3-1-2018

Optimal Stratification and Allocation for the June Agricultural Survey

Jonathan Lisic
Cigna

Hejian Sang
Iowa State University, hjsang@iastate.edu

Zhengyuan Zhu
Iowa State University, zhuz@iastate.edu

Stephanie Zimmer
Iowa State University

Optimal Stratification and Allocation for the June Agricultural Survey

Abstract
A computational approach to optimal multivariate designs with respect to stratification and allocation is investigated under the assumptions of fixed total allocation, known number of strata, and the availability of administrative data correlated with thevariables of interest under coefficient-of-variation constraints. This approach uses a penalized objective function that is optimized by simulated annealing through exchanging sampling units and sample allocations among strata. Computational speed is improved through the use of a computationally efficient machine learning method such as K-means to create an initial stratification close to the optimal stratification. The numeric stability of the algorithm has been investigated and parallel processing has been employed where appropriate. Results are presented for both simulated data and USDA's June Agricultural Survey. An R package has also been made available for evaluation.

Keywords
Area survey, optimal allocation, optimal stratification, multivariate design, simulated annealing

Disciplines
Agricultural Economics | Applied Statistics | Multivariate Analysis | Statistics and Probability

Comments
This article is published as Lisic, Jonathan, Hejian Sang, Zhengyuan Zhu, and Stephanie Zimmer. "Optimal Stratification and Allocation for the June Agricultural Survey." *Journal of Official Statistics* 34, no. 1 (2018): 121-148. DOI: 10.1515/JOS-2018-0007. Posted with permission.

Journal of Official Statistics, Vol. 34, No. 1, 2018, pp. 121–148, http://dx.doi.org/10.1515/JOS-2018-0007

Optimal Stratification and Allocation for the June Agricultural Survey

Jonathan Lisic[1], *Hejian Sang*[2], *Zhengyuan Zhu*[2], *and Stephanie Zimmer*[2]

A computational approach to optimal multivariate designs with respect to stratification and allocation is investigated under the assumptions of fixed total allocation, known number of strata, and the availability of administrative data correlated with thevariables of interest under coefficient-of-variation constraints. This approach uses a penalized objective function that is optimized by simulated annealing through exchanging sampling units and sample allocations among strata. Computational speed is improved through the use of a computationally efficient machine learning method such as K-means to create an initial stratification close to the optimal stratification. The numeric stability of the algorithm has been investigated and parallel processing has been employed where appropriate. Results are presented for both simulated data and USDA's June Agricultural Survey. An R package has also been made available for evaluation.

Key words: Area survey; optimal allocation; optimal stratification; multivariate design; simulated annealing.

1. Introduction

An attribute of many federal surveys is the use of a stratified design. Stratified designs allow the inclusion of knowledge about the Primary Sampling Units (PSU)s in a population through administrative variables that are well correlated with the desired estimators. Many federal surveys have the additional requirements of agency-mandated quality constraints. These constraints are typically based on the Coefficient-of-Variation (CV) or other functions of variance placed on either administrative variables or the survey estimates. Besides quality constraints, financial constraints are also imposed on federal surveys. This brings forth the question, "How can a federal survey practitioner optimally stratify and allocate a survey to meet imposed quality constraints without spending any more money?" In this article, a solution to this question is presented for the case of CV quality constraints and a fixed sample size.

An important aspect of this question is the concept of an optimal stratified design, where an optimal stratified design implies the joint optimization of stratification and allocation with respect to a predeterminated objective function. If an optimal design can be found, then it is optimal over all possible pairings of stratifications and allocations under the design constraints. Joint optimization differs from optimal stratification with an *a priori*

[1] Cigna, 900 Cottage Grove Rd, Bloomfield, CT 06002, U.S.A. Email: jonathan.lisic@cigna.com
[2] Iowa State University, Osborn Dr, Ames, IA 50011 U.S.A. Emails: hjsang@iastate.edu, zhuz@iastate.edu, and sazimme2@iastate.edu.

allocation or optimal allocation conditioned on a prior fixed stratification. Here a priori allocation is defined as designs where the allocations are predeterminated functions of the strata population sizes, such as proportional, uniform or other allocation methods that do not admit administrative data. Both the a priori and conditional allocation place additional constraints on the objective function. In the case of conditional allocation, the objective function used for stratification lacks information about the optimal allocation, making it necessary to optimize an alternative objective function. If an optimal allocation is performed with the desired objective function, then the allocation is restricted by the prior stratification. This restriction can lead to a nonoptimal design. In the case of a priori allocation, only a subset of pairings of all possible stratifications and allocations are considered; this subset is unlikely to contain the optimal allocation for a given objective function due to ignoring administrative data. The importance and improvements provided by assuming neither a priori allocation nor using conditional allocation are discussed and displayed through empirical results in Benedetti et al. (2008), Day (2009), Baillargeon and Rivest (2009), and Ballin and Barcaroli (2013). A comparison of a priori allocated designs for multivariate surveys can be found in Kozak (2006b); further discussion can be found in Gonzalez and Eltinge (2010).

One major advantage that a priori and conditional allocation designs have over optimal stratified designs is that they are easy to obtain. Optimal stratified designs require an exploration of a combinatorial space to find an optimal design. This is a nontrivial problem for even small population and sample sizes. A solution to the problem of finding a univariate optimal stratified design subject to a CV constraint using Neyman allocation for a fixed sample size was proposed by Dalenius and Hodges (1959). This method is commonly known as the cum $\sqrt{f}$ method (Särndal et al. 1991, Section 3.7) (Horgan 2006). Lavallée and Hidiroglou (1988) and the multivariate extensions in Benedetti et al. (2010) and Benedetti and Piersimoni (2012) provide optimal designs under CV constraints, but restrict the strata to either two or three stratum. These stratum in Benedetti and Piersimoni (2012) include a census (take-all) and sampled (take-some) strata and do not restrict the sample size. Benedetti et al. (2010) included a third (take-none) stratum for the purposes of cut-off sampling. These approaches are designed for highly skewed populations, exploiting the similarity of the underlying population to a geometric progression (Gunning et al. 2004). Benedetti and Piersimoni (2012) introduced a method for stratification which uses multiple administrative variables. This method, which is motivated by the Lavallée and Hidiroglou method, partitions the population into two strata, one which is sampled and one, which is a take-all stratum. The partitioning is determined such that the sample size is minimized for a target coefficient of variation of a response variable. In addition to allocations with goals of increasing precision, allocations also consider data collection costs and other practical constraints such as the method proposed by Valliant et al. (2014) to allocate sample in household surveys using Address-Based Sampling Frames and available commercial data.

Other multivariate approaches to optimal stratified designs can be found in Ballin and Barcaroli (2013) and Benedetti et al. (2008). Both of these methods are designed to work on a set of categorical administrative variables. These administrative variables work as a means of data reduction by assigning each PSU to an initial stratum, called an atomic stratum, defined by a unique combination of administrative values. To admit optimization under variance or CV constraints, each atom is assigned a variance estimate. Simultaneous allocation in both cases is performed through optimal allocation as defined in Bethel (1986), where strata allocations are round up to the nearest integer. The major difference between these algorithms is how they explore the combinatorial space of the atomic strata. A divisive tree-based approach is used in Benedetti et al. (2008). In this approach, at each layer of the tree, a stratum is split by the administrative variable that results in the greatest reduction of total sample size according to optimal allocation in Bethel (1986). This is continued until a set of CV quality constraints are met. In Ballin and Barcaroli (2013), a Genetic Algorithm (GA) is used to explore the space of strata formed by merging atomic strata. At each iteration (generation) of the GA, a large collection of possible stratifications are generated and evaluated through minimum sample size under CV constraints. A set of sufficiently well-performing stratifications and allocations and a small number of less optimal stratifications and allocations are retained to contribute to future generations. The less optimal stratifications and allocations are retained to provide genetic diversity. These stratifications and allocations, along with combinations and mutations of these stratifications are then used as the next generation. Combinations of strata are formed by exchanging atomic strata assignments between two stratifications, and mutations are formed by randomly assigning an existing atomic stratum to another stratum. Iteration is continued until changes in the objective function plateau. Both of these approaches consider a variable number of strata with means to specify a maximum number of strata to avoid an unstable stratified design.

A separate but important issue not directly addressed by the previously mentioned works involves the relationship between the administrative variables being optimized and the desired estimators. Even when the administrative variables and the desired estimators are highly correlated, the optimal stratification and allocation under the administrative variables may not be optimal for the desired estimators. In particular, an assumption that meeting quality constraints for the administrative variables may not imply meeting assumed quality constraints for the desired estimators. A discussion of this issue and proposed solution for univariate stratified designs using anticipated moments can be found in Baillargeon and Rivest (2009). Anticipated moments are moments of a random variable calculated under the sample design and the super population model (Isaki and Fuller 1982). When the super population model, referred to as the model in this article, is correctly specified or a sufficiently robust model is used, it is possible to construct strata that on average meet the quality constraints for the desired estimators.

The prior literature on the subject of optimal stratified designs does not address three important use cases; (1) multivariate optimal allocation with continuous administrative variables with more than two strata, (2) multivariate optimal allocation using anticipated moments to attain CV constraints for desired estimators, (3) application to fixed sample sizes with CV constraints. In this article, a method to construct optimal multivariate stratified designs for an arbitrary, but fixed, number of self-representing strata from continuous valued administrative data is presented. This method admits a combination of *hard* and *soft* constraints, where soft constraints are handled by a penalized objective function and hard constraints are handled through traditional nonlinear programming constraints. Anticipated moments can be used within the objective function to account for the relationship between administrative variables and desired response. Accounting for this relationship allows for optimization with CV constraints on the desired estimators. Optimization of this objective function is performed by simulated annealing, by moving individual PSUs between strata. The use of soft constraints allows a survey practitioner to find potential solutions by relaxing less important constraints. Unlike prior multivariate methods, this choice of simulated annealing for optimization provides the theoretical result of guaranteed convergence to the global optima, and good performance characteristics. Simultaneous stratification and allocation is provided by also considering changes in the allocation as part of the simulated annealing algorithm.

The problem of targeting multiple responses that are not-necessarily correlated with each other is a characteristic of area surveys in agriculture. Agricultural production of crops such as corn excludes using the land for another purpose such as growing soybeans. The agricultural area survey examined in this article is the United States Department of Agriculture (USDA) National Agricultural Statistics Service's (NASS) June Agricultural Survey (JAS). In particular, a proposed redesign of JAS using a permanent and fixed area frame is examined. Prior work on optimizing the existing JAS design using simulated annealing procedures has been proposed by Gentle and Perry (2000). This work focused on creating strata that are homogeneous with respect to remote sensing imagery. Given the quality of remotely sensed imagery at the time of publication, this approach provided remarkable improvements in efficiency. However, the approach did not consider optimal allocation, CV constraints, sampling unit dependent costs, or agricultural practices such as crop rotations. The application of the proposed method does consider all four topics and can be consider a modern revisit of the topic with the benefit of higher quality remote sensing data and faster computing resources.

This article is broken into the following sections. In Section 2, details on the proposed method, including the algorithm and objective function, are presented. In Section 3, simulated data are used to illustrate the proposed method and the result is compared to those from other stratification and allocation methods. In Section 4, the JAS is introduced and the results of applying the proposed method to the JAS are compared to those from the current univariate allocation method. The article concludes with a discussion and future extensions to the proposed method that would account for measurement error and improve computational efficiency.

2. Optimal Stratified Design Algorithm

The method proposed here uses a sequence of exchanges of PSUs between a set of initial strata to improve an objective function. The objective function is a weighted vector norm applied to the vector of administrative variable or modeled CVs attained by the current stratification and allocation. A penalty function is added to this objective function, where the penalty function is the sum of the element-wise products of penalty weights and penalty values. The penalty weights serve as importance weights, as in Kozak (2006b) with the exception that the weights can be set to zero once the constraint is met. The penalty values are the difference between the attained CVs and the target CVs for each administrative or modeled variable. This approach can be considered a weighted or approximate constraint satisfaction problem in operations research (Freuder and Wallace 1992). Traditional hard constraints can be considered by setting the penalty weight to infinity. By allowing a combination of hard and soft constraints, survey managers and stakeholders have flexibility in identifying essential CV constraints and nonessential, but desirable, CV constraints. By defining these constraints separately, infeasibility of the constrained design by fixed sample sizes can potentially be avoided. If the nonessential constraints are violated, a design that minimizes the departure from the nonessential CV constraints can be found, and the constraints can be prioritized through the choice of penalty values.

The objective function optimizes the stratification and allocation through functions of moments; in particular, the population total and variance. The population total and variance are calculated from values assigned to individual PSUs. Therefore, either administrative data highly correlated with the desired response or a modeled response is required to find optimal stratification and allocation for a set of desired estimators. In the case of administrative data, it is assumed that the data is complete and available for each PSU. In the case of modeled response, it is similarly assumed that a model can be constructed for each PSU; variances based on this model can be incorporated within the objective function through anticipated moments. The later case will be discussed in Section 3.

The strata formed by PSU exchanges are self-representing. Self-representing strata are defined by PSU assignments, as opposed to any bounds on the administrative variables. Since the self-representing strata are not defined by a set of hyper-planes from administrative data bounds, they allow for strata that have nonlinear partitions and possibly disjoint subsets of the space of the administrative data.

Self-representing strata are also found in Benedetti et al. (2010) and Benedetti and Piersimoni (2012), however the approach presented can exceed three strata and is not restricted to highly asymmetric populations. Instead, it relies on the observation that, given an initial allocation and stratification, then a sequence of exchanges can be taken along primarily stratum boundaries to attain a more optimal design. Optimal allocation is performed simultaneously by potentially changing the sample size at each iteration. Optimization of this problem is performed by a stochastic optimization method known as simulated annealing (Metropolis et al. 1953). This optimization method is also used in Benedetti et al. (2010) and Benedetti and Piersimoni (2012). This stochastic optimization method is a metaheuristic that uses a Monte Carlo method to obtain an optimal solution by generating a sequence of possible solutions that slowly converge to an optimal solution. Simulated annealing has the useful property of being able to explore nonoptimal PSU exchanges and sample size changes, allowing for a more exhaustive search of the feasible region than deterministic optimization methods. This property allows for simulated annealing to guarantee convergence of the sequence to an optimal solution given sufficient run-time and precision.

Computational speed of the algorithm can be accelerated by starting with an initial stratification and allocation close to the optimal stratification and allocation. Such a stratification can be obtained through machine-learning methods such as K-means, and optimal allocation can be provided though the popular multivariate optimal allocation approach of Bethel (1986) with minor adjustments to ensure integer allocation and sample size constraints. It should be noted that, given hard constraints, the initial starting point must be in the feasible region; otherwise, single PSU exchanges are unlikely to result in a finite objective function. Furthermore, the proposed method can also be used for optimal allocation, by running with only allocation changes. Additional hard and soft constraints may also be added to the objective function. Such constraints include costs for PSU collection, bounds on maximum stratum size, or spatial penalty functions.

An implementation of the proposed method has been written in the R Language (R Core Team 2015), as an R package (Lisic 2016). This package supports second order moments through per-PSU additive components and scaling. These second-order moment adjustments allow for the adoption of linear or linearized relationships between administrative variables and survey response. Furthermore, the additive per-PSU component can be used to impose fixed per-PSU costs.

In this section the objective function is presented using administrative data with weighting. Anticipated moments are introduced, followed by a discussion of the simulated annealing algorithm.

2.1. *Objective Function*

Multivariate objective functions are vector valued functions that map from $\mathbb{R}^J \rightarrow \mathbb{R}$, where J is the length of the vector input. If the objective function is bounded over the domain of the problem, then both a maximum and a minimum exist. Constrained objective functions carve out a subset of the unbounded region known as the feasible region. In methods to find optimal designs, the goal is to either minimize or to maximize an objective function over this feasible region. Since all maximization problems can be written as minimization problems by multiplying the objective function by negative one, only minimization problems will be considered in this article.

For constrained objective functions, the feasible region may be empty, implying a solution does not exist. This can be trivially seen when unrealistically small CV constraints are imposed with fixed sample sizes. A way to avoid this issue is through replacing hard constraints with soft constraints. Soft constraints can be violated without reducing the feasible region, but at the expense of increasing the objective function. A standard way of implementing soft objective functions is through the use of a penalty function. Penalty functions impose a positive-valued penalty for violating a constraint.

The unpenalized objective function is a p-norm of the vector of CVs for a stratified survey design with simple random sampling (SRS), SRS with replacement is shown for brevity,

$$\|f(\mathbf{X}|\mathbf{I},\eta)\|_p = \|\left(f(\mathbf{X}_{\cdot,1}|\mathbf{I},\eta),\ldots,f(\mathbf{X}_{\cdot,j}|\mathbf{I},\eta),\ldots,f(\mathbf{X}_{\cdot,J}|\mathbf{I},\eta)\right)\|_p, \tag{1}$$

where

$\mathbf{x}_i =$ the vector valued administrative variable of length J available for all PSUs, identified with row i from matrix $\mathbf{X}$;
$\|\mathbf{x}_i\|_p =$ a p-norm of vector x_i equal to $\left(\sum_{j=1}^J x_{i,j}^p\right)^{1/p}$;
$\mathbf{I} =$ a vector of strata assignment parameters;
$\eta =$ a vector of sample sizes for each stratum;
$f(\mathbf{X}_{\cdot,j}|I,\eta) = \frac{\sqrt{\sum_h^H \eta_h^{-1} N_h^2 S_{h,j}^2}}{T_j}$ the CV for the j^{th} characteristic, $T_j = \sum_{i=1}^N x_{i,j}$ and $S_{h,j}^2 = \text{var}\left(\sum_{i=1}^N x_{i,j}\mathbb{1}_{\mathbf{I}_i=h} N_h^{-1}\right)$ with $h \in \{1,\ldots,H\}$ strata.

It is assumed that the number of strata is known *a priori*, the goal is to estimate the Horvitz-Thompson estimators for population totals, each element of $\mathbf{X}$ is nonnegative, and there is at least one positive valued $x_{i,j}$ for each j. To avoid issues with dividing by zero, it is also assumed that all strata have a minimum sample and population size of two. An extension to probability proportional to size sampling has been developed, but only SRS will be covered in this article.

If a set of quality constraints c are set through target CVs for $J^* \leq J$ administrative variables with a fixed total sample size, then for a given set of strata the problem can be written as

$$\operatorname{argmin}_{\zeta,\mathbf{I}} \|\mathbf{\Phi} f(\mathbf{X}|\mathbf{I}, \eta)\|_p \tag{2}$$

subject to

1. $c_j \geq T_j^{-1}\sqrt{\sum_h^H \zeta_h N_h^2 S_{h,j}^2}, \; j \in \{1, \ldots, J^*\}$;
2. $\sum_h^H \zeta_h^{-1} = n$;
3. ζ_h^{-1} is an integer;
4. each $0 < \zeta_h \leq 1/2$.

where $\mathbf{\Phi}$ is a square matrix of dimension $J \times J$ with a diagonal equal to a vector of positive valued penalty weights. The diagonal of penalty weights from $\mathbf{\Phi}$ is identified as the vector ϕ; elements of ϕ help prioritize reduction of the CVs, regardless of the target CVs being met or not.

Soft constraints can be added to the objective function through the dot product of the penalty weights $\boldsymbol{\lambda}$ and penalty value vector $g(\mathbf{X}|\mathbf{c}, \mathbf{I}, \boldsymbol{\eta})$:

$$\operatorname{argmin}_{\zeta,\mathbf{I}} \|\mathbf{\Phi} f(\mathbf{X}|\mathbf{I}, \boldsymbol{\eta})\|_p + \sum_{j \in J^{**}} \lambda_j g\left(\mathbf{X}_{\cdot,j}|\mathbf{I}, \boldsymbol{\eta}\right) \tag{3}$$

where J^{**} is the set of administrative variables with soft constraints subject to the same population and sample constraints in (2).

Each element of the penalty value vector $g(\mathbf{X}|\mathbf{I}, \boldsymbol{\eta})$ is equal to the maximum of 0 or $f(\mathbf{X}_{\cdot,j}|\mathbf{I}, \boldsymbol{\eta}) - c_j$. By this objective function, if all constraints are met, the problem is simply minimizing the norm of the vector formed by the product of $\mathbf{\Phi}$ and the vector of CVs. It is possible to simplify (3) by removing the hard constraints and replacing them with soft constraints using infinite valued penalty weights.

The choice of penalty weight vector $\boldsymbol{\lambda}$ can be motivated by targeting specific variables over others, or as a method to relate the administrative variables to the targeted response variables $\mathbf{Y}$. In the latter case, there are two potential solutions. The first would be to consider weights proportional to the absolute value of the correlation between the administrative variable and the response. This would favor a reduction in target CVs for variables with stronger relationships between $\mathbf{x_j}$ and $\mathbf{y_j}$ over weaker relationships for $j \in \{1, \ldots, J\}$. The second approach is the use of anticipated moments to explicitly model the response in the objective function. For brevity, only the second approach is covered.

Categorical administrative variables can be used through binning or grouping PSUs into disjoint sets identified by unique categorical values. This allows for the accommodation of industry by occupational groupings in establishment surveys or census blocks in area surveys.

2.2. Anticipated Moments

In practice, we are interested in estimating the Horvitz-Thompson estimators of unknown population characteristics that are correlated with an administrative variable. In Subsection 2.1, CV constraints placed on administrative variables serve as proxies for quality constraints on the estimators for the unknown population characteristics. A more direct approach is to place the CV constraints on the unknown population characteristic $\mathbf{Y}$ by an assumed model. This is accomplished by substituting the moments of $\mathbf{X}$ in (3), namely T_j and $S^2_{h,j}$, of the objective function with the complimentary anticipated moments of $\mathbf{Y}$. This is the multivariate generalization of the univariate approach by Baillargeon and Rivest (2009).

The exact form of the objective function is dependent on the choice of model for $\mathbf{Y}$ given $\mathbf{X}$. For many establishment surveys where y_i is a scalar, the model

$$y_i = x_i\beta + x_i^{\gamma}\epsilon_i, \tag{4}$$

$\mathbb{E}[\epsilon_i] = \mathbb{E}[\epsilon_i, \epsilon_{i'}] = 0$ where $(i \neq i')$, and $\mathbb{E}[\epsilon_i^2] = \sigma^2$ can provide a reasonable model (Kott et al. 2000). In the multivariate case considered, a generalization of this model for vectored value $\mathbf{y}_i$ is

$$\mathbf{y}_i = \mathbf{x}_i\mathbf{B} + \mathbf{V}_i\boldsymbol{\epsilon}_i, \tag{5}$$

$\mathbb{E}[\boldsymbol{\epsilon}_i] = \mathbb{E}[\boldsymbol{\epsilon}_i^T\boldsymbol{\epsilon}_{i'}] = 0$ where $(i \neq i')$, $\mathbb{E}[\boldsymbol{\epsilon}_i^T\boldsymbol{\epsilon}_i] = \Sigma$, and $\mathbf{V}_i$ is a symmetric matrix of heteroscedastic weights for Σ.

To integrate the modeled response into the objective function, Anticipated Variance (AV) is used. Anticipated variance is simply the expectation both using the model ($\mathbb{E}_m$) and design, ($\mathbb{E}_d$),

$$\mathrm{AV}(\hat{T}_j) = \mathbb{E}_m\mathbb{E}_d\left[\hat{T}_j - \mathbb{E}_m\mathbb{E}_d[\hat{T}_j]\right]^2. \tag{6}$$

In the multivariate simple linear regression case with heteroscedastic variance, AV takes the form

$$\mathrm{AV}(\hat{T}_j) = \sum_{h=1}^{H}\left(1 - \frac{n_h}{N_h}\right)\frac{N_h^2}{n_h}\left(\sum_{i\in U_h} v_{i,j}^2\sigma^2 + \sum_{i\in U_h}(x_iB_j - \bar{x}_hB_j)^2\right). \tag{7}$$

Likewise, the Horvitz-Thompson estimator of the population total is,

$$\mathbb{E}_m\mathbb{E}_d[\hat{T}_j] = \sum_{i=1}^{N} = x_iB_j \tag{8}$$

and the anticipated coefficient of variation (ACV) can be estimated as

$$\widehat{ACV}(\hat{T}_j) = \frac{\sqrt{\mathrm{AV}(\hat{T}_j)}}{\sum_{i=1}^{N} x_iB_j} \tag{9}$$

Although not explored in this article, other models could be used in this framework.

2.3. Simulated Annealing

To minimize the objective function (3), we only make changes to $\boldsymbol{\eta}$ and $\mathbf{I}$. Since $\mathbf{I}$ is binary, this optimization is a combinatorial optimization problem, where simulated annealing is applicable. Simulated annealing is a stochastic optimization process that minimizes an objective function (possibly with constraints), and avoids the pitfalls of ending up in a local minima by admitting nonoptimal states. The general form of an algorithm to perform this stochastic process on the objective function m optimized over parameters θ in the finite dimensional parameter space Θ is detailed in Figure 1. Line five is the key component of the simulated annealing algorithm, where nonoptimal states can be accepted with nonzero probability ρ. This probability decreases as the number of iterations increases, allowing for both early exploration and eventual convergence of the simulated annealing sequence. The sequence $t(l)$ is called the cooling schedule and is a nonincreasing function that governs how quickly the probability of accepting a nonoptimal state decreases. Examples of $t(l)$ include $(l+1)^{-1}$ and $(\log(l+1))^{-1}$. The algorithm continues until either a fixed number of iterations L or threshold δ are met.

An advantage that simulated annealing has over other searches of binary spaces as seen in Benedetti et al. (2008) and Ballin and Barcaroli (2013) is the guaranteed theoretical convergence to a global minima by the simple condition that there is a non-zero transition probability between all possible states (Hajek 1988). This is not the case in Benedetti et al. (2008) where strata splits are chosen not by global optimality but local optimality. Similarly, genetic algorithms do not guarantee convergence to a global optimal solution in general.

One disadvantage of simulated annealing is its computational speed. Simulated annealing can be quite slow relative to other methods such as tree based methods that can partition a large number of sampling units at once. Similarly, genetic algorithms easily admit parallel implementations as opposed to simulated annealing which has serial dependence between each iteration (Henderson et al. 2003).

Algorithm 1 Simulated annealing algorithm.

1: **while** $l \leq L$ or $|m(\theta^{(l)}) - m(\theta^{(l-1)})| < \delta$, for objective function m **do**
2: Randomly generate a candidate state $\theta_*^{(l)}$, $l \geq 1$
3: **if** $\theta_*^{(l)}$ has a lower objective function than $\theta^{(l-1)}$ **then**
4: set $\theta^{(l)} = \theta_*^{(l)}$
5: **else if** $\rho = exp\{\Delta h_l / t(l)\} \geq U_l$, where $\Delta h_l = m(\theta^{(l-1)}) - m(\theta_*^{(l)})$ and $U_l \sim \text{Uniform}(0,1)$ **then**
6: set $\theta^{(l)} = \theta_*^{(l)}$
7: **else**
8: set $\theta^{(l)} = \theta^{(l-1)}$
9: **end if**
10: **end while**

Fig. 1. Pseudocode for the simulated annealing algorithm.

Simulated annealing applied to strata formation and allocation is straightforward and detailed in Figure 2. Each new candidate state consists of a PSU exchange and a change in allocation. The PSU exchange is generated by selecting a single PSU and a stratum to move it to. The change in allocation is generated by increasing the sample allocation for a randomly selected stratum by one, and decreasing the stratum allocation for another stratum by one. The new stratum can be the same stratum in which the PSU resides, likewise there may be no change in allocation. To help improve the chance of retaining optimal allocation, multiple allocation exchanges are allowed per PSU change. In this case, subsequent assignment changes are only accepted if they improve the proposed objective function. In practice, the number of assignment changes required to maintain near-optimal allocation in each iteration is small for nonhighly skewed populations. This is due to a change in allocation dominating objective function changes when the sampling fraction is small.

In this application only linear cooling functions will be used, $t(l) = \alpha(l+1)^{-1}$ where α is a tunable parameter. Hard constraints on the objective function are handled by generating states that satisfy the imposed constraints.

Although Benedetti et al. (2010) and Benedetti and Piersimoni (2012) also uses simulated annealing, these methods differ in PSU selection. In the prior two papers, each PSU is iteratively selected ensuring each member of the population is offered a chance to move strata in a finite time-frame. The random search approach does not make this guarantee. Instead, it is assumed that for a sufficient number of iterations all PSUs are likely to be visited at least once. Furthermore, the introduction of nonuniform weighting in PSU selection for random searches could greatly improve performance of the proposed method by considering more likely PSU exchanges near stratum boundaries more frequently than more-extreme valued PSUs.

Algorithm 2 Simulated annealing algorithm for optimal stratification and allocation.

Start with initial stratification $\mathbf{I}^{(0)}$ and allocation $\boldsymbol{\eta}^0$
while $l \leq L$ or $m(\mathbf{y}|\,\mathbf{I},\boldsymbol{\eta}) - m(\mathbf{y}|\,\mathbf{I},\boldsymbol{\eta}) < \delta$, for objective function m **do**
  Randomly generate a candidate state $\mathbf{I}_*^{(l)}$, $l \geq 1$
  Randomly generate a candidate allocation $\boldsymbol{\eta}_*^{(l)}$ with respect to $\mathbf{I}_*^{(l)}$ to obtain $\boldsymbol{\eta}_*^{(l)}$
  if $\left(\mathbf{I}_*^{(l)}, \boldsymbol{\eta}_*^{(l)}\right)$ has a lower objective function than $\left(\mathbf{I}^{(l-1)}, \boldsymbol{\eta}^{(l-1)}\right)$ **then**
    set $\left(\mathbf{I}^{(l)}, \boldsymbol{\eta}^{(l)}\right) = \left(\mathbf{I}_*^{(l)}\boldsymbol{\eta}_*^{(l)}\right)$
  else if $\rho = exp\{\Delta h_l/t(l)\} \geq U_l$, where $\Delta = m\left(\mathbf{I}^{(l)}, \boldsymbol{\eta}^{(l)}\right) - m\left(\mathbf{I}_*^{(l)}\boldsymbol{\eta}_*^{(l)}\right)$ and $U_l \sim \text{Uniform}(0,1)$ **then**
    set $\left(\mathbf{I}^{(l)}, \boldsymbol{\eta}^{(l)}\right) = \left(\mathbf{I}_*^{(l)}\boldsymbol{\eta}_*^{(l)}\right)$
  else
    set $\mathbf{I}^{(l)} = \mathbf{I}^{(l-1)}$
  end if
end while

Fig. 2. Pseudocode for the simulated annealing algorithm applied to joint stratification and allocation.

The allocation approach is similar to the random search of Kozak (2006a), divergence between the two methods occurs in both the use of penalized objective functions and the use of simulated annealing to achieve a final allocation. Instead, Kozak (2006a) considers only a sequence of allocations that are monotone increasing in objective function value. However, both Kozak (2006b) and the proposed method do produce integer based allocations, unlike Ballin and Barcaroli (2013) and Benedetti et al. (2008) that use Bethel (1986) allocation. Where the final allocation from Bethel (1986) is rounded up to the nearest integer. This rounding is generally nonoptimal, particularly when stratum sample sizes are small.

The performance of simulated annealing is governed by three primary factors (Henderson et al. 2003): choice of cooling schedule, the shape of the objective function surface, and the application or domain. The shape of the cooling schedule governs the speed of convergence and the rate of accepting nonoptimal states. The literature on the choice of cooling schedule is largely based on heuristics balancing run-time and acceptable conditions (Romeo and Sangiovanni-Vincentelli 1991). Strenski and Kirkpatrick (1991) provide some theoretical results for extremely small populations with respect to optimal cooling schedules. The results of this analysis suggest that the linear, used here, or geometric cooling schedules tend to out perform more complex methods. Beyond the choice of cooling function, the only other controllable aspect of the optimization is the objective function. Objective functions that have shallow local minima tend to yield shorter run-times and better results due to the ease of escaping from nonoptimal states. In the application to optimal stratification and allocation, the depth of the local minima is a function of the shape of the function. Successful exchanges of PSUs with large values relative to the other PSUs, such as large operations in highly skewed populations, may cause changes in allocation. This can create local minima that are difficult to escape, causing nonoptimal solutions (Hajek 1988).

Implementation of a high-dimensional simulated annealing algorithm for nontrivial cases, however, is not so straightforward. The primary issues are:

- The computational cost of calculating the objective function;
- The likelihood of selecting a move that would reduce the objective function (improving convergence speed).

Calculating the objective function directly is computationally challenging. An alternative is to retain the $S^2_{h,j}$ component and to update a temporary candidate for $S^{2,*}_{h,j}$. Updates are performed through the numerically stable online sample variance calculation algorithm given by (Knuth 1997, 232), where online methods provide an iterative method to update the variance as opposed to recalculation of the variance. Periodic recalculation of the variances is provided to preserve numeric precision over a large number of updates, this recalculation occurs every 1,000,000 accepted exchanges. The numeric stability of the online algorithm was tested on simulated data using 1,000,000 iterations of the algorithm. This result was compared against $S^2_{h,j}$ calculated directly from the current stratification. The difference between these methods was less than e^{-12}. This error rate should be acceptable for most applications. However, care should be taken for exceptionally large populations. The application of a stable online method is also used by Benedetti et al. (2008) and Ballin and Barcaroli (2013) without periodic recalculation of variances.

3. Simulated Examples

The effectiveness of the multivariate joint stratification and allocation method using Simulated Annealing (SA) proposed in this article is compared to other methods in the literature through two examples: A univariate example comparing SA to the univariate joint stratification and allocation method of Lavallee-Hidiroglou (LH) using the R package *stratification* (Baillargeon and Rivest 2011), and a multivariate example comparing SA to the multivariate joint allocation and stratification using a genetic algorithm (GA) provided in theR package *SamplingStrata* (Barcaroli et al. 2014). The tree based method in Benedetti et al. (2008) was not considered due to lack of an available software package. Each of the two examples model PSU response through one of two linear models.

A homoscedastic linear model, and a heteroscedastic linear model where the variance is proportional to administrative data.

The univariate comparison between SA and LH has four goals: (1) Provide a diagnostic to ensure that SA has similar performance to known univariate optimal methods, as in Barcaroli et al. (2014). (2) Provide empirical results with respect to penalty weight selection. (3) Compare results using design variance and anticipated variance. (4) Show improvements that can be obtained over univariate methods with the presence of correlation. Similarly, the multivariate comparison between SA an GA has a two goals: (1) Compare SA and GA with respect to statistical efficiency. (2) As in the univariate example, to illustrate the advantage of using anticipated variance as a criterion for optimization.

In both examples, a population of 5,000 PSUs is simulated from two sets of linear models, a homoscedastic model and a heteroscedastic model. The homoscedastic model provides a simple case where the variance of the response is independent of the administrative data; the heteroscedastic model provides a more complex case where the variance of the response is proportional to the administrative data. This later case is common in many establishment surveys. In each set of linear models, each PSU, indexed by i has a vector valued response $\mathbf{y}_i = \{y_{i,1}, y_{i,2}, y_{i,3}, y_{i,4}, y_{i,5}\}$, and each element of the PSU is correlated with a vector $\mathbf{z}_i = \{z_{i,1}, z_{i,2}\}$ by a varying amount. Both elements of $\mathbf{z}_1$ and $\mathbf{z}_2$ are generated from a Chi-squared distribution with three degrees of freedom and scaled by 50 to produce values largely in the range of 0 to 1,000. This distribution is chosen to mimic the response of skewed populations common in establishment surveys. The relationship between a response y_i and z_i is determined by a linear model. The linear component of these models $x_i = z_i\beta_j \; j \in \{1, \ldots, 5\}$ will be used as an administrative variable for both examples and models, where β_j is assumed to be known. Examples and model will instead vary on the objective functions used for SA and comparisons to alternative methods.

In the homoscedastic linear model, the response vector for the i^{th} observation, y_i, is generated from linear models of the form,

- $y_{i,1} = z_i\beta_1 + \epsilon_{i,1}\bar{z}_1^{\gamma}$ with $\mathbb{E}[\epsilon_{i,1}^2] = \sigma_1^2$,
- $y_{i,p} = z_i\beta_p + \epsilon_{i,p}||(\bar{z}_1, \bar{z}_2)||_2^{\gamma}$ with $\mathbb{E}[\epsilon_{i,p}^2] = \sigma_p^2$ for $p \in \{2, 3, 4\}$, and
- $y_{i,5} = z_i\beta_5 + \epsilon_{i,5}\bar{z}_5\gamma$ with $\mathbb{E}[\epsilon_{i,5}^2] = \sigma_5^2$.

Each model error $\epsilon_i = \{\epsilon_{i,1}, \epsilon_{i,2}, \epsilon_{i,3}, \epsilon_{i,4}, \epsilon_{i,5}\}$ is distributed $\mathcal{N}(0, \Sigma)$ where Σ is a diagonal matrix, and $\bar{z}_j$ is the mean of the variable z_j over all PSUs. To avoid cases where the response is less than zero, all negative values are truncated to zero. To provide similar magnitude of the variances in the heteroscedastic model the variance is scaled by the mean of norm of means from the vectors in $\mathbf{Z}$ raised to $\gamma = 0.75$.

The heteroscedastic model is similar in form and follows from the generalization of the linear model found in Brewer (1963). Specifically,

- $y_{i,1} = z_i\beta_1 + \epsilon_{i,1}z_{i,1}^{\gamma}$ with $\mathbb{E}[\epsilon_{i,1}^2] = \sigma_1^2$,
- $y_{i,p} = z_i\beta_p + \epsilon_{i,p}||z_i||_2^{\gamma}$ with $\mathbb{E}[\epsilon_{i,p}^2] = \sigma_p^2$ for $p \in \{2, 3, 4\}$, and
- $y_{i,5} = z_i\beta_5 + \epsilon_5 z_{i,2}^{\gamma}$ with $\mathbb{E}[\epsilon_{i,5}^2] = \sigma_5^2$.

Each model error ϵ_i is distributed $\mathcal{N}(0, \Sigma)$ where Σ is the identity matrix and γ is also set to 0.75. This value of γ was described by Kott et al. (2000) to be appropriate for many establishment surveys.

In both the homoscedastic and heteroscedastic models β is set to allow for different levels of correlation: β_1 and β_5 are fixed to the vectors $(1, 0)$ and $(0, 1)$ respectfully, and β_p is set at three different levels $(0.75, 0.25)$, $(0.5, 0.5)$, and $(0.25, 0.75)$ respectfully for $p \in \{2, 3, 4\}$. ACV constraints for y in both models are set at 0.04 for all variables. All calculations for SA use ϕ set to one for all variables, 50,000 PSU exchanges with ten iterations of optimal allocation per exchange, and cooling schedule $(l+1)^{-1}$ where l is the current iteration. For simplicity, only soft constraints are used and are specified in each example. Details on the objective functions used in GA and SA, as well as their associated administrative functions are provided in Table 1.

3.1. Univariate Example

In this example, we consider a univariate approach to multivariate response using the LH joint allocation and stratification method, as described in Baillargeon and Rivest (2009), and compare it to the multivariate approach of SA. In this example, LH will be used as a diagnostic measure to ensure that SA can obtain an optimal result in a simple setting, identify performance characteristics using targeted penalty weights and no penalty weights, identify the importance of using anticipated variance, and finally to see if SA can further improve the results of the univariate allocation by finding an allocation that meets the univariate CV target while simultaneously improving the CVs of nontargeted administrative variables. Results of this example can be found in Tables 2 and 3.

Before examining the results of this comparison it is useful to consider some properties of LH relative to SA. LH can use either design variances or anticipated variance to determine the partitioning of the population into a fixed number of strata either with CV constraints without a fixed sample size, or with a fixed sample size and without CV constraints. In this example, we will be using the former case to set an initial sample size for SA.

To form the strata with fixed CV targets, LH uses an iterative algorithm. Since LH only works on univariate administrative data, strata can be identified as a set of disjoint intervals of the real line. Two approaches to find the boundaries of these intervals are found in Baillargeon and Rivest (2011): a model based approach used in the original Lavallée and Hidiroglou (1988) paper, and a random search method proposed in Kozak (2004). Due to the excellent performance characteristics without model assumptions, the random search method was chosen for this example.

Table 1. Objective functions used in the univariate and multivariate examples.

Method	Second moment	Objective function	Administrative data
LH	$\mathbb{E}_m\left[S^2_{h,1}\right]$	Neyman	$\mathbf{x} = \mathbf{Z}\beta_1$
GA	$S^2_{h,j}$	Bethel	$\mathbf{X} = \mathbf{Z}\beta$
SA	$S^2_{h,j}$	Equation (3)	$\mathbf{X} = \mathbf{Z}\beta$
SA (ACV)	$\mathbb{E}_m\left[S^2_{h,j}\right]$	Equation (3)	$\mathbf{X} = \mathbf{Z}\beta$

The random search approach starts with a set of initial intervals. Given these intervals, interval boundaries are perturbed at each iteration. This perturbation is performed by sorting the PSUs by administrative variable values and either moving the boundary forward or backward a random number of places in this sorted order. If the perturbed set of strata is more-optimal than the prior set per an objective function, then the perturbed set of strata is taken; otherwise, the perturbed set is discarded. The algorithm terminates when there is no change in stratification.

The objective function used in LH is simply Neyman allocation with a CV constraint:

$$n_h = n \frac{\sqrt{N_h^2 S_h^2}}{\sum_{k=1}^{H} \sqrt{N_k^2 S_k^2}} \quad \text{s.t. } f\left(\mathbf{X}|\mathbf{I}, \eta\right) \leq c \tag{10}$$

In this example, $c = c_1$ is the univariate target, $S_h^2 = S_{h,1}^2$, and $f\left(\mathbf{x}|\mathbf{I}, \eta\right) =$ is the coefficient of variance of the administrative data $\mathbf{x} = \mathbf{z}_1 \beta_1$. S_h^2 can be substituted with the expected value of S_h^2 with an assumed model to provide an approximation to optimization with the anticipated variance

$$\mathbb{E}_m\left[S_h^2\right] = \frac{1}{N_h - 1}\left(\sum_{i \in U_h} v_i^2 \sigma^2 + \sum_{i \in U_h} (z_{i,1}\beta_1 - \bar{z}_{1,h}\beta_1)^2\right) \tag{11}$$

where $v_i = \bar{z}_1$ in the homoscedastic case and z_i in the heteroscedastic case. All calculations are performed using the R package *stratification* (see Baillargeon and Rivest 2011).

Results for SA are calculated with four possible combinations of objective functions and penalty weighting. The two configurations of the objective function are identified in Tables 2 and 3 as "SA" and "SA (ACV)" with objective function specification identified in Table 1. The first configuration "SA" uses design variance in the objective function (3) with the administrative variables $\mathbf{X} = \mathbf{Z}\beta$. The second configuration "SA (ACV)" uses anticipated variance in the objective function (3) using the homoscedastic or heteroscedastic model. Targeted penalty weighting towards y_1 is used to provide a comparable result to LH, while the nontargeted weighting is used to compare the changes in attained ACV due to targeting a specific variable.

Results in ACV for the homoscedastic and heteroscedastic models are respectively provided in Tables 2 and 3. In both cases, LH chose six total strata and total sample size $n = 23$ in the homoscedastic case and $n = 65$ in the heteroscedastic case. All results are provided in anticipated coefficients of variation.

In addressing the goals of this example, SA (ACV) provides similar results to LH in that both methods attain the desired target CV for the same sample size in both homoscedastic and heteroscedastic cases. However, LH was able to reduce the ACV of y_1 further below the target of SA (ACV). The benefit of this further reduction is debatable, particularly if there is a benefit from reducing the ACV of other characteristics of interest in a survey. When there are other characteristics of interest, as in y_2 through y_5, SA (ACV) clearly outperforms LH.

With respect to penalty weight selection, the targeted weighting clearly had an effect on the result. This effect can be seen through the reduction of SA (ACV) and SA in the

Table 2. Anticipated CVs for simulated population generated from a homoscedastic linear model (univariate example).

Method Target	λ_1	λ_2	λ_3	λ_4	λ_5	ACV(y_1) 0.04	ACV(y_2) 0.04	ACV(y_3) 0.04	ACV(y_4) 0.04	ACV(y_5) 0.04	$\lVert \text{ACV}(y)\rVert_2$
						Targeted weighting on y_1					
LH						0.0391	0.0540	0.0958	0.1406	0.1856	0.2604
SA	100	0	0	0	0	0.0411	0.0487	0.0873	0.1310	0.1754	0.2442
SA (ACV)	100	0	0	0	0	0.0400	0.0492	0.0865	0.1285	0.1713	0.2395
						No targeted weighting					
SA	0	0	0	0	0	0.0820	0.0579	0.0451	0.0507	0.0698	0.1399
SA (ACV)	0	0	0	0	0	0.0693	0.0491	0.0446	0.0580	0.0802	0.1378

Table 3. Anticipated CVs for simulated population generated from a heteroscedastic linear model (univariate example).

Method	λ_1	λ_2	λ_3	λ_4	λ_5	ACV(y_1)	ACV(y_2)	ACV(y_3)	ACV(y_4)	ACV(y_5)	$\|ACV(y)\|_2$
Target						0.04	0.04	0.04	0.04	0.04	
						Targeted weighting on y_1					
LH						0.0397	0.0647	0.0820	0.1050	0.1273	0.1993
SA	100	0	0	0	0	0.0542	0.0576	0.0564	0.0605	0.0609	0.1297
SA (ACV)	100	0	0	0	0	0.0400	0.0619	0.0778	0.0996	0.1204	0.1895
						No targeted weighting					
SA	0	0	0	0	0	0.0553	0.0577	0.0559	0.0594	0.0593	0.1288
SA (ACV)	0	0	0	0	0	0.0553	0.0574	0.0555	0.0591	0.0594	0.1284

direction of y_1. The no target weighting case shows the degree of change that occurs by targeting a particular variable. As would be assumed, the amount of change in attained ACVs is associated with the degree of correlation with the targeted variable. Variables with high positive correlation with y_1 have ACVs that increase the least when y_1 is targeted (e.g., y_2); variables such as y_5 tend to have their ACVs increase the most.

The heteroscedastic case is important, in that strata containing larger values of the administrative variable will have higher model variance. Therefore, the impact of ignoring the model variance is more extreme than only using the design variance. This can clearly be seen in the results of SA compared to those of LH and SA (ACV). The later two results that use anticipated variance tend to consistently meet targets in both cases, while SA using just the design variance almost hits the target using the homoscedastic model, but considerably misses the target in the heteroscedastic case.

3.2. Multivariate Example

In the multivariate example, we reuse the prior population in the univariate example but apply the joint optimal allocation and stratification methods GA and SA. The goal of this example is to compare the statistical efficiency of the resulting survey designs using GA and SA, as well as revisit the topic of using ACVs in optimization.

Because GA as presented in Ballin and Barcaroli (2013) does not support targeting ACVs, the algorithm uses $X = \boldsymbol{Z}\beta$ as known administrative data. As in the univariate example, the results identified as SA are also fit in the same manner; SA (ACV) uses anticipated variance in the objective function. Results are found in Tables 4 and 5. To illustrate the importance of specifying a design using ACVs, the stratifications attained for GA and SA are presented both using attained CVs from the administrative data and ACVs. It is important to note that optimization using design variances are identical in the homogenous and heterogenous case, as they ignore the model variance. Therefore attained CVs are only listed in Table 4.

To provide comparable results between GA and SA the optimal sample size and number of strata from GA are used for the SA based optimizations. In this example, the optimal sample size using GA is 193 and the total number of strata is five.

Individual PSUs are used for atomic strata for GA, and minor performance tuning was performed. Tuning proved problematic due to the long run-time of GA, averaging two hours and 35 minutes per run. Run-times of SA, on the other hand, averaged 25 seconds for both the CV and ACV optimizations.

In both the homoscedastic and heteroscedastic cases, GA was less efficient than SA when only considering CV targets, but produced more robust stratifications. This robustness appears to be a result of attaining a local minima, instead of a feature of the GA algorithm. As in the univariate example, SA (ACV) provided uniformly better results with respect to attained ACV than SA and GA. GA did do reasonably well in the homoscedastic case, meeting all CV targets. With ACV evaluation criteria, GA met one ACV target for both the homoscedastic and heteroscedastic cases, but did not suffer from larger departures from the target as in the case of SA.

The result for SA demonstrate the importance of using ACVs in an objective function. In this result, there was considerable reduction in CV. However, this reduction in CV was

Table 4. CVs and ACVs for simulated population using the homoscedastic model. The first set of results ignore anticipated variance in the objective function and are CV targets are reported, the resulting ACV of the first set of results and SA using anticipated variance in the objective function are provided in the second set of results.

Method	λ_1	λ_2	λ_3	λ_4	λ_5	CV(y_1)	CV(y_2)	CV(y_3)	CV(y_4)	CV(y_5)	$\|CV(y)\|_2$
Target						0.04	0.04	0.04	0.04	0.04	
GA						0.0397	0.0308	0.0274	0.0311	0.0397	0.0763
SA	0	0	0	0	0	0.0274	0.0200	0.0171	0.0202	0.0272	0.0509
Method	λ_1	λ_2	λ_3	λ_4	λ_5	ACV(y_1)	ACV(y_2)	ACV(y_3)	ACV(y_4)	ACV(y_5)	$\|ACV(y)\|_2$
GA						0.0419	0.0418	0.0348	0.0443	0.0420	0.0919
SA	0	0	0	0	0	0.0630	0.0493	0.0439	0.0490	0.0616	0.1205
SA (ACV)	0	0	0	0	0	0.0238	0.0242	0.0235	0.0252	0.0255	0.0547

Table 5. ACVs for simulated population using the heteroscedastic model.

Method	λ_1	λ_2	λ_3	λ_4	λ_5	ACV(y_1)	ACV(y_2)	ACV(y_3)	ACV(y_4)	ACV(y_5)	$\|ACV(y)\|_2$
GA						0.0451	0.0423	0.0397	0.0422	0.0451	0.0960
SA	0	0	0	0	0	0.0673	0.0590	0.0544	0.0584	0.0660	0.1369
SA (ACV)	0	0	0	0	0	0.0344	0.0346	0.0328	0.0344	0.0342	0.0762

done at the expense of ACV in both the homoscedastic and heteroscedastic cases, where the attained ACV was over double the attained CV. This is an over fitting and model misspecification problem, where assuming the control data as the response yielded a nondesirable outcome. In practice, care should be taken to test multiple potential response models on a potential stratification.

4. June Agricultural Survey

Application of the simulated annealing based multivariate optimal stratification and allocation method is examined in the proposed redesign of the United States Department of Agriculture (USDA) National Agricultural Statistics Service's (NASS) June Agricultural Survey (JAS). In this section, JAS and the proposed redesign are introduced along with a discussion of administrative variables and implementation details. A proxy of JAS provides a comparable design using the same administrative data and PSUs. This proxy is then compared to the simulated annealing based stratification and allocation method in this article, followed by a discussion of the results.

JAS is a two-stage annual area survey of the continental United States, producing estimates of acreage devoted to various agricultural land uses and other spatially associated estimates (Davies 2009). JAS is administered by NASS, with data collected by The National Association of State Departments of Agriculture (NASDA) employees. The first stage of JAS is a stratified simple random sample design with replacement, where strata are formed by grouping PSUs based on the percentage of cultivated acres within each PSU. When needed, specialty strata are used to target rare commodities or demographic groups. Each PSU in the first stage is a contiguous one-to-eight square mile region of land manually delineated along permanent geographic features such as roads. Cultivated acreage for each PSU is calculated using NASS's Cropland Data Layer (CDL), a remotely-sensed administrative data set of land-cover and land-use (Boryan et al. 2011). PSUs are sampled using systematic sampling of a spatial index, allowing for a spatially well distributed sample. In the second stage, selected PSUs are partitioned into smaller areas of land known as segments, serving as Secondary Sampling Units (SSU)s. Segments are manually formed by the delineation of PSUs into approximately one-square-mile contiguous regions of high agricultural production; larger segments can be drawn in areas with no-to-low agricultural activity. A single segment is selected randomly from each PSU, and all land within the selected segment is fully enumerated. Nonresponse is handled through observation, remote sensing, or subject matter experts. Allocation in JAS is performed by Bethel (1986) using historic data with equal cost per PSU.

To lower design costs and to allow for estimation of year-to-year change JAS is replicated. A set of replicates are created every five years and all of these replicates are rotated into the sample one year at a time. Each replicate is collected for approximately five years, and then rotated out of the sample.

National level target CVs are chosen by NASS to ensure quality estimates. The CV targets are predictive in nature, as they are set for the estimates, not the administrative data. Target CVs are considered satisfied if on-average the attained CVs are less than the target CVs. This comparison occurs at the level of precision of the target CV. Target CVs are typically met each year, but occasionally some targets cannot be attained for a given sample size.

Iowa State in cooperation with NASS, considered an update of the current JAS design (Zimmer et al. 2013). In this proposed redesign, the two-stage design is replaced by a single-stage design with optimal stratification based on areal units of one-square-mile in size. These PSUs, known as sections, are part of a permanent area frame based on the Public Land Survey System (PLSS). This permanent frame greatly reduces survey cost, as the current JAS requires the labor intensive manual delineation of PSUs and SSUs. Stratification of the proposed redesign's area frame is based on the optimal joint allocation and stratification algorithm described in this article. Like JAS, this design is calculated using equal PSU costs. Spatial balance is attained by using the local pivotal method in Grafström et al. (2012) and implemented using the *BalancedSampling* R package (Grafström and Lisic 2016). Unfortunately, the current implementation of the simulated annealing procedure only supports the Sen-Yates-Grundy variance estimates (Sen 1953), over estimating the variance when using locally balanced sampling; instead, simple random sampling with replacement is used as an upper bound for the variance with an assumptions of spatial clustering (Grafström et al. 2012).

As in the current JAS design, remotely sensed CDL data is used as administrative data. The CDL has accuracy above 90% for corn and soybeans as well as above 80% for winter and spring wheat for all years used in this research (2008–2015). This makes the CDL a fairly useful tool in evaluating surveys, unfortunately, linear models of section acreage are not particularly good at predicting future land use. This is due to the agricultural practice of crop rotations, where individual fields within a section tend to follow crop specific sequences to maximize yield, mitigate pests, and reduce erosion. Instead of directly modeling these crop sequences, an assumption is made that sequences observed within a period of time are likely to re-occur within a future window of time. Using this assumption, we use the prior four years to predict the next four years. This is similar to the current JAS practice where a single stratification is used for multiple years.

Due to lack of correlation between nonacreage based estimators, such as number of farms and livestock, with available administrative data, the joint optimal allocation and stratification method is only used for acreage based estimates. To ensure that quality constraints are met for nonacreage estimators, prior JAS response is used to calculate historical variances. These historical variances are used to ensure that the total sample size is of sufficient size to meet the imposed quality constraints. To check for any potential deleterious effects caused by unforeseen relationships with the nonacreage responses and the optimal stratification and allocation, quality constraints are checked by post stratifying geo-referenced, but not complete, historical data by the new design.

In this article, multivariate stratification and allocation are only performed on the PSUs of the redesign (sections). The original JAS design is proxied by a set of univariate bounds based on cultivated acres (Table 6). A proxy is used, instead of the original JAS design,

Table 6. Approximation of JAS stratification.

Stratum	Definition
11	75% or more cultivated land
20	50–74% cultivated land
30	15–49% cultivated land
40	1–14% cultivated land
50	0% cultivated land

due to the intractability of modeling the manual segmentation of secondary segmentation units. For the purpose of evaluation, only results for South Dakota were considered with the acreages of interest including corn, soybeans, winter wheat, spring wheat, and cultivated acres. South Dakota provides a reasonable use case for multivariate allocation with a large number of crop types and large frame of close to 80,000 sections.

National level target CVs from JAS cannot be applied to a single state; therefore, historical JAS CVs (2008–2011) for South Dakota are used as CV targets. Multiple years of land use are used to form strata to account for year-to-year land cover variability. In this example 2008–2011 are used to form the stratification, and 2012–2015 are used to evaluate future land use. The CDL variables used for stratification are cultivated, corn, soybeans, winter wheat, and spring wheat acreages. These variables were used for all segments in the population, and optimization was applied to each variable and year combination from 2008–2011 simultaneously treating each combination as a separate variable. The algorithm is run for 5,000,000 iterations with five sample size optimization steps per allocation. Initial stratification is performed by K-means to accelerate convergence. The penalty weights φ and λ were set to 1 and 1,000 respectively for each combination of year and land cover. The total run-time using this parameterization using the proposed method is 30 minutes. In both the univariate and the multivariate cases all 80,000 segments were assigned to five strata. Allocation for the JAS strata is performed by the multivariate allocation method described in (Bethel 1986). The total sample size from this allocation is used for the simulated annealing based approach. The highly correlated administrative data is used as a proxy for the true response. The resulting CVs from both methods provided in Table 8 for the multivariate method and Table 7 for the method approximating the current JAS.

Table 7. CVs for specified variables based on an approximate JAS stratification and allocation of South Dakota.

	Cultivated	Corn	Soybeans	Winter wheat	Spring wheat
Target	0.01	0.05	0.05	0.19	0.16
2008	0.0162	0.0470	0.0524	0.1078	0.1129
2009	0.0153	0.0453	0.0510	0.1148	0.1153
2010	0.0168	0.0455	0.0484	0.1239	0.1151
2011	0.0137	0.0410	0.0483	0.1139	0.1275
2012	0.0147	0.0395	0.0477	0.1378	0.1389
2013	0.0158	0.0409	0.0500	0.1546	0.1404
2014	0.0167	0.0428	0.0481	0.1514	0.1327
2015	0.0168	0.0457	0.0488	0.1606	0.1310

Table 8. CVs for specified variables based on multivariate joint stratification and allocation of South Dakota. Changes relative to the univariate method (Table 7) in parenthesis.

	Cultivated	Corn	Soybeans	Winter wheat	Spring wheat
Target	0.01	0.05	0.05	0.19	0.16
2008	0.0137 (−15.43%)	0.0496 (5.53%)	0.0534 (1.91%)	0.1138 (5.57%)	0.1200 (6.29%)
2009	0.0118 (−22.88%)	0.0475 (4.86%)	0.0508 (−0.39%)	0.1197 (4.27%)	0.1219 (5.72%)
2010	0.0138 (−17.86%)	0.0465 (2.20%)	0.0475 (−1.86%)	0.1293 (4.36%)	0.1207 (4.87%)
2011	0.0119 (−13.14%)	0.0429 (4.63%)	0.0482 (−0.21%)	0.1178 (3.42%)	0.1338 (4.94%)
2012	0.0125 (−14.97%)	0.0403 (2.03%)	0.0492 (3.14%)	0.1449 (5.15%)	0.1452 (4.54%)
2013	0.0137 (−13.29%)	0.0419 (2.44%)	0.0511 (2.20%)	0.1599 (3.43%)	0.1466 (4.42%)
2014	0.0144 (−13.77%)	0.0436 (1.87%)	0.0487 (1.25%)	0.1571 (3.76%)	0.1352 (1.88%)
2015	0.0146 (−13.09%)	0.0469 (2.63%)	0.0494 (1.23%)	0.1623 (1.06%)	0.1339 (2.21%)

The results of this empirical example showed a general improvement in the CVs for the multivariate method relative to the univariate method in areas where the CV targets were difficult to attain. Considering JAS rounding rules, this improvement allows the multivariate method to meet the target CVs of all crops both on average (2012–2015) and per-year. The rounded CV targets in the univariate case are generally met, with the exception of total cultivated acreage that was only met in 2012. However, the rounding rule tends to favor the multivariate method over stating its performance relative to the univariate method.

For both methods, the most difficult to attain target CV is total cultivated acreage, where the multivariate method averaged a -13.78% decrease in CV relative to the univariate method in the evaluation years (2012–2015). The univariate method has lower CVs for other crops within the evaluation years, but most of these CVs for other crops are well below the target CVs for both methods. Other results included indications of model misspecification in the multivariate method through the general increase in attained CVs for the predicted years.

5. Discussion

In this article, a method to construct optimal multivariate stratified designs for an arbitrary, but fixed, number of self-representing strata was presented. This method admits a combination of hard and soft constraints, where soft constraints are handled using a penalized objective function and hard constraints are handled through traditional nonlinear programming constraints. Optimization of the objective function is performed using simulated annealing, by moving individual PSUs between strata. Simultaneous allocation is provided by also considering changes in the allocation as part of the simulated annealing algorithm. Penalized weighting in the objective function allowed for flexibility in design specification, allowing for penalty weights to target specific commodities based on preference or correlation with administrative variables. The use of anticipated variance in the objective function was shown to account for uncertainty in the relationship between administrative data and targeted estimates, and opens the door to modeling nonsampling error. Applications to both a simulated population and the proposed redesign of JAS were provided. Important issues with the proposed method, beyond the scope of this research, include investigations of nonsampling error, handling poor quality administrative data, model misspecification when using a model-assisted objective function, and improvements to computational speed. Future application specific research with respect to the JAS redesign, and potentially other spatial surveys, includes improvements to the objective function to reflect better the variance using spatially balanced sampling methods and the development of better prediction models for agricultural land use for individual PSUs.

In application, the proposed method was shown to be computationally tractable for reasonably large populations and more flexible than existing methods through the use of soft constraints and the use of anticipated variance in the objective function. The two examples are chosen to show utility of the method in existing establishment and area surveys. Application to more complex designs has not been considered in this research, but the model-assisted objective function could be extended to account for subsampling and other traits of complex designs. Application to more complicated designs requiring optimization of multiple samples may also be possible for paneled or split-questionnaire designs as in Ioannidis et al. (2016).

In a univariate example, both homoscedastic and heteroscedastic populations are investigated to describe establishment surveys with different dispersion patterns. In both the anticipated moments-based approach, SA performed better than the existing anticipated moments-based LH approach. Provided that the relationship between the administrative variables and the target variables are reasonably well known, the proposed method should provide improvements over LH in multivariate scenarios.

In the multivariate example, the proposed method greatly out-performs the genetic algorithm approach in optimization. Improvements to the genetic algorithm-based approach, such as the adoption of anticipated moments in the objective function and finer tuning of parameters, may provide parity between the results. Furthermore, model misspecification, and prospective use of related goodness-of-fit diagnostics should be explored for both methods. However, the proposed method may be more applicable for larger populations due to the long run-time of the genetic algorithm relative to the proposed method.

In the JAS redesign, the proposed method met or exceeded the attained CVs of the JAS approximation under JAS rounding rules, and in general had lower target CVs for hard-to-attain targets, providing a strong argument for the use of multivariate designs onthis survey. This method was also shown to be computationally feasible for population sizes of 80,000 with a run-time of thirty minutes: computation for larger populations should be possible at the expense of longer run-times. The computational speed and stability of the proposed method improves on existing methods through the use of online-variance calculation with periodic recalculation of variances. The use of prior information, as in the case of JAS may not be possible for other area surveys, nor advisable if the underlying stochastic process changes over time.

The applicability to other area surveys is largely dependent on the variance estimation method employed, number of PSUs, availability of administrative data, and the ability to model individual PSUs. The current objective function only considers SRS with or without replacement, not accounting for increases in efficiency that could be attained using more advanced sampling methods. The application of the proposed method to a population of 280,000 PSUs has been explored by Lisic et al. (2015), but general applicability to surveys with considerably larger populations has not been explored. Modeling individual PSUs is not needed to apply the proposed method to a survey using administrative data based quality constraints. However, if quality constraints are placed on the estimates either correlation-based weights or a model should be introduced. The correlation-based weights may be useful under linear relationships. However, their use for more complicated relationships is uncertain. The applicability in the case of a model would depend on how well the model describes the uncertainty in the relationship between the administrative data and the response.

Although not explored in this research, this anticipated moment approach allows incorporation of estimates of nonsampling errors such as assumed nonresponse in the response variable. This feature can already be found in (Baillargeon and Rivest 2014) for univariate optimal allocation and stratification. Correlation-based penalty weighting can also incorporate nonsampling error within the correlation function. However, this approach may be limited to cases when the relationships between the administrative variables and the survey response is fairly linear.

A similar problem addressed within the context of the JAS redesign, but not in general is the handling of poor quality administrative variables. This can occur when only a subset of the frame has complete records, such as using prior survey data or incomplete databases. Provided that an accurate measure of uncertainty can be obtained for each observation, the anticipated variance framework can provide an optimal allocation and stratification. However, this question is beyond the scope of this research in this article.

Another interesting, but unexplored, area of research within this article is the importance of model specification for the anticipated moment approach. In the simulated populations, it is assumed that the model is known. In application, this is an unlikely case. Therefore, future analysis of the effect of model misspecification, and prospective use of related goodness-of-fit diagnostics should be explored more thoroughly.

Further acceleration of the proposed method may extend the applicability to larger populations. Two ways to improve the computational speed of the presented method for larger populations is through selecting PSUs near stratum boundaries with greater probability and exchanging multiple PSUs. These PSUs are more likely to be accepted during exchanges, allowing for faster convergence of the algorithm. The current implementation already supports the use of static weights to increase the probability that a particular PSU is selected. However, finding the ideal properties of these weights has not been considered yet. For these large population sizes moving individual PSUs between strata may result in infeasible run-time. One solution to this problem is by exchanging clusters of PSUs or partitioning strata by identifying useful hyperplanes in the space of administrative variables.

6. References

Baillargeon, S. and L.-P. Rivest. 2009. "A General Algorithm for Univariate Stratification." *International Statistical Review* 77(3): 331–344. Doi: http://dx.doi.org/10.1111/j.1751-5823.2009.00093.x.

Baillargeon, S. and L.-P. Rivest. 2011. "The Construction of Stratified Designs in R with the Package Stratification." *Survey Methodology* 37(1): 53–65.

Baillargeon, S. and L.-P. Rivest. 2014. "Stratification: Univariate Stratification of Survey Populations." Available at: https://CRAN.R-project.org/package=stratification, r package version 2.2-5 (accessed January 18, 2017).

Ballin, M. and G. Barcaroli. 2013. "Joint Determination of Optimal Stratification and Sample Allocation Using Genetic Algorithm." *Survey Methodology* 39(2): 369–393.

Barcaroli, G. 2014. "Sampling Strata: An R Package for the Optimization of Stratified Sampling." *Journal of Statistical Software* 61(4): 1–24. Available at: https://www.jstatsoft.org/article/view/v061i04/v61i04.pdf (accessed January 20, 2018).

Benedetti, R., M. Bee, and G. Espa. 2010. "A Framework for Cut-Off Sampling in Business Survey Design." *Journal of Official Statistics* 26(4): 651–671.

Benedetti, R., G. Espa, and G. Lafratta. 2008. "A Tree-Based Approach to Forming Strata in Multipurpose Business Surveys." *Survey Methodology* (34): 195–203.

Benedetti, R. and F. Piersimoni. 2012. "Multivariate Boundaries of a Self Representing Stratum of Large Units in Agricultural Survey Design." *Survey Research Methods* 6: 125–135. Doi: http://dx.doi.org/10.18148/srm/2012.v6i3.5127.

Bethel, J.W. 1986. *An Optimum Allocation Algorithm for Multivariate Surveys*. Technical report, United States Department of Agriculture, Statistical Reporting Service. Available at: https://naldc.nal.usda.gov/naldc/download.xhtml?id=43260&content=PDF (accessed January 20, 2017).

Boryan, C., Z. Yang, R. Mueller, and M. Craig. 2011. "Monitoring US Agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program." *Geocarto International* 26: 341–358. Doi: http://dx.doi.org/10.1080/10106049.2011.562309.

Brewer, K.R.W. 1963. "Ratio Estimation and Finite Populations: Some Results Deducible from the Assumption of an Underlying Stochastic Process." *Australian Journal of Statistics* 5(3): 93–105. Doi: http://dx.doi.org/10.1111/j.1467-842X.1963.tb00288.x.

Dalenius, T. and J.L.J. Hodges. 1959. "Minimum Variance Stratification." *Journal of the American Statistical Association* 54: 88–101.

Davies, C. 2009. *Area Frame Design for Agricultural Surveys*. Technical report, United States Department of Agriculture, National Agricultural Statistics Service. Available at: https://www.nass.usda.gov/Publications/Methodology_and_Data_Quality/Advanced_Topics/AREA%20FRAME%20DESIGN.pdf (accessed January 20, 2018).

Day, C.D. 2009. "Evolutionary Algorithms for Optimal Sample Design." A paper presented at the 2009 Federal Committee on Statistical Methodology Conference, November 4, 2009. Washington DC: Federal Committee on Statistical Methodology. Available at: https://fcsm.sites.usa.gov/files/2014/05/2009FCSM_Day_XI-A.pdf (accessed January 20, 2018).

Freuder, E.C. and R.J. Wallace. 1992. "Partial Constraint Satisfaction." *Artificial Intelligence* 58: 21–70. Doi: https://doi.org/10.1016/0004-3702(92)90004-H.

Gentle, J. and C. Perry. 2000. "Optimal Stratification of Area Frames." In Proceedings of the second international conference on establishment surveys - II, June 17–21, 2000. 1354–1382. Buffalo, NY: American Statistical Association. Available at: https://ww2.amstat.org/meetings/ices/2000/proceedings/S04.pdf (accessed January 20, 2018).

Gonzalez, J.M. and J.L. Eltinge. 2010. "Optimal Survey Design: A Review." In Proceedings of the Survey Research Methods Section of the American Statistical Association. Joint Statistical Meeting, October 4, 2010. 4970–83. Vancouver, BC: American Statistical Association. Available at: https://www.bls.gov/osmr/pdf/st100270.pdf (accessed January 20, 2018).

Grafström, A. and J. Lisic. 2016. *Balanced Sampling: Balanced and Spatially Balanced Sampling*. Available at: https://CRAN.R-project.org/package=BalancedSampling, r package version 1.5.1. (accessed March 8, 2016).

Grafström, A., N.L. Lundström, and L. Schelin. 2012. "Spatially Balanced Sampling through the Pivotal Method." *Biometrics* 68: 514–520. Doi: http://dx.doi.org/10.1111/j.1541-0420.2011.01699.x.

Gunning, P., J. Horgan, and W. Yancey. 2004. "Geometric Stratification of Accounting Data." *Contaduría y Administración*. Available at: http://www.ejournal.unam.mx/rca/214/RCA21401.pdf (accessed January 20, 2018).

Hajek, B. 1988. "Cooling Schedules for Optimal Annealing." *Mathematics of Operations Research* 13(2): 311–329. Doi: https://doi.org/10.1287/moor.13.2.311.

Henderson, D., S.H. Jacobson, and A.W. Johnson. 2003. "The Theory and Practice of Simulated Annealing." In *Handbook of Metaheuristics*, edited by F. Glover and G.A. Kochenberger, 287–319. Boston, MA: Springer.

Horgan, J.M. 2006. "Stratification of Skewed Populations: A Review." *International Statistical Review* 74: 67–76. Doi: http://dx.doi.org/10.1111/j.1751-5823.2006.tb00161.x.

Ioannidis, E., T. Merkouris, L.-C. Zhang, M. Karlberg, M. Petrakos, F. Reis, and P. Stavropoulos. 2016. "On a Modular Approach to the Design of Integrated Social Surveys." *Journal of Official Statistics* 32(2): 259–286. Doi: https://doi.org/10.1515/jos-2016-0013.

Isaki, C.T. and W.A. Fuller. 1982. "Survey Design under the Regression Superpopulation Model." *Journal of the American Statistical Association* 77(377): 89–96. Doi: http://dx.doi.org/10.2307/2287773.

Knuth, D.E. 1997. *The Art of Computer Programming: Seminumerical Algorithms*, volume 2. Addison-Wesley, 3 edition.

Kott, P.S. and J.T. Bailey. 2000. "The Theory and Practice of Maximal Brewer Selection with Poisson PRN Sampling." In Proceedings of the second international conference on establishment surveys - II, June 17–21, 2000. 269–304. Buffalo, NY: American Statistical Association. Available at: https://ww2.amstat.org/meetings/ices/2000/proceedings/S04.pdf (accessed January 20, 2018).

Kozak, M. 2004. "Optimal Stratification Using Random Search Method in Agricultural Surveys." *Statistics in Transition* 6(5): 797–806. Available at: http://omega.sggw.waw.pl/~m.kozak/SIT2004_6_5.pdf (accessed November 16, 2017).

Kozak, M. 2006a. "Multivariate Sample Allocation: Application of Random Search Method." *Statistics in Transition* 7(4): 889–900. Available at: https://www.researchgate.net/profile/Marcin_Kozak/publication/242329930_MULTIVARIATE_SAMPLE_ALLOCATION_APPLICATION_OF_RANDOM_SEARCH_METHOD/links/02e7e51cc8ede254cc000000.pdf (accessed November 16, 2017).

Kozak, M. 2006b. "On Sample Allocation in Multivariate Surveys." *Communications in Statistics-Simulation and Computation* 35: 901–910. Doi: http://dx.doi.org/10.1080/03610910600880286.

Lavallée, P. and M. Hidiroglou. 1988. "On the Stratification of Skewed Populations." *Survey Methodology* 14: 33–43.

Lisic, J.J. 2016. "saAlloc: Stratification and Allocation of Sampling Units Using Simulated Annealing." Available at: https://github.com/jlisic/saAlloc, r package version 2.0 (accessed January 20, 2018).

Lisic, J.J., H. Sang, Z. Zhu, and S. Zimmer. 2015. "Optimal Stratification and Allocation for the June Area Survey." A paper presented at the 2015 Federal Committee on Statistical Methodology Conference, December 2, 2015. Washington, DC: Federal Committee on Statistical Methodology. Available at: https://fcsm.sites.usa.gov/files/2016/03/E2_Lisic_2015FCSM.pdf (accessed January 20, 2018).

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. 1953. "Equation of State Calculations by Fast Computing Machines." *The Journal of Chemical Physics* 21: 1087–1092. Doi: https://doi.org/10.1063/1.1699114.

R Core Team. 2015. "R: A Language and Environment for Statistical Computing." *R Foundation for Statistical Computing*, Vienna, Austria, Available at: https://www.R-project.org/ (accessed July 1, 2017).

Romeo, F. and A. Sangiovanni-Vincentelli. 1991. "A Theoretical Framework for Simulated Annealing." *Algorithmica* 6(1): 302–345. Doi: https://doi.org/10.1007/BF01759049.

Särndal, C.-E., B. Swensson, and J. Wretman. 1991. *Model Assisted Survey Sampling*. New York: Springer.

Sen, A.R. 1953. "On the Estimate of the Variance in Sampling with Varying Probabilities." *Journal of the Indian Society of Agricultural Statistics* 5(1194): 127.

Strenski, P.N. and S. Kirkpatrick. 1991. "Analysis of Finite Length Annealing Schedules." *Algorithmica* 6(1): 346–366. Doi: https://doi.org/10.1007/BF01759050.

Valliant, R., F. Hubbard, S. Lee, and C. Chang. 2014. "Efficient Use of Commercial Lists in US Household Sampling." *Journal of Survey Statistics and Methodology* 2(2): 182–209. Doi: http://dx.doi.org/10.1093/jssam/smu006.

Zimmer, S., J.-K. Kim, and S. Nusser. 2013. "Automatic Stratification for an Agricultural Area Frame Using Remote Sensing Data." In Proceedings of the 59th ISI World Statistics Congress, 25–30 August 2013, Hong Kong, 1952–1957 Available at: www.statistics.gov.hk/wsc/STS043-P5-S.pdf (accessed June 12, 2014).

Received May 2016
Revised July 2017
Accepted November 2017