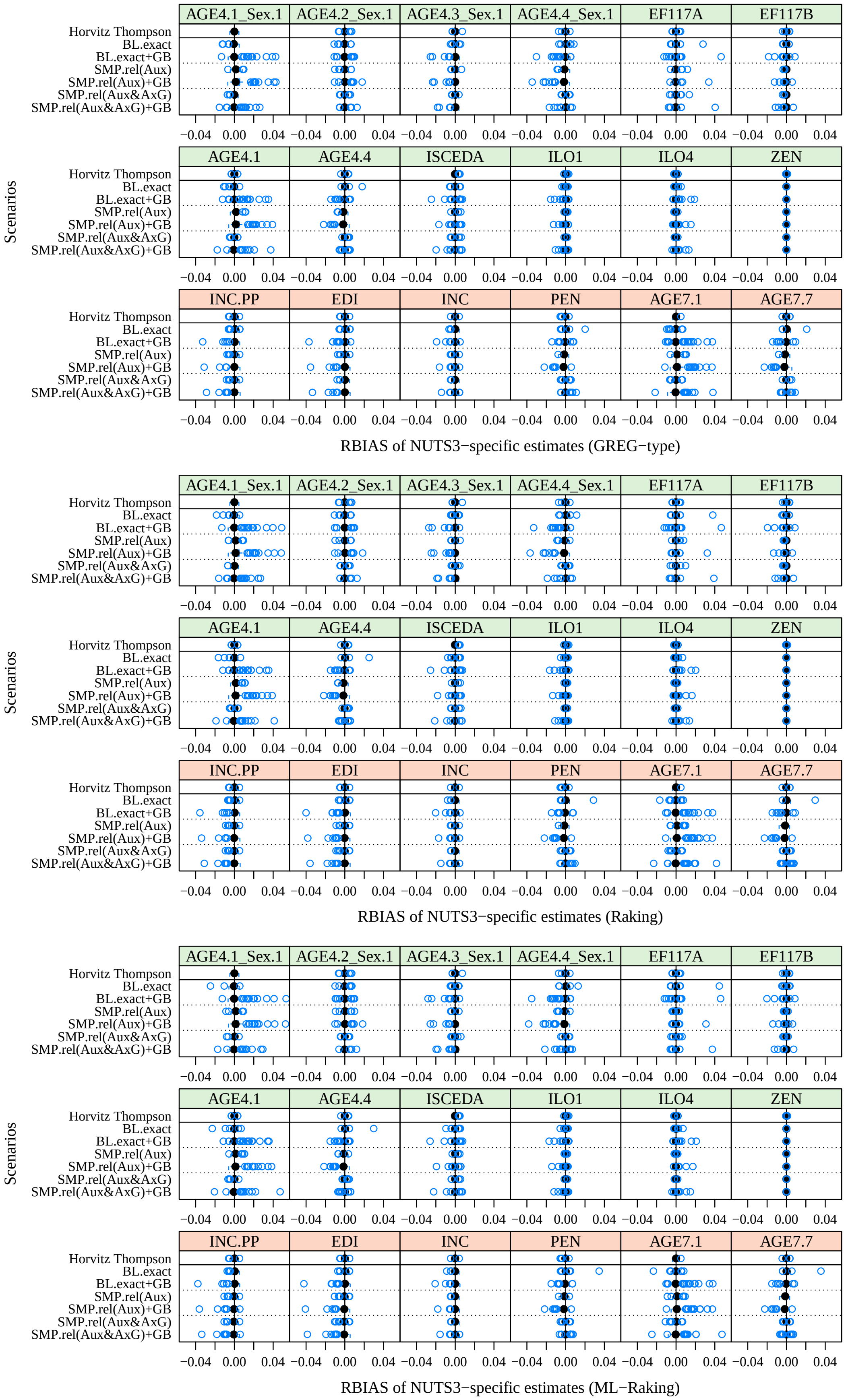


Figure B.10: Relative bias of NUTS3-specific point estimates.

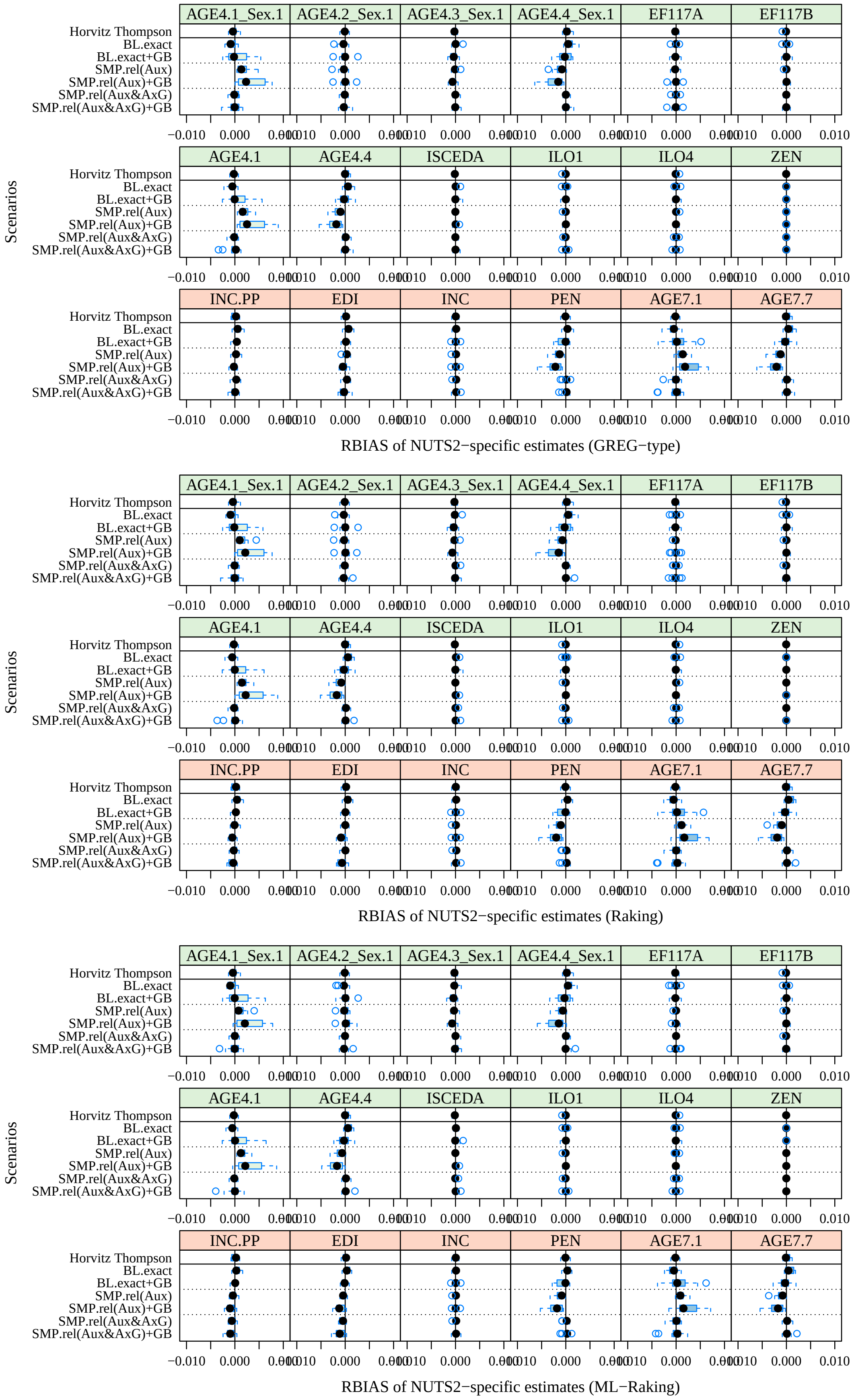


Figure B.11: Relative bias of NUTS2-specific point estimates.

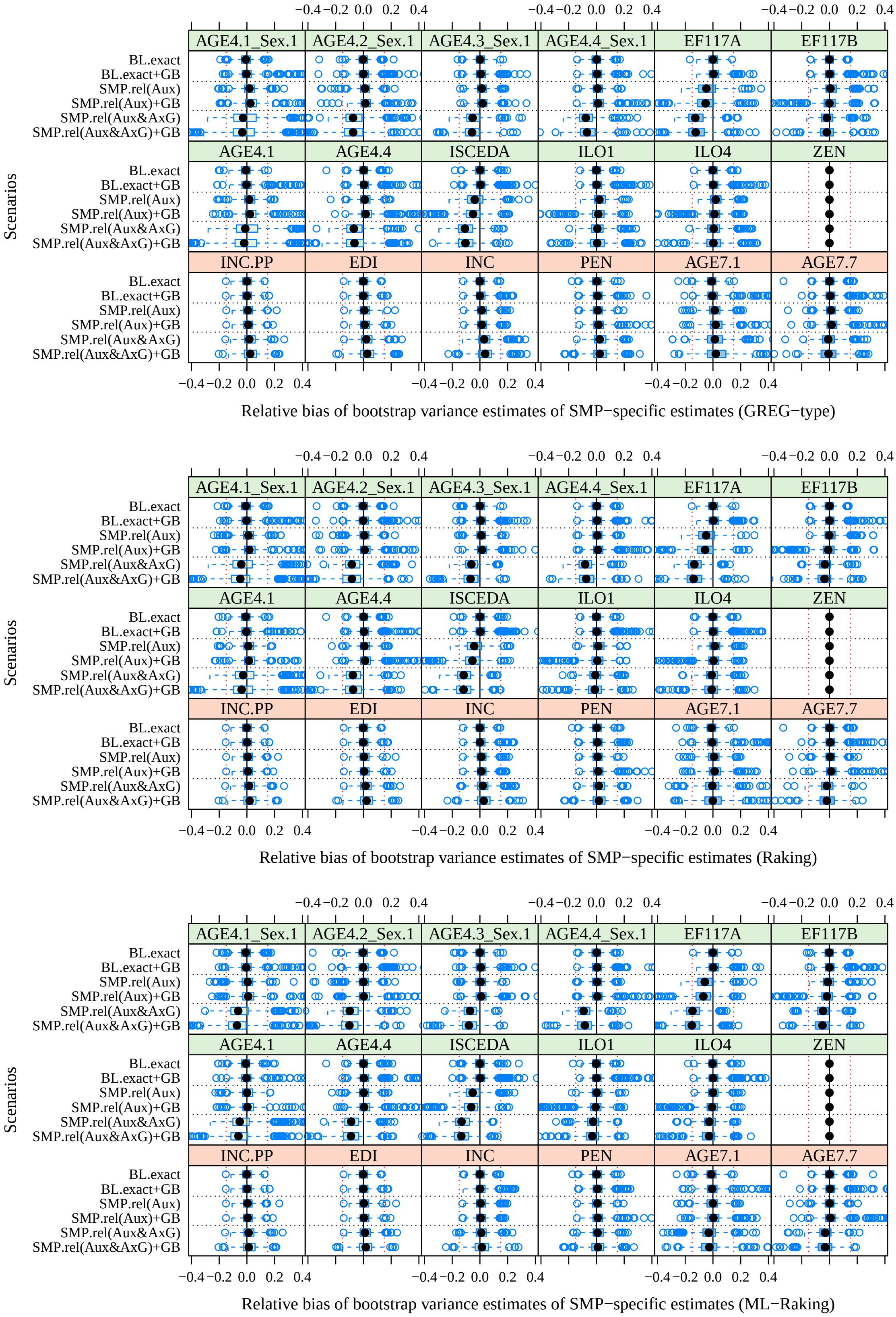


Figure B.12: Relative bias of SMP-specific variance estimates computed by the rescaling bootstrap.

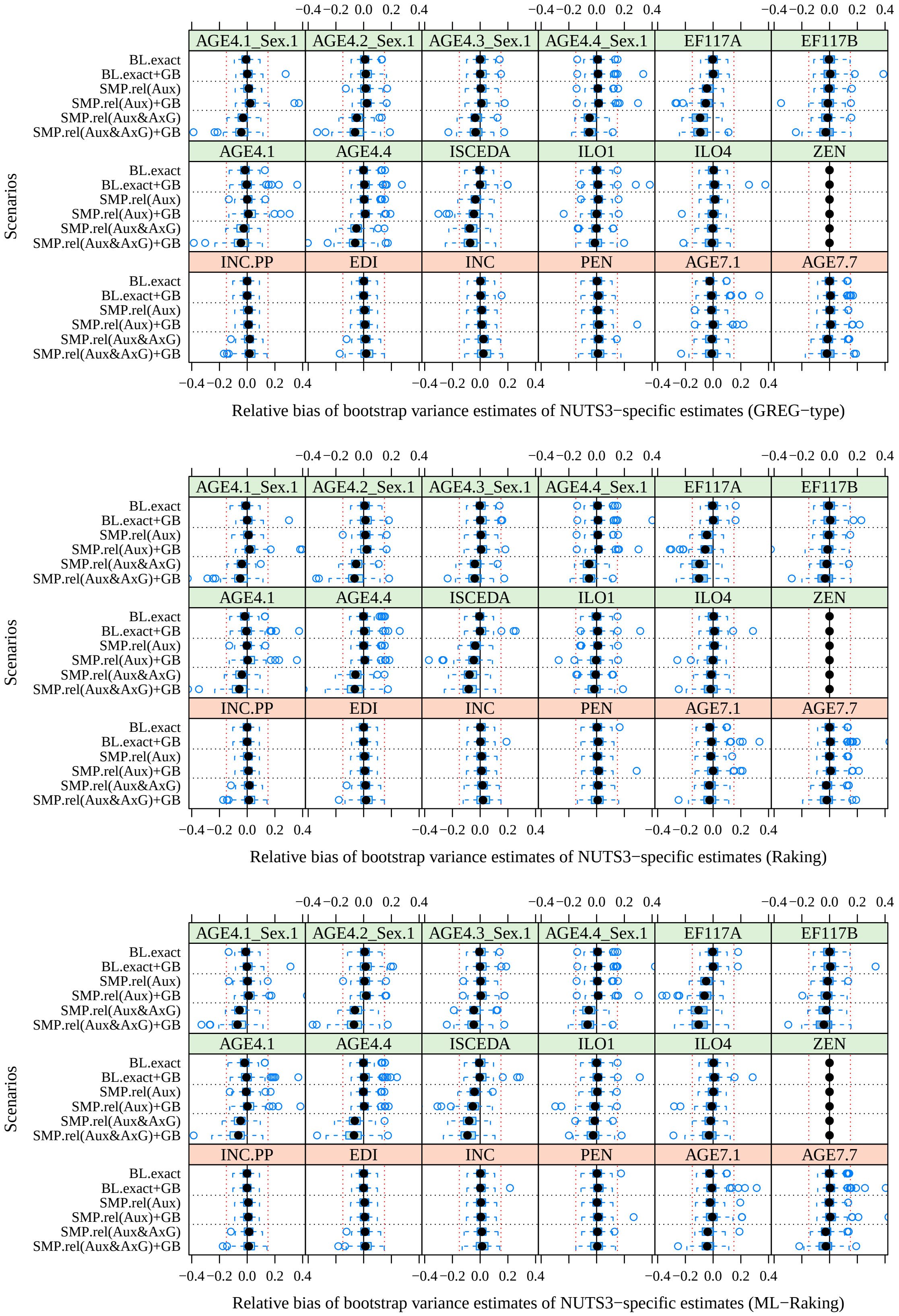


Figure B.13: Relative bias of NUTS3-specific variance estimates computed by the rescaling bootstrap.

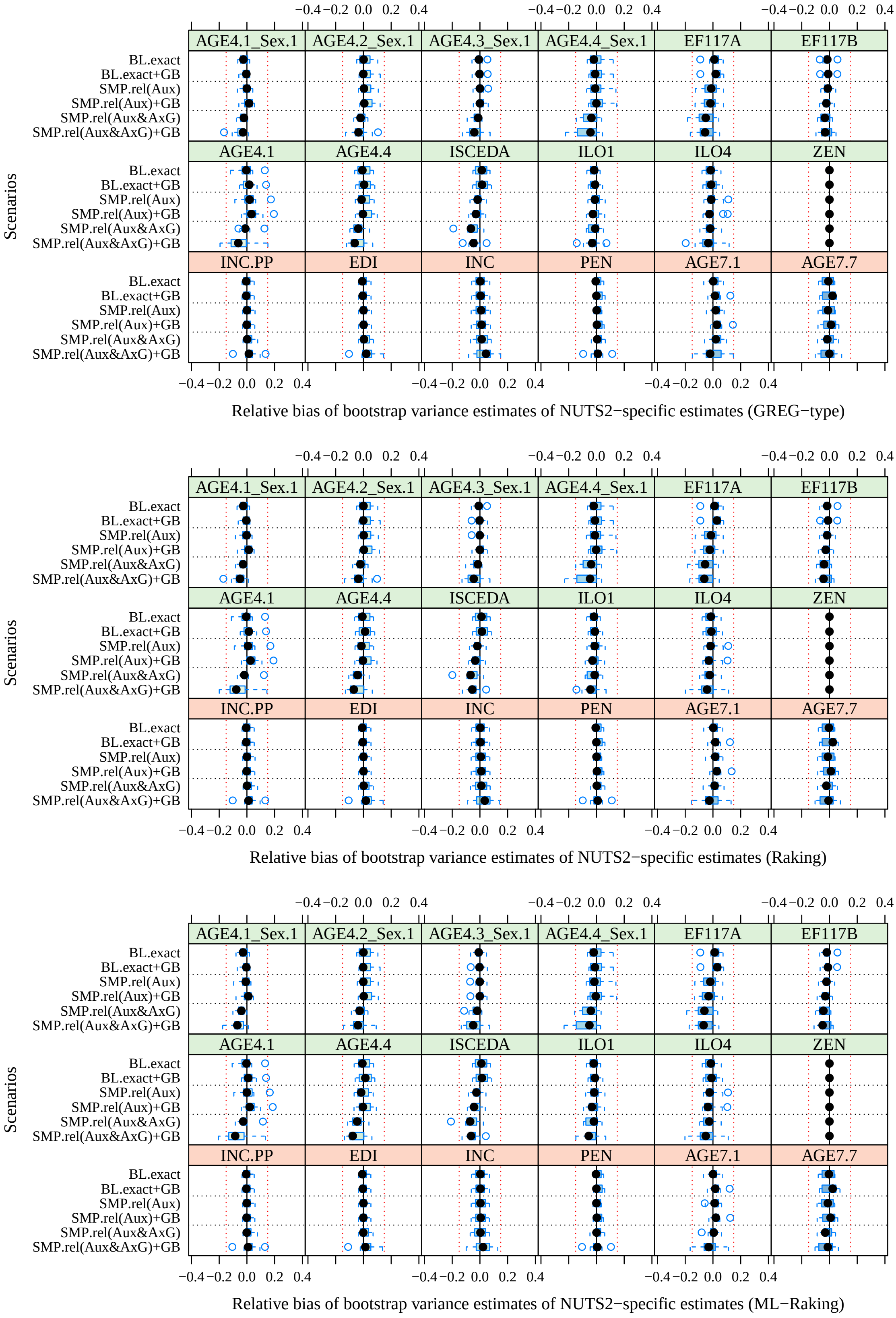


Figure B.14: Relative bias of NUTS2-specific variance estimates computed by the rescaling bootstrap.

Appendix C

R-Packages

In this chapter, two `R` packages are presented which contain the optimal multivariate and multidomain allocation method `MMDopt` and the generalized calibration method `GCAL`. Both the package `MultOptAlloc` for `MMDopt` and the package `genCalib` for the calibration method `GCAL` are still under development. In the following, the input and output structure of both packages is briefly sketched and some possible options are mentioned.

The `R` package `MultOptAlloc`

Input:

- Auxiliary data and structure of stratification (on unit level or aggregated stratum-values)
- Restrictions for regional efficiency and box-constraints
- Sampling fraction, scalarization, standardization, weights

Output:

- Optimal multivariate and univariate allocations, summaries, performance of convergence
- Optional: plots to analyze results (heatmaps, boxplots, Pareto frontier)

The `R` package `genCalib`

Input:

- Design weights, stratification structure, auxiliaries, benchmarks
- Box-constraints, allowed perturbations for relaxed benchmarks
- Optional: rescaling weights for variance estimation

Output:

- Calibration weights, point estimates, variance estimates, summaries

Bibliography

Ahsan, M. J. and Khan, S. U. (1982). Optimum allocation in multivariate stratified random sampling with overhead cost. *Metrika*, 29:71–78.

Alfons, A., Kraft, S., Templ, M., and Filzmoser, P. (2011). Simulation of close-to-reality population data for household surveys with application to eu-silc. *Statistical Methods & Applications*, 20(3):383–407.

Ardilly, P. (2006). *Les techniques de sondage*. Paris: Editions Technip.

Armijo, L. (1966). Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16(1):2–5.

Arthanari, T. S. and Dodge, Y. (1981). *Mathematical Programming in Statistics*, volume 341 of *Wiley Series in Probability and Statistics*. Wiley, New York.

Bankier, M. D. (1988). Power allocations: Determining sample sizes for subnational areas. *American Statistical Association*, 42(3):174–177.

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401):28–36.

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

Bokrantz, R. and Forsgren, A. (2013). An algorithm for approximating convex pareto surfaces based on dual techniques. *INFORMS Journal on Computing*, 25(2):377–393.

Bonnans, J. F., Gilbert, J. C., Lemaréchal, C., and Sagastizábal, C. A. (2006). *Numerical Optimization: Theoretical and Practical Aspects*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2nd edition.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Bretthauer, K. M., Ross, A., and Shetty, B. (1999). Nonlinear integer programming for optimal allocation in stratified sampling. *European Journal of Operational Research*, 116(3):667–680.

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.

Burgard, J. P., Münnich, R., and Rupp, M. (2018). A generalized calibration approach ensuring coherent estimates with small area constraints. Working paper (in submission).

Burgard, J. P., Münnich, R., and Zimmermann, T. (2014). The impact of sampling designs on small area estimates for business data. *Journal of Official Statistics*, 30(4):749–771.

Cassel, C. M., Särndal, C. E., and Wretman, J. H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63(3):615–620.

Cassel, C. M., Särndal, C. E., and Wretman, J. H. (1977). *Foundations of inference in survey sampling*. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley.

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

Chatterjee, S. (1968). Multivariate stratified surveys. *Journal of the American Statistical Association*, 63(322):530–534.

Chatterjee, S. (1972). A study of optimum allocation in multivariate stratified surveys. *Scandinavian Actuarial Journal*, 1972(1):73–80.

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.

Chipperfield, J. and Preston, J. (2007). Efficient bootstrap for business surveys. *Survey Methodology*, 33(2):167–172.

Choudhry, G. H., Rao, J., and Hidiroglou, M. A. (2012). On sample allocation for efficient domain estimation. *Survey Methodology*, 38(1):23–29.

Clarke, F. H. (1983). *Optimization and Nonsmooth Analysis*. Wiley, New York.

Cochran, W. G. (1977). *Sampling Techniques*. Wiley, New York, 3rd edition.

Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. (1996). On the lambertw function. *Advances in Computational Mathematics*, 5(1):329–359.

Craft, D. L., Halabi, T. F., Shih, H. A., and Bortfeld, T. R. (2006). Approximating convex pareto surfaces in multiobjective radiotherapy planning. *Medical Physics*, 33(9):3399–3407.

Dalenius, T. (1953). The multivariate sampling problem. *Scandinavian Actuarial Journal*, 36:92–102.

Dalenius, T. and Hodges, J. L. (1959). Minimum variance stratification. *Journal of the American Statistical Association*, 54(285):88–101.

D'Arrigo, J. and Skinner, C. J. (2010). Linearization variance estimation for generalized raking estimators in the presence of nonresponse. *Survey Methodology*, 36(2):181–192.

Demnati, A. and Rao, J. N. K. (2004). Linearization Variance Estimators for Survey Data. *Statistics Canada*, 30(1):17–26.

Destatis (2016). Bundesstatistikgesetz – BStatG. Retrieved from https://www.destatis.de/DE/Methoden/Rechtsgrundlagen/Statistikbereiche/Inhalte/010_BStatG.pdf?__blob=publicationFile. visited 08/01/2018.

Destatis (2018). Zensus 2011. Available online at https://www.zensus2011.de/EN/Home/home_node.html. visited 08/01/2018.

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

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

Deville, J.-C. and Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91(4):893–912.

Díaz-García, J. A. and Cortez, L. U. (2006). Optimum allocation in multivariate stratified sampling: Multi-objective programming. Technical report, Centro de Investigación en Matemáticas, Guanajuato, México.

Díaz-García, J. A. and Ramos-Quiroga, R. (2014). Optimum allocation in multivariate stratified random sampling: A modified Prékopa's approach. *Journal of Mathematical Modelling and Algorithms in Operations Research*, 13(3):315–330.

Dippo, C. S., Fay, R. E., and Morganstein, D. H. (1984). Computing variances from complex samples with replicate weights. In *Proceedings of the Survey Research Methods Section*, pages 489–494.

Durrett, R. (2010). *Probability: Theory and Examples, 4. Edition*. Cambridge University Press.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Ehrgott, M. (2005). *Multicriteria Optimization*. Springer, Heidelberg, 2nd edition.

Estevao, V. and Särndal, C. E. (2006). Survey estimates by calibration on complex auxiliary information. *International Statistical Review*, 74:127–147.

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 08/01/2018.

Falorsi, P. D. and Righi, P. (2008). A balanced sampling approach for multi-way stratification designs for small area estimation. *Survey Methodology*, 34(2):223–234.

Falorsi, P. D. and Righi, P. (2015). Generalized framework for defining the optimal inclusion probabilities of one-stage sampling designs for multivariate and multi-domain surveys. *Survey Methodology*, 41(1):215–236.

Falorsi, P. D. and Righi, P. (2016). A unified approach for defining optimal multivariate and multi-domains sampling designs. In *Topics in Theoretical and Applied Statistics*, pages 145–152. Springer, Heidelberg.

Fay, R. E. (1989). Theory and application of replicate weighting for variance calculations. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*, volume 12, pages 212–217.

Fay, R. E. and Herriot, R. A. (1979). Estimates of income for small places: An application of james-stein procedures to census data. *Journal of the American Statistical Association*, 74(366):269–277.

Fischer, A. (1997). Solution of monotone complementarity problems with locally lipschitzian functions. *Mathematical Programming*, 76(3):513–532.

Folks, J. L. and Antle, C. E. (1965). Optimum allocation of sampling units to strata when there are R responses of interest. *Journal of the American Statistical Association*, 60(309):225–233.

Friedrich, U. (2016). *Discrete Allocation in Survey Sampling and Analytic Algorithms for Integer Programming*. PhD thesis, Trier University.

Friedrich, U., Münnich, R., de Vries, S., and Wagner, M. (2015). Fast integer-valued algorithms for optimal allocations under constraints in stratified sampling. *Computational Statistics and Data Analysis*, 92:1–12.

Friedrich, U., Münnich, R., and Rupp, M. (2018). Multivariate optimal allocation with box-constraints. *Austrian Journal of Statistics*, 47(2):33–52.

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

Geiger, C. and Kanzow, C. (2002). *Theorie und Numerik restringierter Optimierungsaufgaben*. Springer. Springer Berlin Heidelberg.

Gelman, A. (2007). Struggles with survey weighting and regression modeling. *Statistical Science*, 22(2):153–164.

Godfrey, J., Roshwalb, A., and Wright, R. L. (1984). Model-based stratification in inventory cost estimation. *Journal of Business & Economic Statistics*, 2(1):1–9.

Guggemos, F. and Tillé, Y. (2010). Penalized calibration in survey sampling: Design-based estimation assisted by mixed models. *Journal of Statistical Planning and Inference*, 140(11):3199 – 3212.

Han, S.-P., Pang, J.-S., and Rangaraj, N. (1992). Globally convergent newton methods for nonsmooth equations. *Mathematics of Operations Research*, 17(3):586–607.

Hidiroglou, M. A. and Lavallée, P. (2009). Chapter 17 - sampling and estimation in business surveys. In Rao, C. R., editor, *Handbook of Statistics*, volume 29 of *Handbook of Statistics*, pages 441 – 470. Elsevier.

Hochbaum, D. S. (1995). A nonlinear Knapsack problem. *Operations Research Letters*, 17:103–110.

Hohnhold, H. (2009a). Gerneralized power allocations. Technical report, Statistisches Bundesamt, Wiesbaden.

Hohnhold, H. (2009b). Variants of optimal allocation in stratified sampling. Technical report, Statistisches Bundesamt, Wiesbaden.

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

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685.

Hu, Y., Yang, X., and C.-K., S. (2015). Inexact subgradient methods for quasi-convex optimization problems. *European Journal of Operational Research*, 240(2):315 – 327.

Huddleston, H. F., Claypool, P. L., and R., H. R. (1970). Optimal sample allocation to strata using convex programming. *Journal of the Royal Statistical Society Series C*, 19(3):273–278.

Isaki, C. T. and Fuller, W. A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77(377):89–96.

Ito, K. and Kunisch, K. (2009). On a semi-smooth Newton method and its globalization. *Mathematical Programming*, 118(2):347–370.

Jahn, J. (1986). *Mathematical Vector Optimization in Partially Ordered Linear Spaces*, volume 31 of *Methoden und Verfahren der mathematischen Physiks*. Verlag Peter Lang.

Jahn, J. (2007). *Introduction to the Theory of Nonlinear Optimization*. Springer Berlin Heidelberg, 3rd edition.

Johnson, S. G. (2018). The NLopt nonlinear-optimization package. Available online at `http://ab-initio.mit.edu/nlopt`. visited 08/01/2018.

Khan, M. F., Ali, I., Raghav, Y. S., and Bari, A. (2012). Allocation in multivariate stratified surveys with non-linear random cost function. *American Journal of Operations Research*, 2:100–105.

Khan, M. G. M., Khan, E. A., and Ahsan, M. J. (2003). An optimal multivariate stratified sampling design using dynamic programming. *Australian and New Zealand Journal of Statistics*, 45(1):107–113.

Kish, L. (1965). *Survey sampling*. Wiley classics library. John Wiley and Sons.

Kish, L. (1976). Optima and proxima in linear sample designs. *Journal of the Royal Statistics Society Series A*, 139(1):80–95.

Kokan, A. R. (1963). Optimum allocation in multivariate surveys. *Journal of the Royal Statistics Society Series A*, 126(4):557–565.

Kokan, A. R. and Khan, S. (1967). Optimum allocation in multivariate surveys: An analytical solution. *Journal of the Royal Statistics Society Series B*, 29(1):115–125.

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

Kovar, J. G., Rao, J. N. K., and Wu, C. F. J. (1988). Bootstrap and other methods to measure errors in survey estimates. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 16:25–45.

Krug, W., Nourney, M., and Schmidt, J. (2001). *Wirtschafts- und Sozialstatistik*. Oldenbourg Verlag.

Laarhoven, P. J. M. and Aarts, E. H. L. (1987). *Simulated Annealing: Theory and Applications*. Kluwer Academic Publishers.

Lange, K. (2013). *Optimization*. Springer New York, 2nd edition.

Lehtonen, R. and Veijanen, A. (2009). Chapter 31 - design-based methods of estimation for domains and small areas. In Rao, C. R., editor, *Handbook of Statistics*, volume 29 of *Handbook of Statistics*, pages 219 – 249. Elsevier.

Lin, J. G. (2005). On min-norm and min-max methods of multi-objective optimization. *Mathematical Programming*, 103:1–33.

Lohr, S. L. (2009). *Sampling: Design and Analysis, 2nd Edition*. Advanced (Cengage Learning). Cengage Learning.

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

Merkouris, T. (2010). Combining information from multiple surveys by using regression for efficient small domain estimation. *Journal of the Royal Statistical Society: Series B*, 72(1):27–48.

Mifflin, R. (1977). Semismooth and semiconvex functions in constrained optimization. *SIAM Journal on Control and Optimization*, 15:959–972.

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., 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.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97:558–625.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, New York, NY, USA, second edition.

Ortega, J. M. (1968). The newton-kantorovich theorem. *The American Mathematical Monthly*, 75(6):658–660.

Pang, J.-S. (1990). Newton's method for b-differentiable equations. *Mathematics of Operations Research*, 15(2):311–341.

Preston, J. (2009). Rescaled bootstrap for stratified multistage sampling. *Survey Methodology*, 35(2):227–234.

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.

Qi, L. and Sun, J. (1994). A trust region algorithm for minimization of locally lipschitzian functions. *Mathematical Programming*, 66(1-3):25–43.

Rademacher, H. (1919). Über partielle und totale Differenzierbarkeit von Funktionen mehrerer Variabeln und über die Transformation der Doppelintegrale. *Collected Papers of Hans Rademacher*, 1(4).

Rahman, A. and Harding, A. (2016). *Small Area Estimation and Microsimulation Modeling*, volume 1. Chapman and Hall/CRC Press.

Rao, J. and Molina, I. (2015). *Small Area Estimation, 2nd edition*. 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.

Rao, J. N. K. and Singh, A. C. (2009). Range restricted weight calibration for survey data using ridge regression. *Pakistan Journal of Statistics*, 25:371–384.

Rao, J. N. K. and Wu, C. F. J. (1988). Resampling inference with complex survey data. *Journal of the american statistical association*, 83(401):231–241.

Rao, J. N. K., Wu, C. F. J., and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey methodology*, 18(2):209–217.

Renssen, R. H. and Nieuwenbroek, N. J. (1997). Aligning estimates for common variables in two or more sample surveys. *Journal of the American Statistical Association*, 92(437):368–374.

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

Robinson, S. M. (1987). *Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity*, pages 45–66. Springer Berlin Heidelberg, Berlin, Heidelberg.

Ruszczynski, A. (2006). *Nonlinear Optimization*. Princeton University Press, Princeton.

Sachs, E. W. and Sachs, S. M. (2011). Nonmonotone line searches for optimization algorithms. *Control and Cybernetics*, 40:1059–1075.

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

Schaich, E. and Münnich, R. (1993). Zum Allokationsproblem bei mehreren Untersuchungsvariablen. *Allgemeines Statistisches Archiv*, 77:390–405.

Shapiro, A. (1990). On concepts of directional differentiability. *Journal of Optimization Theory and Applications*, 66(3):477–487.

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

Spellucci, P. (1993). *Numerische Verfahren der nichtlinearen Optimierung*. Birkhäuser Basel.

Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer.

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

Stukel, D., Hidiroglou, M., and Särndal, C. E. (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. (2006). *Sampling Algorithms*. Springer Science and Business Media, Inc.

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

Tschuprow, A. A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, 2:461–493.

Vanderhoeft, C. (2001). *Generalised Calibration at Statistics Belgium: SPSS Module G-CALIB-S and Current Practices*. Statistics Belgium. Institute National de Statistique.

Varian, H. R. (2010). *Intermediate Microeconomics: A Modern Approach*. W.W. Norton & Company.

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

Walter, W. (2002). *Analysis 2, 5. erweiterte Auflage*. Springer.

Werner, D. (2007). *Funktionalanalysis*. Springer-Lehrbuch. Springer Berlin Heidelberg.

Witting, H. (1978). *Mathematische Statistik, 3. Auflage*. Teubner Studienbücher.

Wolter, K. (2007). *Introduction to Variance Estimation*. Springer Series in Statistics. Springer New York.

Wright, R. L. (1983). Finite population sampling with multivariate auxiliary information. *Journal of the American Statistical Association*, 78(384):879–884.

Yamamuro, S. (1974). *Differential calculus in topological linear spaces*. Lecture notes in mathematics. Springer.

You, Y. and Rao, J. N. K. (2002). A pseudo-empirical best linear unbiased prediction approach to small area estimation using survey weights. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 30(3):431–439.

Ypma, J., Borchers, H. W., and Eddelbuettel, D. (2017). R package nloptr: R interface to NLopt. `http://CRAN.R-project.org/package=nloptr`. R package version 1.04.

Zieschang, K. D. (1990). Sample weighting methods and estimation of totals in the consumer expenditure survey. *Journal of the American Statistical Association*, 85(412):986–1001.