

# Multivariate Optimal Allocation with Box-Constraints

Ulf Friedrich  
Trier University

Ralf Münnich  
Trier University

Martin Rupp  
Trier University

---

## Abstract

Modern surveys aim at fostering accurate information on demographic and other variables. The necessity for providing figures on regional levels and on a variety of subclasses leads to fine stratifications of the population. Optimizing the accuracy of stratified random samples requires incorporating a vast amount of strata on various levels of aggregation. Accounting for several variables of interest for the optimization yields a multivariate optimal allocation problem in which practical issues such as cost restrictions or control of sampling fractions have to be considered. Taking advantage of the special structure of the variance functions and applying Pareto optimization, efficient algorithms are developed which allow solving large-scale problems. Additionally, integrality- and box-constraints on the sample sizes are considered. The performance of the algorithms is presented comparatively using an open household dataset illustrating their advantages and relevance for modern surveys.

*Keywords:* stratified random sampling, multi-criteria optimization, linear constraints, integer optimization, Pareto optimality, semismooth Newton.

---

## 1. Introduction

Accurate population figures provide an important basis for political and economic decision processes. In light of urban audits and regional policies, these figures, however, have to be made available in sufficient regional detail as well as for many sub-classes, which requires introducing a vast number of strata by regions and content. Censuses, registers, or adequate surveys can provide the information necessary for such research. Using surveys, stratified random sampling provides an adequate basis that allows integrating further optimization techniques while considering practical settings with various constraints. Additionally, several variables of interest may be incorporated in the optimization process which either contains complementary or conflictory information. This finally leads to a multivariate optimal allocation problem under constraints regarding regional as well as context-specific stratifications.

For the stratified random sampling problem we assume a finite population  $\mathcal{U}$  of size  $N$  with disjoint cross-classification strata  $h = 1, \dots, H$ . Let  $\tau_Y$  denote the total of a variable of interest  $Y$ . In stratified random sampling, an unbiased estimator is  $\hat{\tau}_Y^{\text{StrRS}} = \sum_{h=1}^H N_h \bar{y}_h$ , where  $\bar{y}_h$  is the sample mean of variable  $Y$  and  $N_h$  is the population size in stratum  $h$ . Its

variance is

$$\text{Var}(\hat{\tau}_Y^{\text{StrRS}}) = \sum_{h=1}^H \frac{N_h^2 S_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \quad (1)$$

with known stratum-specific variances  $S_h^2$  of variable  $Y$  and stratum-specific sample sizes  $n_h$  for all strata  $h = 1, \dots, H$  (Lohr 2010, chapter 4). In practice, earlier surveys or highly correlated variables yield the necessary information for  $S_h^2$ . In our presentation, we tacitly assume that adequate proxies are available as a discussion of proxy quality and its implications is beyond the scope of this work.

Minimizing the variance (1) with respect to the stratum-specific sample sizes  $n_h$  while respecting a given total sample size  $n_{\max}$  leads to the (univariate) optimal allocation introduced by Tschuprow (1923) and Neyman (1934). In contrast to the equal and proportional allocation, see Cochran (1977), the optimal allocation depends on the variable of interest. The resulting optimal allocation is given in closed form by

$$n_h^* = \frac{N_h S_h}{\sum_{k=1}^H N_k S_k} \cdot n_{\max}. \quad (2)$$

This allocation method is extended in Gabler, Ganninger, and Münnich (2012) and Münnich, Sachs, and Wagner (2012b), such that for each stratum-specific sample size  $n_h$  box-constraints  $m_h, M_h$  with

$$2 \leq m_h \leq n_h \leq M_h \leq N_h \quad (3)$$

are added to the optimization problem. As zero sample sizes in single strata lead to biased estimates and variance estimation of the total estimate requires stratum-specific sample sizes of at least two, a lower constraint  $m_h \geq 2$  is applied. Upper constraints  $M_h \leq N_h$  have to be introduced to avoid overallocation in strata where  $n_h$  given by (2) exceeds  $N_h$ . A further reduction of  $M_h$  allows to control sample fractions, for example to avoid highly different response burdens in various regions or strata. In addition,  $M_h$  prevents a stratum-specific full census which is prohibited by law in specific surveys, for example by judgment of the German Federal Administrative Court (BVerwG, 03/15/2017, 8 C 6.16).

Altogether, the optimal allocation problem under box-constraints is given by

$$\begin{aligned} \min_{n \in \mathbb{R}_+^H} \quad & \text{Var}(\hat{\tau}_Y^{\text{StrRS}}) \\ \text{s.t.} \quad & \sum_{h=1}^H n_h = n_{\max} \\ & m_h \leq n_h \leq M_h \quad \forall h = 1, \dots, H. \end{aligned} \quad (4)$$

The problem can equivalently be stated with the inequality constraint  $\sum_{h=1}^H n_h \leq n_{\max}$ , but equality holds at every optimal solution.

Friedrich, Münnich, de Vries, and Wagner (2015) provide a further extension that ensures integrality of the solution of the optimal allocation. Using Gabler *et al.* (2012) and by separating different sub-regions, the method can be rewritten to a simultaneous optimal allocation for multiple areas. This is achieved by using a quadratic separable decision function.

In the multivariate generalization of the optimal allocation problem,  $K$  different variances  $\text{Var}(\hat{\tau}_1^{\text{StrRS}}), \dots, \text{Var}(\hat{\tau}_K^{\text{StrRS}})$  are considered simultaneously. Dalenius (1953) discusses this problem in detail and distinguishes two solution strategies. In the first, one or more of the variances  $\text{Var}(\hat{\tau}_k^{\text{StrRS}})$ ,  $k = 1, \dots, K$ , 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 optimization problem with non-linear constraints. In the second, the variances are minimized simultaneously subject to linear size (or cost) constraints. This

perspective leads to a multi-objective optimization problem with conflicting objectives that requires an appropriate mathematical theory. In particular, an adequate notion of optimality, such as Pareto optimality, is essential and the problem has to be transformed into a form that is solvable by optimization algorithms. Most of the literature dealing with multivariate allocation splits up depending on which of these two formulations is used.

Indeed, Chatterjee (1968), Chatterjee (1972), and Huddleston, Claypool, and Hocking (1970) use the first variant. Multivariate optimal allocation problems are addressed in the same way in Kokan (1963) and supplemented by existence and uniqueness results in Kokan and Khan (1967). Introducing overhead costs, Ahsan and Khan (1982) discuss the problem with variance constraints for a more general objective function. More recently, Bankier (1988), Hohnhold (2009a), and Hohnhold (2009b) have published allocation techniques with more than one level of strata. These techniques are based on a compensation of the accuracy of regional estimates and population total estimates and, hence, also belong to the first class of methods. Falorsi and Righi (2015) present a generalized framework for defining the optimal inclusion probabilities in multivariate and multi-domain surveys. Falorsi and Righi (2008) and Falorsi and Righi (2016) introduce a solution method using a balanced sampling design. Combining aspects from both strategies, Kish (1976) proposes to combine aspects of variance and cost minimization in a non-linear model with the help of loss functions and discusses various choices for the objective function within his model.

Turning to the second solution strategy introduced by Dalenius (1953), the multivariate optimal allocation is threatened in Folks and Antle (1965) as a multi-objective optimization problem with linear constraints. They discuss the mathematical theory of scalarization and the relationship between the multi-objective problem and the scalarized problem. Moreover, they prove a sufficiency result for the set of efficient (or Pareto optimal) solutions for the simple problem without box-constraints and neglecting the integrality of the solutions. Díaz-García and Ramos-Quiroga (2014) solve the multivariate allocation as a multi-objective problem as well but with the help of stochastic programming. Khan, Ali, Raghav, and Bari (2012) use stochastic programming on another model. Both methods lead to non-linear integer optimization problems which are hard to solve even for small instances. Khan, Khan, and Ahsan (2003) solve multivariate allocation problems by exploiting the separability of the objective function and 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 the computational study of Bretthauer, Ross, and Shetty (1999) shows.

All strategies using the multi-objective perspective on the problem have to use scalarization techniques to combine the variances for the variables in a 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.

We also 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 result in Folks and Antle (1965) by giving a (necessary and sufficient) characterization of *all* Pareto optimal points. Moreover, in contrast to earlier publications, we solve the problem while respecting integrality and box-constraints. We compute the set of Pareto optimal solutions, the so-called Pareto frontier, for this refined problem formulation which allows decision makers to choose a personally specified preference from this set.

The solution of allocation problems under the box-constraints (3) may yield non-differentiable points and 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 (Münnich, Sachs, and Wagner 2012a and Wagner 2013). Because stratum-specific sample sizes are integer values, we also provide an alternative algorithm to derive a multivariate optimal allocation in which all stratum-specific samples are integer-valued. This strategy avoids rounding and is based on the integer optimal allocation techniques published

in Friedrich *et al.* (2015).

Each scalarization for multivariate optimal 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. We present an alternative solution that extends the techniques published in Schaich and Münnich (1993).

Finally, it is of great importance for practical applications to solve large problem instances in appropriate time. Our methods solve multivariate optimal 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 that are generally computationally tractable for only a small number of strata. The computationally efficient solution of large (integer) multivariate optimal allocation problems supplements the theoretical discussion and certainly is another central innovation of our methods.

In Section 2, we use the method of box-constrained optimal allocation presented in Münnich *et al.* (2012b) as a starting point to derive a generalized multivariate box-constrained optimal allocation problem with various decision-making functions and standardization techniques. Moreover, we establish the link between the multivariate allocation problem and the theory of Pareto optimization. In Section 3, we provide efficient numerical algorithms for selected variants of the developed problem. These are fast enough to solve even large problem instances and avoid rounding the solution by finding the globally optimal integer-valued solution. In Section 4, we present selected performance and simulation results based on the open AMELIA household dataset (Alfons, Burgard, Filzmoser, Hulliger, Kolb, Kraft, Münnich, Schoch, and Templ 2011, as well as Merkle, Burgard, and Münnich 2016).

## 2. Multivariate optimal allocation

### 2.1. Preliminaries

In a multivariate optimal allocation problem, several variables of interest are considered simultaneously. The resulting optimization problem has several conflicting objective functions. Thereby, the correlation between the variables of interest, the variable types as well as the purpose of the survey are decisive factors. The use of a scalarization technique is mandatory to treat this conflict of objectives and to solve the optimization problem numerically. The choice of a scalarization technique is not clear in advance, depends on the application, and has a considerable influence on the solution of the problem.

The most intuitive scalarization technique is the weighted sum method, for which each objective is weighted and the weighted objectives are cumulated (Jahn 1986). Another widespread technique is the epsilon-constraint method, which corresponds to minimizing the cost while respecting variance restrictions (Ehrgott 2005 and Falorsi and Righi 2015). As we focus on the minimization of the variance, we do not consider the epsilon-constraint method here. Moreover, we propose a  $p$ -norm of the objectives ( $p = 1, 2, 4, 8, \infty$ ), which is discussed in Lin (2005). Schaich and Münnich (1993) study the particular case  $p = \infty$  which is equivalent to the so-called *min-max* method.

In addition to scalarization, standardization techniques are also important for standardizing variances of various types of variables of interest. Schaich and Münnich (1993) suggest 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

$$CV^2(\hat{\tau}_Y^{\text{StrRS}}) := \frac{\text{Var}(\hat{\tau}_Y^{\text{StrRS}})}{\tau_Y^2}$$

with the population total  $\tau_Y$  of variable  $Y$  for the (CV2)-standardization. Although the principal effect is similar, squaring the coefficient of variation may lead to small differences

in some settings. A drawback of using the squared coefficient of variation is the requirement for the population total  $\tau_Y$  of variable  $Y$ , which is generally not given in advance and which, as it is a ratio, is even more demanding than using only the proxies for the stratum-specific variances.

Furthermore, we propose the alternative (opt) standardization, in which the objectives are standardized by the unique univariate optimal allocations as standardization factors. The standardized objective for the variable of interest  $Y$  is given by

$$\text{opt}(\hat{\tau}_Y^{\text{StrRS}}) := \frac{\text{Var}(\hat{\tau}_Y^{\text{StrRS}})}{\text{Var}_Y^{\text{opt}}}$$

where  $\text{Var}_Y^{\text{opt}}$  is the univariate optimal allocation for the variable  $Y$  computed, for example, with the box-constraint optimal allocation by [Münnich et al. \(2012b\)](#). This standardization technique reflects the relative loss for each variable under consideration when using the compromise allocation rather than the single variable optimized allocation. In contrast to (CV2), an advantage of this technique is that the total  $\tau_Y$  of variable  $Y$  is not required. Moreover, if  $S_h^2$  has to be estimated, the uncertainty and blur of this estimation is symmetrically present in the numerator and denominator of the objectives, and, thus, eliminated. Hence, a standardization by the univariate optimal variances results in a more robust multivariate optimal allocation.

## 2.2. Methods of multivariate optimal allocation

The optimal allocation with respect to only one variable of interest  $Y$  with box-constraints for stratum-specific sample sizes is given by (4). The simultaneous consideration of several variables of interest  $Y_1, \dots, Y_K$  yields the following multi-criteria optimization problem

$$\begin{aligned} \min_{n \in \mathbb{R}_+^H} \quad & (\text{Var}(\hat{\tau}_1^{\text{StrRS}}), \dots, \text{Var}(\hat{\tau}_K^{\text{StrRS}})) \\ \text{s.t.} \quad & \sum_{h=1}^H n_h = n_{\max} \\ & m_h \leq n_h \leq M_h \quad \forall h = 1, \dots, H \end{aligned} \quad (5)$$

where

$$\text{Var}(\hat{\tau}_k^{\text{StrRS}}) = \sum_{h=1}^H \frac{N_h^2 (S_h^k)^2}{n_h} \left(1 - \frac{n_h}{N_h}\right)$$

with  $H$  cross-classification strata, stratum sizes  $N_h$ , and stratum-specific variances  $(S_h^k)^2$  given for each stratum  $h = 1, \dots, H$  and variables of interest  $k = 1, \dots, K$ . To prove the existence of a solution, we refer to [Jahn \(1986, Theorem 6.3\)](#). The multivariate allocation problem (5) can be reformulated as a single-objective optimization problem with objective function  $f : \mathbb{R}_+^H \rightarrow \mathbb{R}_+$  by combining the  $K$  original objective functions in one scalar expression. In this scalarization the objective functions are also standardized to make them comparable. Next, we explain the standardized scalarization in detail.

### *Weighted sum scalarization*

Using the weighted sum scalarization method, we obtain the objective function  $f$  given by

$$\begin{aligned} f(n) &:= \sum_{k=1}^K w_k \frac{\text{Var}(\hat{\tau}_k^{\text{StrRS}})}{\alpha_k} \\ &= \sum_{k=1}^K w_k \frac{\sum_{h=1}^H \frac{N_h^2 (S_h^k)^2}{n_h} \left(1 - \frac{n_h}{N_h}\right)}{\alpha_k}, \end{aligned} \quad (6)$$

which depends on externally given weights  $w_1, \dots, w_K \in \mathbb{R}_+$  with  $\sum_{k=1}^K w_k = 1$ . Using the squared coefficient of variation as a standardization method, factor  $\alpha \in \mathbb{R}^K$  is defined by  $\alpha_k := \tau_k^2$  for all  $k = 1, \dots, K$ . Alternatively, if we apply the single unique optimal allocations as a standardization technique, we set  $\alpha_k := \text{Var}_k^{\text{opt}}$  for all  $k = 1, \dots, K$ .

#### Alternative scalarization techniques

Alternatively, by using the  $p$ -norm ( $p < \infty$ ) as scalarization method, we obtain the objective

$$f(n) := \left( \sum_{k=1}^K \left( \sqrt{\frac{\text{Var}(\hat{\tau}_k^{\text{StrRS}})}{\alpha_k}} \right)^p \right)^{\frac{1}{p}}. \quad (7)$$

If we define  $f$  by the 2-norm, it is equivalent to the weighted sum with equal weights. Finally, using the *min-max* method,  $f$  is given by

$$f(n) := \max_{k=1, \dots, K} \sqrt{\frac{\text{Var}(\hat{\tau}_k^{\text{StrRS}})}{\alpha_k}}. \quad (8)$$

#### Properties of the objective function

In the case of the weighted sum scalarization, the objective function  $f$  in (6) is continuously differentiable, strictly convex, and separable (Münnich *et al.* 2012b). These properties are essential for the fast algorithms presented in Section 3. If the alternative scalarization methods are used,  $f$  changes and may lose some of these properties. In particular, the objective  $f$  in (7) is continuously differentiable and strictly convex, but only separable if  $p = 2$ . If  $f$  is not separable, as in the case  $p \neq 2$ , special attention must be paid to the selection of the solution algorithm.

Furthermore, for  $p = \infty$  the objective (8) is not continuously differentiable. However, many classical optimization methods, such as the Newton method, rely on differentiability and are not applicable in this case. For more details we refer to Section 3.

### 2.3. Weighted sum and Pareto optimization

The scalarization by the weighted sum fits in the theory of Pareto optimality. When optimizing competing objectives, the *Pareto frontier* describes the set of all efficient solutions in the sense that for all points in the frontier one objective can only be improved by diminishing another. Therefore, the Pareto frontier gives a very suitable characterization of all those points decision makers should consider in a multi-criteria optimization problem. On the other hand, it is not advisable to choose an allocation which is not on the Pareto frontier, because it could be improved without cost.

Moreover, the Pareto frontier describes the optimal solutions independently from the weighting, that means independently from the ranking of the variables of interest by decision makers. Instead of determining the ranking in advance, our method allows users to select a preferred solution among all Pareto optimal points *after* the optimization step. Advantages of this procedure are the ability to optimize without a known priority ranking of the variables of interest and the possibility to use additional information at the time of decision, for example variance structures or sensitivity, and the robustness of the solution with respect to the weights.

We describe the entire frontier of Pareto optimal solutions to the multivariate allocation problem (5) mathematically in Appendix A and extend the results by Folks and Antle (1965). We prove that each optimal solution of the weighted sum reformulation for an arbitrary choice of weights is a Pareto optimal solution for (5). Moreover, if we solve the weighted sum problem for all possible choices of weights, we obtain all Pareto optimal solutions of the original problem (subject only to the discretization of the weights). This way, we compute the whole



Pareto frontier of the multivariate allocation problem. We refer to Sections 4.2 and 4.3 for computation algorithms of the Pareto frontier, their implementation, and exemplary numerical results.

## 2.4. Multivariate optimal integer allocation

So far, we have ignored the requirement that the calculated stratum-specific sample sizes in an (univariate or multivariate) optimal allocation problem have to be in the set of non-negative integers for almost all application problems because, for example, a fraction of a person cannot be drawn in a sample. In general, the solution of the allocation problems in continuous variables presented in Section 2.2 is not an integer but a fractional number. In practical applications 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 as in the example data presented in Section 4.4. Therefore, we also discuss an algorithm for the computation of the globally optimal solution in integer variables.

## 3. Algorithmic solution of allocation problems

In this section we present two efficient algorithms for the numerical solution of (5) in continuous and integer variables. The strict convexity and separability of the scalarized objective function  $f$  is crucial for the correctness of both algorithms. Concerning the scalarization and standardization techniques presented in Section 2,  $f$  is only separable for the weighted sum or the 2-norm but not for the other  $p$ -norms or *min-max*.

### 3.1. Semismooth Newton

The algorithm is based on developments and derivations published in Münnich *et al.* (2012b) who consider a univariate optimal allocation problem with box-constraints. After scalarization and standardization with the techniques described in Section 2, it is also applicable to the multivariate problem. The main characteristic of the algorithm is to express the stratum-specific sample sizes  $n_h$  as a function of the Lagrange multiplier  $\lambda \in \mathbb{R}$  by transforming the Karush-Kuhn-Tucker optimality conditions. Then, the expression for  $n_h(\lambda)$  is substituted into the equality-constraint of the original problem, which leads to a one-dimensional system of equations depending on  $\lambda$

$$\Phi(\lambda) := \sum_{h=1}^H n_h(\lambda) - n_{\max} = 0 \quad (9)$$

with  $n_h(\lambda) := \text{Proj}_{[m_h, M_h]} \left( \frac{S_h^2 N_h^2}{\lambda} \right)^{\frac{1}{2}}$ , where  $\text{Proj}_{[m_h, M_h]}$  denotes the projection into the interval  $[m_h, M_h]$ . Due to this cut-off,  $\Phi$  is not continuously differentiable. Nevertheless, Qi and Sun (1993) show semismoothness for  $\Phi$ . Münnich *et al.* (2012b) suggest a fixed-point iteration to solve (9). We chose a semismooth Newton method because it allows for additional generalizations. For a detailed presentation of the semismooth Newton method in the context of survey statistics, we refer to Münnich *et al.* (2012a).

It is also necessary to solve non-separable settings of the continuous allocation problem for a complete comparison of the methods in Section 4. These instances are solved with the R package `nloptr` (Ypma, Borchers, and Edelbuettel 2014).

### 3.2. Solution as integer optimization problem

As in the continuous case, the multivariate optimal *integer* allocation problem is algorithmically tractable whenever the objective function  $f$  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.

Friedrich *et al.* (2015) present three algorithms for the problem that use the fact that the minimization of a separable and convex function is polynomially solvable in integer variables if the feasible set is a so-called *polymatroid*, which is a convex polytope with strong combinatorial properties. An exhaustive discussion of the mathematical background is given in Friedrich (2016). The algorithms are based on so-called *Greedy* strategies and find the globally optimal integer solution.

In the case of convex objective functions that are not necessarily separable, the problem can still be reformulated as a single-objective integer optimization problem, but the fast Greedy algorithms do not find the optimal solution. 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 (many hours instead of seconds). Therefore, we do not solve the integer version of the non-separable problems in Section 4.

## 4. Simulation study and results

We use the synthetic AMELIA dataset (Merkle *et al.* 2016) for a simulation study to verify and compare the presented methods. It is a household dataset reflecting the household structure of Europe containing 3 781 289 households and 10 012 600 individuals. We use the household structure with stratification levels districts (DIS – 40 strata), household size (HHS – 6 strata), and degree of urbanization (DOU – 3 strata). This results in  $40 \cdot 6 \cdot 3 = 720$  cross-classification strata. As variances have to be compared, cross-classification strata with a total size of  $N_h < 2$  are eliminated, so that the simulation only contains 676 strata. The size distribution of the 676 strata is shown in Figure 1 clustered by the classes of household size. Classes 1 up to 5 contain households with the respective number of persons, class 6 contains households with more than five persons.

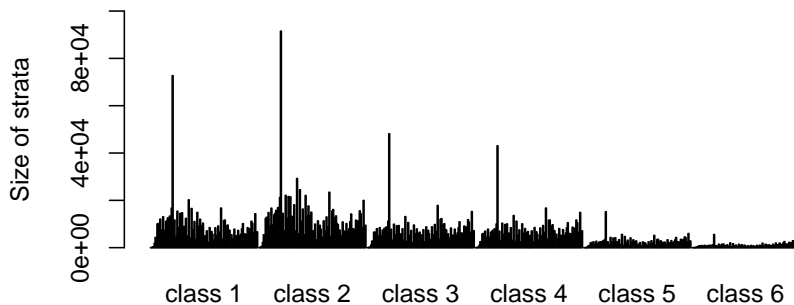


Figure 1: Size distribution of the 676 cross-classification strata clustered by the classes of household size.

We choose the total household income (INC), the social income (SOC), and the age of the main income earner (HAGE) as variables of interest. The social income of a household is defined as the sum of unemployment, old-age, survivors, sickness, disability, and education-related benefits of all people within the household, see Merkle *et al.* (2016). Although the three variables of interest are all continuous, our method is applicable to proportions of categorical variables as well.

The correlations within each district are presented in the boxplots in Figure 2. The correlations over the population are depicted as vertical lines. In particular, concerning variable HAGE, we observe some differences between the overall correlation and the separated district



correlations.

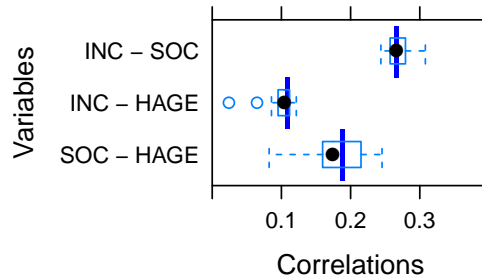


Figure 2: Boxplots of the correlations of the variable of interest per district in contrast to the total population (vertical lines).

We compare the estimates under various settings with the relative mean squared error (RMSE). Because the sampling design is stratified random sampling, the estimates are unbiased and the RMSE comparison is equivalent to the variance comparison (Lohr 2010, chapter 2). Since the AMELIA dataset is used, the true values of the RMSEs can be computed directly for the comparative analysis rather than the Monte Carlo equivalences. Furthermore, most of the following figures and graphs do not contain absolute values of errors, variances or sample sizes, but relative values compared to the case of an independent univariate optimal allocation of the three variables of interest.

#### 4.1. Comparison of variances depending on the decision-making strategy

In the following, we compare the results of the four decision-making functions 2-norm, 4-norm, 8-norm, and min-max as well for the (CV2)-standardization and the alternative (opt)-standardization presented in Section 2.1. The variance functions for each variable of interest have equal weights in these settings and, as pointed out before, the 2-norm is equivalent to the weighted sum with equal weights.

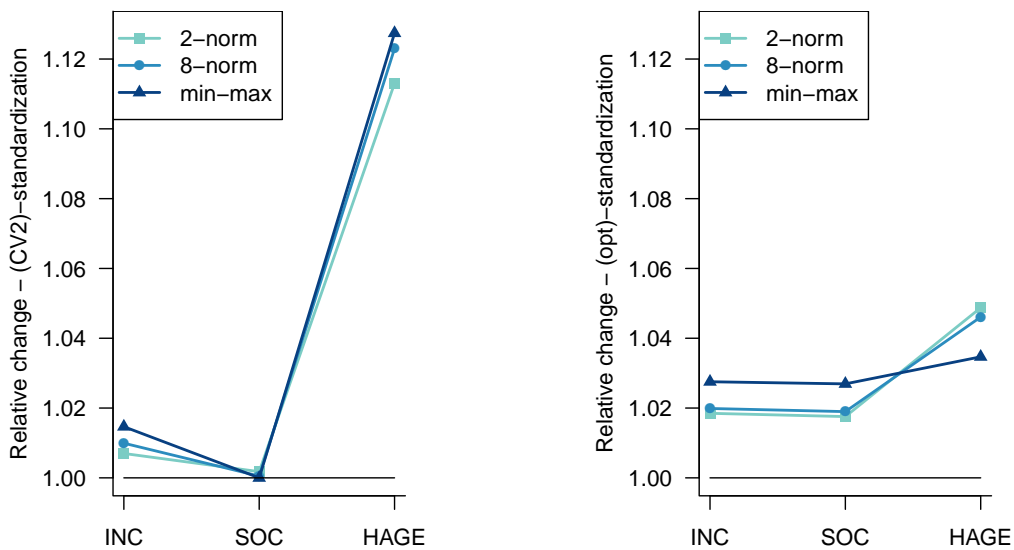


Figure 3: Relative change of RMSE for the estimated population totals for various standardization and scalarization techniques.

In Figure 3 we show the relative increase of the RMSE for the total population estimates of the three variables of interest compared to the optimal univariate allocation computed by the

method of Münnich *et al.* (2012b). For a better visibility, the 4-norm is not displayed in the figure. Its graph is between the graphs of the 2- and 8-norm. As every univariate optimal allocation is optimal, the RMSEs have to be higher or equal in the multivariate case compared to the univariate RMSEs. As a consequence, the graphs in Figure 3 are located on or above the horizontal one-line. In the settings with (CV2)-standardization, the error-increases are not well balanced. Because

$$CV^2(\hat{\tau}_{SOC}^{StrRS}) > CV^2(\hat{\tau}_{INC}^{StrRS}) > CV^2(\hat{\tau}_{HAGE}^{StrRS})$$

for all appropriate allocations, the increase of the RMSE is smallest in variable SOC. In the min-max case ( $p = \infty$ ), the increase for SOC is zero, which means that the multivariate optimal allocation is equal to the univariate optimal allocation with respect to SOC.

In contrast, we observe a well balanced increase of the RMSEs for the (opt)-standardization because the  $p$ -norm of the relative change of the variances compared to the univariate optimal allocations is minimized. This results in a well compensated allocation. For  $p = \infty$  we obtain an almost equal increase.

In Figure 4 we plot the same settings as in Figure 3, but the RMSEs of the subtotal estimates of each of the 40 districts are presented. As before, the errors are illustrated relative to the errors when using the univariate optimal allocations. Dots which are located to the right of the vertical one-line correspond to subtotal estimates with an increase of the district specific RMSEs. Accordingly, dots to the left of the one-line correspond to estimates with a decrease. Again, the settings with the most compensated errors are those corresponding to the (opt)-standardization. Although the RMSEs in Figure 3 are higher than the univariate RMSEs, the multivariate allocation also leads to error-decreases in some districts, shown as points located to the left of the one-line in the boxplots.

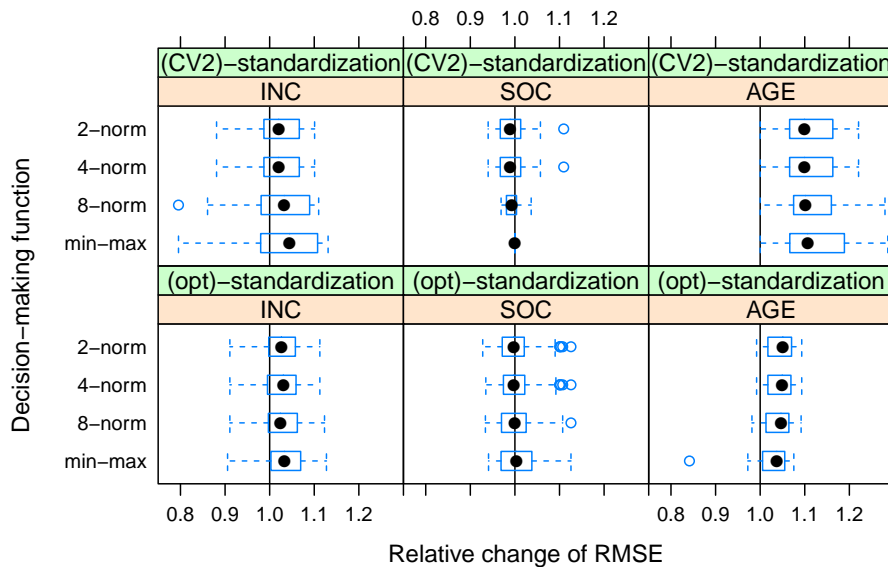


Figure 4: Relative change of RMSE for the estimated district totals for various standardization and scalarization techniques.

Figures 3 and 4 show that the (CV2)-standardization and a scalarization with a larger  $p$  accentuate single variables, in particular those with a comparably high CV. This contrasts compensatory methods which may be preferable in cases where no most important variable is obvious.

### 4.2. Comparison of variances depending on the chosen weights

#### Predefined weights

Here, we focus on the weighted sum as decision-making function. As illustrated in Section 2.3, this decision-making function facilitates the computation of the whole set of Pareto optimal solutions. We plot the relative increases of the district specific RMSEs for ten combinations of weights for the (CV2)-standardization in Figure 5 and for the (opt)-standardization in Figure 6. The relative error-increases of the total population estimates are comparatively

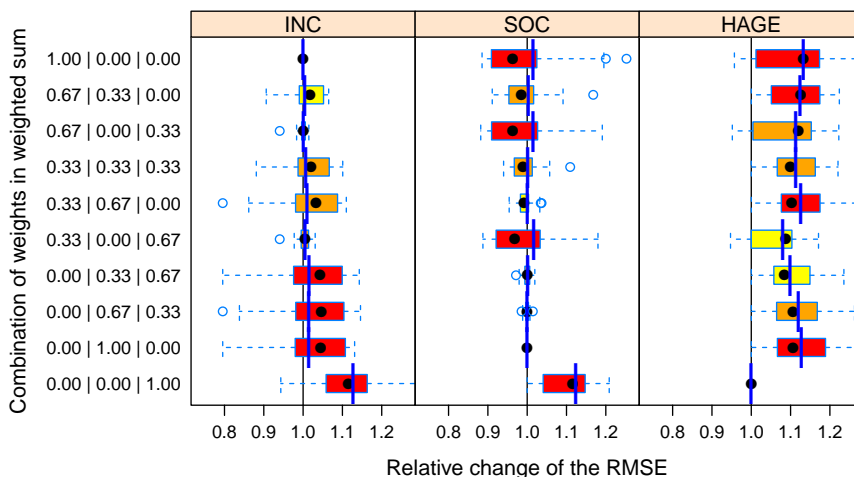


Figure 5: Relative change of the RMSE for the estimated district totals for ten combinations of weights with (CV2)-standardization. Red boxes correspond to a weight of 0.00, orange boxes to a weight of 0.33, and yellow boxes to a weight of 0.67 for the respective variable.

shown as vertical lines. The settings in row one, nine, and ten are equal to the univariate optimal allocations with respect to one of the three variables of interest, which is why the boxplots for the corresponding variables have no spread. In most cases, higher weights coincide with lower estimation errors of the district totals. Nevertheless, this coincidence is not a general statement and depends, among others, on the correlation structure of the variables of interest. Comparing the results of Figures 5 and 6, we observe more compensated error-

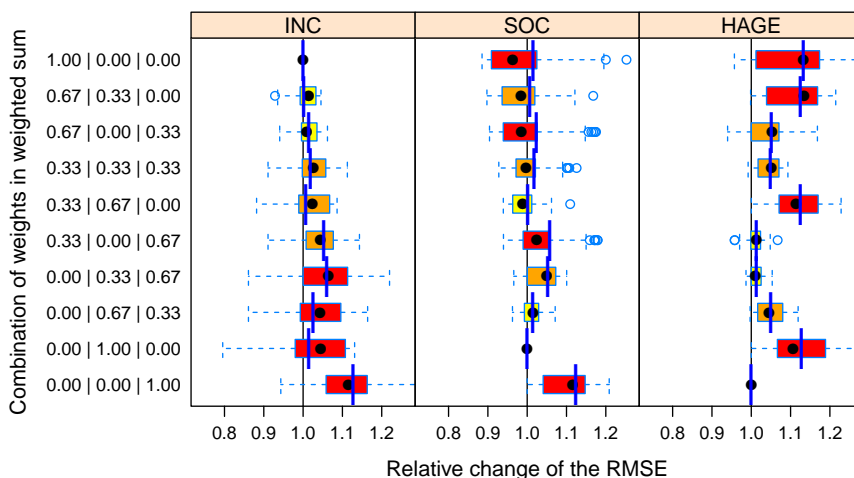


Figure 6: Relative change of the RMSE for the estimated district totals for ten combinations of weights with (opt)-standardization. Red boxes correspond to a weight of 0.00, orange boxes to a weight of 0.33, and yellow boxes to a weight of 0.67 for the respective variable.

increases over all variables and all districts using the (opt)-standardization than using the (CV2)-standardization. For example, in Figure 5 the variable SOC is dominant. If SOC is assigned any weight higher than zero, the increase in the error of the estimates is low for SOC, but high for the other variables of interest. This effect does not occur for the (opt)-standardization and the particular weight combinations  $w = (1/3, 1/3, 1/3)$  and  $w = (0, 1/3, 2/3)$  in Figure 6.

### Pareto optimization

To obtain a characterization of the Pareto frontier, we compute the multivariate optimal allocations for all possible combinations of weights with a resolution of 0.1.

In the heatmaps in Figure 7 we plot the increase of the variances of the total estimates with respect to the univariate optimal allocation variances. Each dot represents one combination of weights. The percentage weight for each separate variable is marked on the related axis. In consequence of the scaling resolution of 0.1, the dots which represent the equal weighting

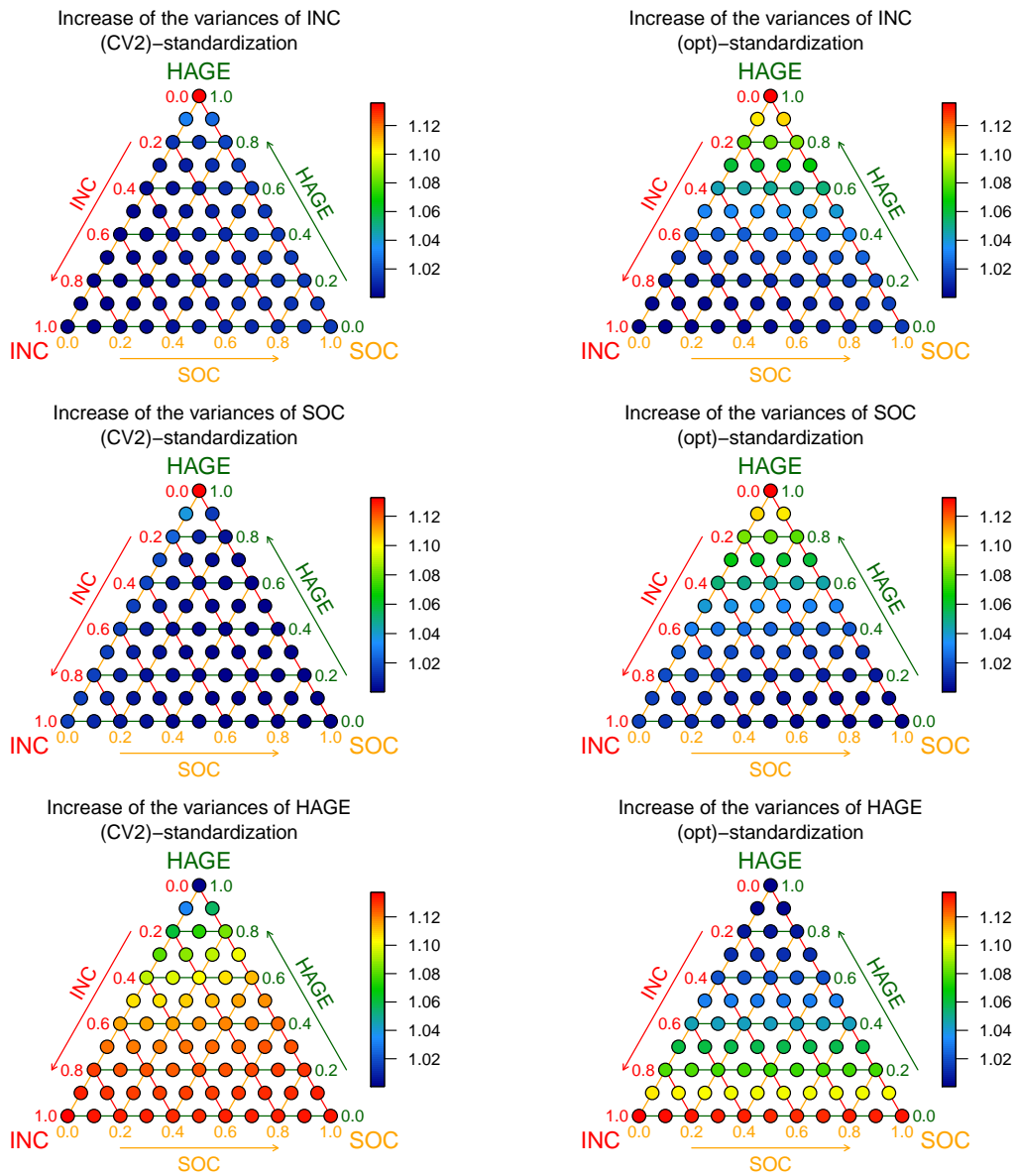


Figure 7: Relative increase of the variances of the population estimates under (CV2)- and (opt)-standardization for 66 combinations of weights for each variable of interest. The percentage weight for each separate variable is marked on the related axis.

$w = (1/3, 1/3, 1/3)$  and the weighting  $w = (0, 1/3, 2/3)$  are not contained in the heatmaps.

However, they can be accurately approximated by the surrounding dots. Blue dots are favorable because they represent combinations of weights with a lower increase of the variances. For example, the minimal variance for variable INC is located at the vertex where variable INC is given the full weight 1.00. The variances differ depending on the choice of the standardization strategy.

A similar behavior between the variances of INC and SOC can be observed because of their positive correlation of 0.27. In addition to that, the correlation between HAGE and INC as well as HAGE and SOC is smaller, which results in a higher error-increase of the total estimate of HAGE, even for roughly equal weights. Similarly to Section 4.1, the setting with (opt)-standardization is more balanced in the overall comparison of the heatmaps.

The structure of the heatmaps in Figure 8 is equivalent to Figure 7, but the *cumulated error-increase* of the total estimates of the three variables of interest is plotted. In the case of the (CV2)-standardization, the best choice is an asymmetric weighting. In contrast, the setting with (opt)-standardization is more balanced, so the best choice has roughly equal weights.

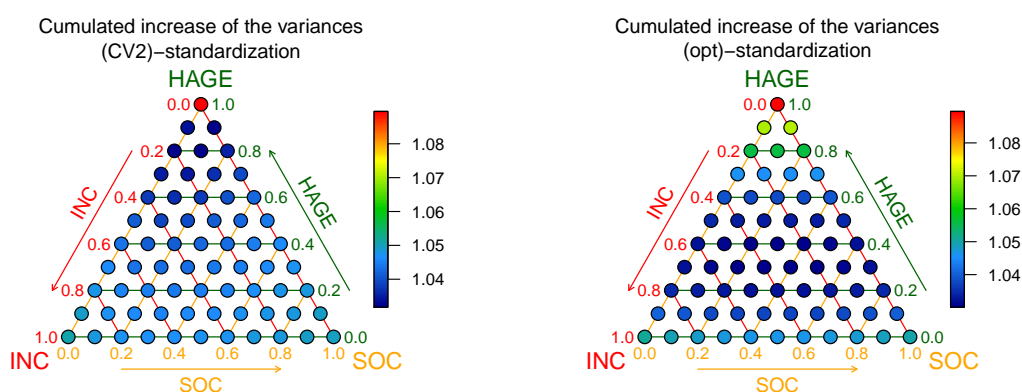


Figure 8: Relative cumulated increase of the variances of the population estimates under (CV2)- and (opt)-standardization for 66 combinations of weights. The percentage weight for each separate variable is marked on the related axis.

By Theorem A.1, each dot in the heatmaps represents the variance of one Pareto optimal solution. To be precise, the dots along the edges (where at least one weight is zero) are weakly Pareto optimal. Combining the heatmaps of the three variables of interest in one plot, we can display the Pareto frontier in Figure 9. Each dot in the three-dimensional space represents one Pareto optimal solution. Each of the three axes represents the error-increase for the corresponding variable. As already observed before, the (opt)-standardization results in a more balanced Pareto frontier.

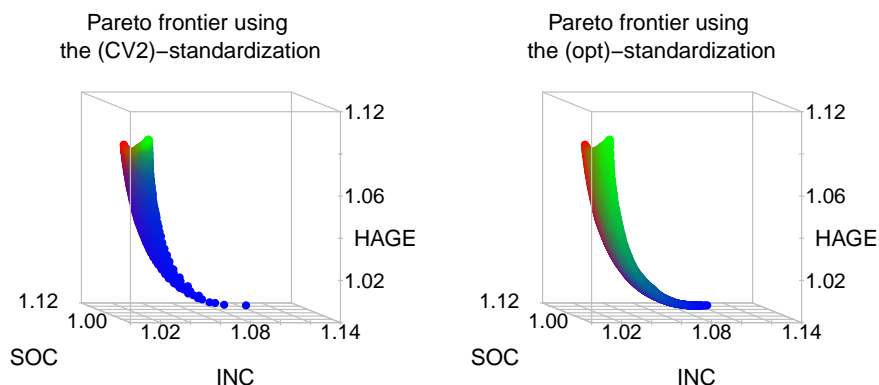


Figure 9: Pareto frontiers for the (CV2)-standardization and (opt)-standardization.

The evaluation of these plots offers valuable support for the decision maker to select the

preferred solution among all efficient solutions. By using the weighted sum and calculating the Pareto frontier, the decision is based on a higher level of reliable information. This contrasts using a  $p$ -norm that does not give the user the possibility to choose his preferred solution.

As the computation of the Pareto frontier requires the solution of many optimal allocation problems, it is only realizable in a practical time frame if efficient algorithms are used. We show in Section 4.5 that our algorithms are fast enough to facilitate this analysis of the Pareto frontier for multivariate allocation problems even for large problem instances. Moreover, this finding holds for both the continuous and integer problem.

### 4.3. Comparison of stratum-specific sample sizes

In Figure 10 the stratum-specific sample sizes are presented on the district level. Each boxplot contains the 40 districts and shows the relative change of the stratum-specific sample sizes compared to the equal weighting  $w = (1/3, 1/3, 1/3)$  in line four.

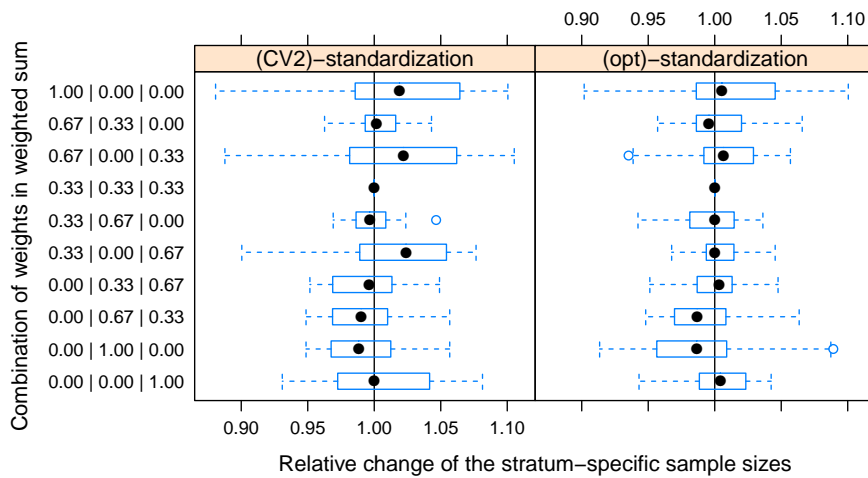


Figure 10: Relative differences in stratum-specific sample sizes of the districts.

On the one hand, there is a spread in the stratum-specific sample sizes depending on the weights (up to 12%), which illustrates the effect of the chosen weighting on the structure of the optimal allocation and the advantage of knowing the Pareto frontier. On the other hand, we recognize clear differences between the (CV2)- and (opt)-standardization. The relative changes of the sample sizes using (opt)-standardization is smaller.

### 4.4. Comparison of continuous and integer solution

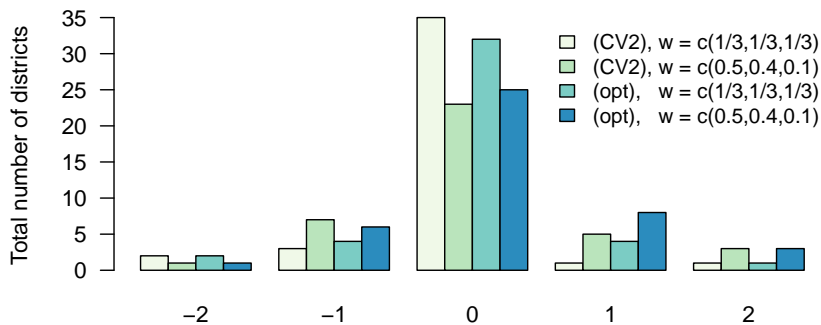


Figure 11: Absolute differences between district-specific sample sizes for rounded and integer allocation.



In Figure 11 we compare the rounded optimal continuous solution of the multivariate allocation problem, computed by the semismooth Newton method, with the optimal integer solution, computed by the Greedy method. We plot the cumulative differences in each district exemplary for two combinations of weights and both standardization techniques.

The rounded sample sizes differ from the optimal integer sample sizes and the differences vary from  $-2$  up to  $+2$  per district. The associated RMSEs are also different. This shows the advantage of the integer optimal allocation – especially when considering that the computing times of the continuous and integer solver are of the same magnitude (Friedrich *et al.* 2015).

#### 4.5. Performance of the algorithms

All the numerical results are computed in R on a desktop PC with an Intel Core i7-6700 CPU at  $3.40\text{GHz} \times 8$  and an internal memory of 32 GB.

Münnich *et al.* (2012b) and Friedrich *et al.* (2015) show that the fixed-point iteration or semismooth Newton method for the continuous problem as well as the Greedy algorithm for the integer problem have huge advantages in computing time compared to the R package `nloptr` which provides an R interface to the open-source library `NLopt` for nonlinear optimization. However, as pointed out in Section 2.2, the separability of the objective function is mandatory for these algorithms, so they can only be applied using a weighted sum or 2-norm as standardization technique. The following results are based on the 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  $\lambda^0$  for the continuous solvers is calculated as the mean of the three separate univariate optimal stratum-specific sample sizes.

Table 1: Performance of the semismooth Newton algorithm compared to `nloptr`.

|                     | R package<br><code>nloptr</code> | Semismooth<br>Newton |
|---------------------|----------------------------------|----------------------|
| Computing time [ms] | 6 801                            | 1                    |
| Iterations          | 242                              | 4                    |

The performance of the semismooth Newton algorithm is shown in Table 1 and Table 2. Using similar predefined precisions, we observe distinct improvements in computing time and the number of iterations compared to the R package `nloptr`. Moreover, by analyzing column two of Table 2, we can numerically confirm a quadratic convergence rate, which is proved (locally) for the semismooth Newton method in Qi and Sun (1993).

Table 2: Convergence of the semismooth Newton algorithm.

| Iteration $i$ | Residual $\ \Phi(\lambda^i)\ _2$ | Objetive $f(n(\lambda^i))$ |
|---------------|----------------------------------|----------------------------|
| 0             | $2.0 \cdot 10^2$                 | $6.0263 \cdot 10^{12}$     |
| 1             | $6.1 \cdot 10^0$                 | $6.1512 \cdot 10^{12}$     |
| 2             | $6.0 \cdot 10^{-3}$              | $6.1552 \cdot 10^{12}$     |
| 3             | $5.8 \cdot 10^{-9}$              | $6.1553 \cdot 10^{12}$     |
| 4             | $1.3 \cdot 10^{-11}$             | $6.1553 \cdot 10^{12}$     |

For the computing times of the Greedy methods for the integer allocation problem we refer to the detailed analysis in Friedrich *et al.* (2015), who, in particular, prove a worst-case bound on the running time. The computing times are generally longer than those for the fixed-point iteration in the continuous case, but still well below one second.

## 5. Conclusion

The optimization of stratified sampling designs must consider many requirements, such as conflicting variables of interest, cost restrictions, or the control of sampling fractions. This results in a multivariate optimal allocation problem under constraints.

We have proposed several scalarization and standardization techniques for the efficient solution of multivariate allocation problems. Whereas the scalarization reflects the decision function when evaluating conflicting goals, the standardization of the variances yields a rescaling of the variables fostering comparability. Furthermore, we have shown how the entire Pareto frontier as the set of all Pareto optimal solutions can be computed. The major benefit is the possibility of an a posteriori choice of a weighting scheme of the variables of interest, so that the decision maker is able to incorporate additional information to achieve the 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. Additionally, we have observed considerable differences in estimation errors and stratum-specific sample sizes when varying the weighting schemes. We can underline the importance of the chosen scalarization, standardization, and weighting in multivariate optimal allocation.

We have computed solutions for instances of the continuous and integer allocation problem using the AMELIA dataset. This simulation study presents the algorithms comparatively, underlines their advantages, and allows recommendations for their practical use. In contrast to standard solvers, using the separability and convexity of the given problem yields a substantial increase in the numerical performance, which enables calculating the Pareto frontier in high resolution. The integer algorithm avoids rounding. Moreover, the semismooth Newton method supports extensions with more general restrictions.

## Acknowledgements

This research was supported within the project *Research Innovation for Official and Survey Statistics* (RIFOSS), funded by the German Federal Statistical Office, and by the research training group 2126 *Algorithmic Optimization* (ALOP), funded by the German Research Foundation DFG. Finally, we thank the editor for his support and encouragement.

## References

- Ahsan MJ, Khan SU (1982). “Optimum Allocation in Multivariate Stratified Random Sampling with Overhead Cost.” *Metrika*, **29**, 71–78.
- Alfons A, Burgard J, Filzmoser P, Hulliger B, Kolb JP, Kraft S, Münnich R, Schoch T, Templ M (2011). “The AMELI Simulation Study. Research Project Report WP6–D6.1.” *Technical report*, AMELI. URL <http://ameli.surveystatistics.net>.
- Arthanari TS, Dodge Y (1981). *Mathematical Programming in Statistics*, volume 341 of *Wiley Series in Probability and Statistics*. Wiley, New York.
- Bankier MD (1988). “Power Allocations: Determining Sample Sizes for Subnational Areas.” *American Statistical Association*, **42**(3), 459–472.
- Brethauer KM, Ross A, Shetty B (1999). “Nonlinear Integer Programming for Optimal Allocation in Stratified Sampling.” *European Journal of Operational Research*, **116**(3), 667–680.
- 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.
- Cochran WG (1977). *Sampling Techniques*. 3rd edition. Wiley, New York.
- Dalenius T (1953). “The Multivariate Sampling Problem.” *Scandinavian Actuarial Journal*, **36**, 92–102.
- Díaz-García JA, Cortez LU (2006). “Optimum Allocation in Multivariate Stratified Sampling: Multi-Objective Programming.” *Technical report*, Centro de Investigación en Matemáticas, Guanajuato, México.
- Díaz-García JA, 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.
- Ehrgott M (2005). *Multicriteria Optimization*. 2nd edition. Springer, Heidelberg.
- Falorsi PD, Righi P (2008). “A Balanced Sampling Approach for Multi-Way Stratification Designs for Small Area Estimation.” *Survey Methodology*, **34**(2), 223–234.
- Falorsi PD, 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 PD, Righi P (2016). “A Unified Approach for Defining Optimal Multivariate and Multi-Domains Sampling Designs.” In *Topics in Theoretical and Applied Statistics*, pp. 145–152. Springer, Heidelberg.
- Folks JL, Antle CE (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*. Ph.D. thesis, Trier University.
- Friedrich U, Münnich R, de Vries S, Wagner M (2015). “Fast Integer-Valued Algorithms for Optimal Allocations under Constraints in Stratified Sampling.” *Computational Statistics and Data Analysis*, **92**, 1–12.
- Gabler S, Ganninger M, Münnich R (2012). “Optimal Allocation of the Sample Size to Strata under Box Constraints.” *Metrika*, **75**(2), 151–161.
- Hochbaum D (1995). “A Nonlinear Knapsack Problem.” *Operations Research Letters*, **17**, 103–110.
- Hohnhold H (2009a). “Generalized Power Allocations.” *Technical report*, Statistisches Bundesamt, Wiesbaden.
- Hohnhold H (2009b). “Variants of Optimal Allocation in Stratified Sampling.” *Technical report*, Statistisches Bundesamt, Wiesbaden.
- Huddleston HF, Claypool PL, Hocking RR (1970). “Optimal Sample Allocation to Strata Using Convex Programming.” *Journal of the Royal Statistical Society Series C*, **19**(3), 273–278.
- Jahn J (1986). *Mathematical Vector Optimization in Partially Ordered Linear Spaces*. Verlag Peter Lang, Frankfurt am Main.

- Khan MF, Ali I, Raghav YS, Bari A (2012). “Allocation in Multivariate Stratified Surveys with Non-Linear Random Cost Function.” *American Journal of Operations Research*, **2**, 100–105.
- Khan MGM, Khan EA, Ahsan MJ (2003). “An Optimal Multivariate Stratified Sampling Design Using Dynamic Programming.” *Australian and New Zealand Journal of Statistics*, **45**(1), 107–113.
- Kish L (1976). “Optima and Proxima in Linear Sample Designs.” *Journal of the Royal Statistics Society Series A*, **139**(1), 80–95.
- Kokan AR (1963). “Optimum Allocation in Multivariate Surveys.” *Journal of the Royal Statistics Society Series A*, **126**(4), 557–565.
- Kokan AR, Khan S (1967). “Optimum Allocation in Multivariate Surveys: An Analytical Solution.” *Journal of the Royal Statistics Society Series B*, **29**(1), 115–125.
- Lin JG (2005). “On Min-Norm and Min-Max Methods of Multi-Objective Optimization.” *Mathematical Programming*, **103**, 1–33.
- Lohr SL (2010). *Sampling: Design and Analysis*. 2nd edition. Cengage Learning, Boston.
- Merkle H, Burgard JP, Münnich R (2016). “The AMELIA Dataset - A Synthetic Universe for Reproducible Research.” In YG Berger, JP Burgard, A Byrne, A Cernat, C Giusti, P Koksel, S Lenau, S Marchetti, H Merkle, R Münnich, I Permanyer, M Pratesi, N Salvati, N Shlomo, D Smith, N Tzavidis (eds.), *Deliverable 23.1: Case studies*, volume WP23 – D23.1. URL <http://inclusivegrowth.be>.
- Münnich R, Sachs E, Wagner M (2012a). “Calibration of Estimator-Weights via Semismooth Newton.” *Journal of Global Optimization*, **52**(3), 471–485.
- Münnich R, Sachs EW, Wagner M (2012b). “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.
- Qi L, Sun J (1993). “A Nonsmooth Version of Newton’s Method.” *Mathematical Programming*, **58**(1), 353–367.
- Schaich E, Münnich R (1993). “Zum Allokationsproblem bei Mehreren Untersuchungsvariablen.” *Allgemeines Statistisches Archiv*, **77**, 390–405.
- Tschuprow A (1923). “On the Mathematical Expectation of the Moments of Frequency Distributions in the Case of Correlated Observations.” *Metron*, **2**, 461–493.
- Wagner M (2013). *Numerical Optimization in Survey Statistics*. Ph.D. thesis, Trier University.
- Ypma J, Borchers H, Eddelbuettel D (2014). “R Package nloptr: R Interface to NLOpt.” <http://CRAN.R-project.org/package=sampling>. R package version 1.04.

## A. Appendix: Optimality in multi-criteria optimization

It is in general not possible to find all Pareto-optimal points for a multi-criteria optimization problems by solving the weighted sum problem. In this section we analyze the weighted sum method for the multivariate allocation problem mathematically. While the sufficient condition of Theorem A.1 holds in a very general setting, see for example Folks and Antle (1965), this is not true for the necessary condition of Theorem A.3. We assume that the reader is familiar with the concept of optimality in multi-objective optimization and in particular with (weak) Pareto optimality. We refer to Ehrgott (2005, chapter 2) or Jahn (1986, chapter 4) for a detailed presentation.

**Theorem A.1** (Sufficient Condition). *Let  $D \subseteq \mathbb{R}^H$  and let  $f_k : D \rightarrow \mathbb{R}$ ,  $k = 1, \dots, K$ . For every optimal solution  $\bar{n}$  of  $\min_{n \in D} \sum_{k=1}^K w_k f_k(n)$  with weights  $w \in \mathbb{R}^K$ , the following statements hold.*

1.  $\bar{n}$  is a weakly Pareto optimal solution for  $\min_{n \in D} (f_1(n), \dots, f_K(n))$  if  $w \geq 0$ .
2.  $\bar{n}$  is a Pareto optimal solution for  $\min_{n \in D} (f_1(n), \dots, f_K(n))$  if  $w > 0$ .

*Proof.* Proposition 3.9 in Ehrgott (2005). □

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

**Lemma A.2.** *Let  $D \subseteq \mathbb{R}^H$  be convex and let  $f_k : D \rightarrow \mathbb{R}$ ,  $k = 1, \dots, K$ , be convex functions. Then the set  $C_+(f) := \{(f_1(n), \dots, f_K(n))^T \mid n \in D\} + \mathbb{R}_+^K$  is convex.*

*Proof.* Theorem 2.6 in Jahn (1986). □

**Theorem A.3** (Necessary Condition). *Let  $D \subseteq \mathbb{R}^H$  be convex and let  $f_k : D \rightarrow \mathbb{R}$  for  $k = 1, \dots, K$  be convex functions. Then, for each Pareto optimal solution  $\bar{n}$  of the problem  $\min_{n \in D} (f_1(n), \dots, f_K(n))$  there exist weights  $\bar{w} \in \mathbb{R}_+^K \setminus \{0\}$  such that  $\bar{n}$  is an optimal solution of the weighted sum problem  $\min_{n \in D} \sum_{k=1}^K \bar{w}_k f_k(n)$ .*

*Proof.* Using the convexity of the objective function, Lemma A.2 shows that the set  $C_+(f)$  mentioned in the lemma is convex. Using this property, the result follows directly from Theorem 5.4 in Jahn (1986). □

As the convexity assumption of Theorem A.3 holds for the optimal allocation problem formulated in (5), we can apply the theorem and we have proved that (up to discretization) the computations in Sections 3 and 4 describe the entire Pareto frontier of the problem.

**Affiliation:**

Ulf Friedrich  
DFG-RTG Algorithmic Optimization  
Trier University  
D-54286 Trier, Germany  
E-mail: [friedrich@uni-trier.de](mailto:friedrich@uni-trier.de)  
URL: <https://www.optmath.de>

Ralf Münnich  
Economic and Social Statistics Department  
Trier University  
D-54286 Trier, Germany  
E-mail: [muennich@uni-trier.de](mailto:muennich@uni-trier.de)  
URL: <https://www.uni-trier.de/index.php?id=58137>

Martin Rupp  
Economic and Social Statistics Department  
Trier University  
D-54286 Trier, Germany  
E-mail: [ruppm@uni-trier.de](mailto:ruppm@uni-trier.de)  
URL: <https://www.uni-trier.de/index.php?id=54776>