Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology 41-1

Generalized framework for defining the optimal inclusion probabilities of one-stage sampling designs for multivariate and multi-domain surveys

by Piero Demetrio Falorsi and Paolo Righi

Release date: June 29, 2015

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at infostats@statcan.gc.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "About us" > "The agency" > "Providing services to Canadians."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for a specific reference period
...	not applicable
0	true zero or a value rounded to zero
0^s	value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p}	preliminary
^{r}	revised
x	suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E}	use with caution
F	too unreliable to be published
*	significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2015
Vol. 41, No. 1, pp. 215-236
Statistics Canada, Catalogue No. 12-001-X

Generalized framework for defining the optimal inclusion probabilities of one-stage sampling designs for multivariate and multi-domain surveys

Piero Demetrio Falorsi and Paolo Righi[1]

Abstract

This paper introduces a general framework for deriving the optimal inclusion probabilities for a variety of survey contexts in which disseminating survey estimates of pre-established accuracy for a multiplicity of both variables and domains of interest is required. The framework can define either standard stratified or incomplete stratified sampling designs. The optimal inclusion probabilities are obtained by minimizing costs through an algorithm that guarantees the bounding of sampling errors at the domains level, assuming that the domain membership variables are available in the sampling frame. The target variables are unknown, but can be predicted with suitable super-population models. The algorithm takes properly into account this model uncertainty. Some experiments based on real data show the empirical properties of the algorithm.

Key Words: Optimal Allocation; Multi-way stratification; Domain estimates; Balanced Sampling.

1 Introduction

Surveys conducted in the context official statistics commonly produce a large number of estimates relating to both different parameters of interest and highly detailed estimation domains. When the domain indicator variables are available for each sampling unit in the sampling frame, the survey sampling designer could attempt to select a sample in which the size for each domain is fixed. Thus, direct estimates can be obtained for each domain and sampling errors at the domain level would be controlled. We hereby present a *unified* and *general* framework for defining the *optimal inclusion probabilities* for *uni-stage sampling designs* when the domain membership variables are known at the design stage. This case may be the most recurrent scenario in establishment surveys and other survey contexts, such as agricultural surveys or social surveys if the domains are geographical (e.g., type of municipality, region, province, etc.). The growing development of data integration among administrative registers and survey frames may also increase the applicability of the approach presented herein in social surveys too. The proposal may be useful for planning an optimal second phase survey if, during the first phase, the domain membership variables have been collected.

The problem of defining optimal sampling designs has been addressed in some recent papers. Gonzalez and Eltinge (2010) present an interesting overview of the approaches for defining optimal sampling strategies. The optimization problem is usually dealt with in stratified sampling designs with a fixed sample size in each stratum. The optimal allocation in stratified samplings for a univariate population is well-known in sampling literature (Cochran 1977). In multivariate cases, where more than one characteristic is to be measured on each sampled unit, the optimal allocation for individual characteristics is of little practical use unless the various characteristics under study are highly correlated. This is because an allocation which is optimal for one characteristic is generally far from being optimal for others. The

1. Piero Demetrio Falorsi, FAO, Viale delle Terme di Caracalla, Roma. E-mail: piero.falorsi@fao.org; Paolo Righi, ISTAT Via C. Balbo 16, 00184 Roma. E-mail: parighi@istat.it.

multidimensionality of the problem leads to definition of a compromise allocation method (Khan, Mati and Ahsan 2010) with a loss of precision compared to the individual optimal allocations. Several authors have discussed various criteria for obtaining a feasible compromise allocation - see e.g., Kokan and Khan (1967), Chromy (1987), Bethel (1989), Falorsi and Righi (2008), Falorsi, Orsini and Righi (2006) and Choudhry, Rao and Hidiroglou (2012).

Recently, some papers have focused on finding optimal inclusion probabilities in balanced sampling (Tillé and Favre 2005; Chauvet, Bonnéry and Deville 2011), a general class of sampling designs that includes stratified sampling designs as special cases. In particular, Chauvet et al. (2011) propose the adoption of the fixed point algorithm for defining the optimal inclusion probabilities. Nevertheless, the above mentioned papers do not address the case in which the balancing variables depend on the inclusion probabilities and present only a partial solution to the problem related to the fact that the sampling variance is an **implicit** function of the inclusion probabilities. Choudhry et al. (2012) propose an optimal allocation algorithm for domain estimates in stratified sampling (if the estimation domains do not cut across the strata). Their algorithm represents a special case of the approach proposed herein. The methodological setting illustrated here is a substantial improvement with respect to the earlier version of the methodology described in Falorsi and Righi (2008) which only accounted for the case in which the values of the variables of interest were known and the measure of accuracy was expressed by the design variance; furthermore, the previous version did not consider the fact that the design variance, bounded in the optimization problem, is an implicit function of the inclusion probabilities. This paper studies the more realistic case in which the variables of interest are not known and must be estimated. Moreover, it explicitly deals with the problem that the anticipated variances are implicit functions of the inclusion probabilities. The new optimization algorithm can be easily performed because it is based on a general decomposition of the measure of accuracy. A general sampling design which includes most of the one-stage sampling designs adopted in actual surveys is proposed, e.g., Simple Random Sampling Without Replacement (SRSWOR), Stratified SRSWOR, Stratified PPS, Designs with incomplete stratification, etc. The framework is based on a joint use of *balanced sampling designs* (Deville and Tillé 2004) which, depending upon the different definitions of the balancing equations, represents a wide-ranging sampling design and *superpopulation models for predicting* the unknown values of the variables of interest. The paper is structured as follows. Section 2 introduces definitions and notations. Section 3 and Section 4 illustrate the sampling design and the Anticipated Variance. The algorithm for defining the optimal inclusion probabilities is described in Section 5. In Section 6, some experiments based on real business data show the empirical properties of the algorithm. The conclusions are given in Section 7.

2 Definitions and notation

In this section, we introduce the concepts of ***estimation domain*** and ***planned domain*** which play a key role in the framework presented herein.

Let U be the reference population of N elements and let $U_d\ (d = 1,\ldots,D)$ be an ***estimation domain***, i.e., a generic sub-population of U with N_d elements, for which separate estimates must be calculated. Let y_{rk} denote the value of the r^{th} $(r = 1,\ldots,R)$ variable of interest attached to the k^{th} population unit and let γ_{dk} denote the domain membership indicator for unit k defined as

$$\gamma_{dk} = \begin{cases} 1 & \text{if } k \in U_d \\ 0 & \text{otherwise} \end{cases}. \tag{2.1}$$

We assume that the γ_{dk} values are available in the sampling frame and more than one value γ_{dk} $(d = 1,\ldots,D)$ can be 1 for each unit k; therefore, the estimation domains can overlap.

The parameters of interest are the $D \times R$ domain totals

$$t_{(dr)} = \sum_{k \in U} y_{rk}\gamma_{dk}\ (r = 1,\ldots,R; d = 1,\ldots,D). \tag{2.2}$$

Let $p(\cdot)$ be a single-stage without replacement sampling design and $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_k,\ldots,\pi_N)'$ be the N-vector of inclusion probabilities. Let s be the sample selected with probability $p(s)$. Denote by U_h $(h = 1,\ldots,H)$ the subpopulation of size $N_h = \sum_{k \in U_h} \delta_{hk}$ where $\delta_{hk} = 1$ if $k \in U_h$ and $\delta_{hk} = 0$ otherwise.

We focus on fixed size sampling designs which are those satisfying

$$\sum_{k \in s} \boldsymbol{\delta}_k = \mathbf{n}, \tag{2.3}$$

where $\boldsymbol{\delta}_k = (\delta_{1k},\ldots,\delta_{hk},\ldots,\delta_{Hk})'$ and $\mathbf{n} = (n_1,\ldots,n_h,\ldots,n_H)'$ is the vector of integer numbers defining the sample sizes fixed at the design stage. Since the sample size n_h, corresponding to U_h, does not vary among sample selections, the subpopulation U_h will be referred to as a ***planned domain*** in the sequel. A necessary but not sufficient condition for ensuring that (2.3) is satisfied is that the vector $\boldsymbol{\pi}$ is such that

$$\sum_{k \in U} \pi_k \boldsymbol{\delta}_k = \mathbf{n}. \tag{2.4}$$

In our setting, the planned domains can overlap; therefore, the unit k may have more than one value $\delta_{hk} = 1$ (for $h = 1,\ldots,H$). Let us suppose that the δ_{hk} values are known, and available in the sampling frame, for all population units. We suppose furthermore that the $N \times H$ matrix $(\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_k,\ldots,\boldsymbol{\delta}_N)'$ is non-singular.

The planned domains and their relationship with the estimation domains play a central role in our generalized framework. We assume that the estimation domains may be defined as an aggregation of complete planned domains, which ensure that the *expected* sample size in the d^{th} estimation domain U_d, say n_d, can be obtained as a simple aggregation of the expected sample sizes of the planned domains that are included within it. Finally, let $\hat{t}_{(dr)}$ be the Horvitz-Thompson (HT) estimator of $t_{(dr)}$ with

$$\hat{t}_{(dr)} = \sum_{k \in s} \frac{1}{\pi_k} y_{rk}\gamma_{dk}. \tag{2.5}$$

An example from business surveys. Suppose that the survey estimates must be calculated separately considering three domain types: *region* (with 20 modalities), *economic activity* (2 modalities: goods and services) and *enterprise size* (3 modalities: small, medium and large enterprises). That is, there are

$D = 20 + 2 + 3 = 25$ possible overlapping estimation domains. The planned domains can be defined with different options.

Option 1. The single planned domain U_h is identified by a specific intersection of the categories of the estimation domains. In this case $H = 20 \times 2 \times 3 = 120$ planned domains are defined. They represent a specific partition of U. The planned domains do not overlap and $\sum_h \delta_{hk} = 1$.

Option 2. The planned domains U_h coincide with the estimation domains. Therefore, $H = D = 25$ and the $\boldsymbol{\delta}'_k$ are defined as vectors with three 1's, so that $\sum_h \delta_{hk} = 3$. Recall that the planned domains overlap.

Option 3. The planned domains U_h are defined as (*i*) region by economic activity and (*ii*) economic activity by enterprise size; then, $H = (20 \times 2) + (2 \times 3) = 46$ with $\sum_h \delta_{hk} = 2$.

Other intermediate relationships among estimation and planned domains are possible.

It is emphasised that the planned domains represent the basis for defining broad classes of sampling designs. For instance, ***stratified sampling designs*** require that the planned domains do not overlap, as $\sum_h \delta_{hk} = 1$ and each U_h is referred to as a stratum. Therefore, Option 1 in the example above leads us to define a stratified sampling design. Furthermore, the strata defined as in Option 1 are the basis of the so-called "multi-way stratified sampling design" (Winkler 2001).

If $\sum_h \delta_{hk} > 1$, the sample sizes of the planned domains identified in Option 1 (strata) are not strictly controlled. Nevertheless, the sample sizes are still controlled at an aggregated level. In Option 2 of the example above, the sample sizes are controlled only for the estimation domains; while in Option 3, the sample sizes are controlled for the subsets of two different partitions, defined by (*i*) the region by economic activity and (*ii*) the economic activity by enterprise size. On the basis of the Winkler's definition, we denote the designs using these types of planned domains as ***Incomplete multi-way Stratified Sampling*** **(ISS)** ***designs***.

3 Sampling

Let $\mathbf{z}_k$ be a vector of auxiliary variables available for all $k \in U$. A sampling design $p(s)$ is said to be balanced on the auxiliary variables if and only if it satisfies the following ***balancing equations***

$$\sum_{k \in s} \frac{\mathbf{z}_k}{\pi_k} = \sum_{k \in U} \mathbf{z}_k \tag{3.1}$$

for each sample s such that $p(s) > 0$ (Deville and Tillé 2004). Depending on the auxiliary variables and the inclusion probabilities, equation (3.1) can be exactly or approximately satisfied in each possible sample; therefore, a balanced sampling design does not always exist. By specifying

$$\mathbf{z}_k = \pi_k \boldsymbol{\delta}_k, \tag{3.2}$$

equations (3.1) become

$$\sum_{k\in s}\boldsymbol{\delta}_k = \sum_{k\in U}\pi_k\boldsymbol{\delta}_k. \tag{3.3}$$

In this case, the balancing equations state that the sample size achieved in each subpopulation U_h is equal to the expected size. In different contexts, Ernst (1989) and Deville and Tillé (2004; page 905 Section 7.3), have proved that, (*i*) with the specification (3.2) and (*ii*) if the vector of the expected sample sizes, given by $\mathbf{n} = \sum_{k\in U}\pi_k\boldsymbol{\delta}_k$, includes only integer numbers, then a balanced sampling design always exists. Specification (3.2) defines sampling designs that guarantee equation (2.4), upon which we wish to focus on. Deville and Tillé (2004, pages 895 and 905), Deville and Tillé (2005, page 577) and Tillé (2006, page 168) have shown that several customary sampling designs may be considered as special cases of balanced sampling, by properly defining the vectors $\boldsymbol{\pi}$ and $\boldsymbol{\delta}_k$ of equation (3.2). These issues are illustrated in Remark 4.2 and in Section 6. Balanced samples may be drawn by means of the Cube method (Deville and Tillé 2004). This strongly facilitates the sample selection of incomplete stratified sampling designs that overcome the computational drawbacks of methods based on linear programming algorithms (Lu and Sitter 2002). The Cube method satisfies (3.1) exactly when (3.2) holds and $\mathbf{n}$ is a vector of integers. In the cases of SRSWOR and SSRSWOR, the standard sample selection methods can be used, as well as the Cube method. Deville and Tillé (2005) propose as approximation of the variance for the HT estimator, in the balanced sampling

$$E_p\left(\hat{t}_{(dr)} - t_{(dr)}\right)^2 \cong [N/(N-H)]\left[\sum_{k\in U}(1/\pi_k - 1)\,\eta^2_{(dr)k}\right] \tag{3.4}$$

where E_p denotes the sampling expectation and

$$\eta_{(dr)k} = y_{rk}\gamma_{dk} - \pi_k\boldsymbol{\delta}_k'\left[\mathbf{A}(\boldsymbol{\pi})\right]^{-1}\sum_{j\in U}\pi_j\left(1/\pi_j - 1\right)\boldsymbol{\delta}_j y_{rk}\gamma_{dk} \tag{3.5}$$

with

$$\mathbf{A}(\boldsymbol{\pi}) = \sum_{j\in U}\boldsymbol{\delta}_j\boldsymbol{\delta}_j'\pi_j\left(1-\pi_j\right). \tag{3.6}$$

Recently, the simulation results in Breidt and Chauvet (2011) confirm that equation (3.4) represents a good approximation of the sampling variance when the balanced equations are satisfied exactly. Variance estimation is studied in Deville and Tillé (2005).

4 Anticipated variance

Prior to sampling, the y_{rk} values are not known and the variance expressed in formula (3.4) cannot be used for planning the sampling precision at the design phase. In practice, it is necessary to either obtain some proxy values or predict the y_{rk} values based on superpopulation models that exploit auxiliary information. The increasing availability of auxiliary information (deriving by integration of administrative registers and survey frames) facilitates the use of predictions. Under a model-based inference, the y_{rk} values are assumed to be the realization of a superpopulation model M. The model we study has the following form:

$$\begin{cases} y_{rk} = f_r(\mathbf{x}_k;\boldsymbol{\beta}_r) + u_{rk} \\ E_M(u_{rk}) = 0 \ \ \forall k;\ E_M(u_{rk}^2) = \sigma_{rk}^2;\ E_M(u_{rk},u_{rl}) = 0 \ \ \forall k \neq l \end{cases}, \tag{4.1}$$

where $\mathbf{x}_k$ is a vector of predictors (available in the sampling frame), $\boldsymbol{\beta}_r$ is a vector of regression coefficients and $f_r(\mathbf{x}_k;\boldsymbol{\beta}_r)$ is a known function, u_{rk} is the error term and $E_M(\cdot)$ denotes the expectation under the model. The parameters $\boldsymbol{\beta}_r$ and the variances σ_{rk}^2 are assumed to be known, although in practice they are usually estimated. The model (4.1) is variable-specific and different models for different variables may be used and this does not create additional difficulty. As a measure of uncertainty, we consider the *Anticipated Variance* (AV) (Isaki and Fuller 1982):

$$\mathrm{AV}(\hat{t}_{(dr)}) = E_M E_p(\hat{t}_{(dr)} - t_{(dr)})^2. \tag{4.2}$$

A general expression for the AV under linear models was derived by Nedyalkova and Tillé (2008). Their formulation is obtained by considering a linear function $f_r(\cdot)$ and a unique set of auxiliary variables, $\mathbf{x}_k$, used for both the prediction of the y values and for balancing the sample. In our context, we have introduced $\mathbf{x}_k$ and $\mathbf{z}_k = \pi_k\boldsymbol{\delta}_k$, highlighting that the auxiliary variables can be different for prediction and balancing. The variables $\mathbf{x}_k$ must be as predictive of y_{rk} as possible, while the variables $\mathbf{z}_k$ play an instrumental role in controlling the sample sizes for sub-populations.

In the context considered here, inserting the approximate variance (3.4) in the equation (4.2), we obtain the approximate expression of the AV:

$$\mathrm{AAV}(\hat{t}_{(dr)}) = [N/(N-H)]\sum_{k\in U}(1/\pi_k - 1)E_M(\eta_{(dr)k}^2), \tag{4.3}$$

where the terms $\eta_{(dr)k}^2$ in (3.4) are replaced by $E_M(\eta_{(dr)k}^2)$. By defining

$$\tilde{y}_{rk} = f_r(\mathbf{x}_k;\boldsymbol{\beta}_r), \tag{4.4}$$

the equation (4.3) may be reformulated as

$$\mathrm{AAV}(\hat{t}_{(dr)}) = [N/(N-H)]\left[\sum_{k\in U}\frac{1}{\pi_k}(\tilde{y}_{rk}^2 + \sigma_{rk}^2)\gamma_{dk} - \sum_{k\in U}(\tilde{y}_{rk}^2 + \sigma_{rk}^2)\gamma_{dk} - \mathrm{AAV}_{3(dr)}\right], \tag{4.5}$$

where the third variance component of $\mathrm{AAV}(\hat{t}_{(dr)})$ is

$$\begin{aligned} \mathrm{AAV}_{3(dr)} &= \sum_{k\in U}(1-\pi_k)a_{(dr)k}(\boldsymbol{\pi})[2\tilde{y}_{rk}\gamma_{dk} - \pi_k a_{(dr)k}(\boldsymbol{\pi})] \\ &+ \sum_{k\in U}(1-\pi_k)[2b_{(dr)k}(\boldsymbol{\pi}) - \pi_k c_{(dr)k}(\boldsymbol{\pi})] \end{aligned} \tag{4.6}$$

and $a_{(dr)k}(\boldsymbol{\pi})$, $b_{(dr)k}(\boldsymbol{\pi})$ and $c_{(dr)k}(\boldsymbol{\pi})$ are real numbers defined respectively in equations (A1.4), (A1.7) and (A1.8) of Appendix A1.

Remark 4.1. Expression (4.5) is a cumbersome formula but, for all practical purposes, calculations may be simplified by considering a slight upward approximation by setting $b_{(dr)k}(\boldsymbol{\pi}) = c_{(dr)k}(\boldsymbol{\pi}) = 0$ in (4.6). The proof is given in Appendix A3. An upward approximation is a safe choice in this setting, since it averts from the risk of defining an insufficient sample size for the expected accuracy.

Remark 4.2. The SSRSWOR design is obtained if the planned domains define a unique partition of population (Option 1 of the example in Section 2) and the model (4.1) is specified so that the predicted values are: $\tilde{y}_{rk} = \bar{Y}_{rh}$ with $\sigma^2_{rk} = \sigma^2_{rh}$ (for $k \in U_h$). The AAV becomes

$$\mathrm{AAV}(\hat{t}_{(dr)}) = [N/(N-H)] \sum_{d=1}^{D} \sum_{h \in H_d} \sigma^2_{rh} N_h (N_h/n_h - 1), \tag{4.7}$$

where H_d is the set of planned domains included in U_d (see Appendix A4). Note that the expression (4.7) agrees with the *Result 2* of Nedyalkova and Tillé (2008), but for the term $N/(N-H)$. If $[N/(N-H)](1/N_h) \approx 1/(N_h - 1)$ the expression (4.7) would approximate the variance of the HT estimate in the SSRSWOR design. The above approximation is proved true when the number of domains H remains small compared to the overall population size N, and when the domain sizes N_h are large.

5 Determination of the optimal inclusion probabilities

The vector of π-values is determined by solving the following optimization problem:

$$\begin{cases} \mathrm{Min}\left(\sum_{k \in U} \pi_k c_k\right) \\ \mathrm{AAV}(\hat{t}_{(dr)}) \le \bar{V}_{(dr)} \quad (d = 1, \ldots, D; r = 1, \ldots, R), \\ 0 < \pi_k \le 1 \qquad\qquad (k = 1, \ldots, N) \end{cases} \tag{5.1}$$

where c_k is the cost for collecting information from unit k and $\bar{V}_{(dr)}$ is a fixed variance threshold corresponding to $\hat{t}_{(dr)}$. System (5.1) minimizes the expected cost ensuring that the anticipated variances are bounded and that the inclusion probabilities lie between 0 and 1. If all the c_k values are constants equal to 1, then the problem (5.1) minimizes the sample size. We note that in problem (5.1) the variances σ^2_{rk} in $\mathrm{AAV}(\hat{t}_{(dr)})$ are treated as known; in practice they must be estimated. In Section 6, an empirical evaluation is conducted in order to study the sensitivity of the overall sample size with different estimated values of σ^2_{rk}.

To solve (5.1), we rearrange the inequality constraints to obtain

$$\sum_{k \in U} \frac{(\tilde{y}^2_{rk} + \sigma^2_{rk})\gamma_{dk}}{\pi_k} \le \frac{N-H}{N} \bar{V}_{(dr)} + \sum_{k \in U} (\tilde{y}^2_{rk} + \sigma^2_{rk})\gamma_{dk} + \mathrm{AAV}_{3(dr)}. \tag{5.2}$$

By fixing the values of $\mathrm{AAV}_{3(dr)}$ appropriately, the optimization problem becomes a classical Linear Convex Separate Problem (LCSP; Boyd and Vanderberg 2004). Figure 5.1 depicts the flow chart of the algorithm (A prototype software implementing the algorithm is available at http://www.istat.it/it/strumenti/metodi-e-software/software.), which is organized into two nested loops: the **Outer Loop** (OL) and the **Inner Loop** (IL). The two loops are updated according to a *fixed point* algorithm scheme. The convergence under some approximations is shown in Appendix A2.

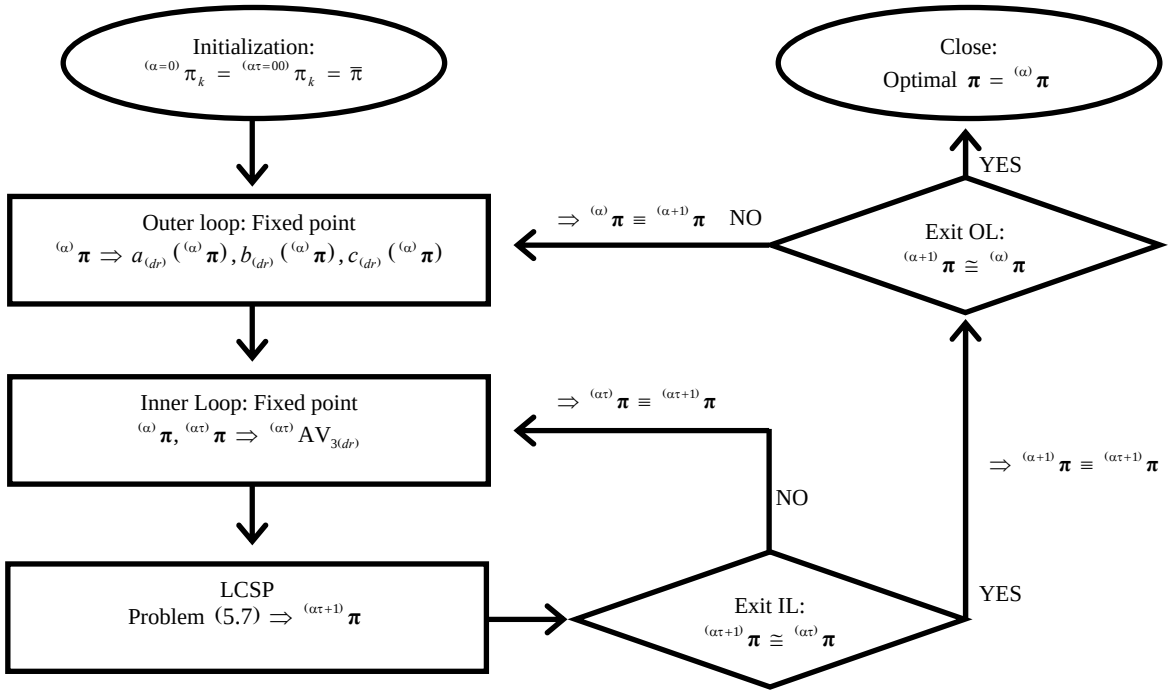

Figure 5.1 Algorithm flowchart

Initialization. At iteration $\alpha = 0$ of the OL, set $^{(\alpha=0)}\boldsymbol{\pi} = \left\{^{(\alpha=0)}\pi_k = \overline{\pi}; k = 1, \ldots, N\right\}$ with $0 < \overline{\pi} \leq 1$. A reasonable choice is $\overline{\pi} = 0.5$. At iteration $\tau = 0$ of the Inner Loop, set $^{(\alpha\tau=0)}\boldsymbol{\pi} = {}^{(\alpha)}\boldsymbol{\pi}$. Fix the N vector, $\boldsymbol{\varepsilon}$, of small positive values.

Outer loop

- ***Fixing the values for the Inner Loop***. In accordance with expressions (A1.4), (A1.7) and (A1.8) given in Appendix A1, the following real scalar values are computed

$$a_{(dr)k}(^{(\alpha)}\boldsymbol{\pi}) = \boldsymbol{\delta}_k' [\mathbf{A}(^{(\alpha)}\boldsymbol{\pi})]^{-1} \sum_{j \in U} \boldsymbol{\delta}_j \tilde{y}_{rj} \gamma_{dj} \left(1 - {}^{(\alpha)}\pi_j\right), \tag{5.3}$$

$$b_{(dr)k}(^{(\alpha)}\boldsymbol{\pi}) = \boldsymbol{\delta}_k' [\mathbf{A}(^{(\alpha)}\boldsymbol{\pi})]^{-1} \boldsymbol{\delta}_k \sigma_{rk}^2 \gamma_{dk} \left(1 - {}^{(\alpha)}\pi_k\right), \tag{5.4}$$

$$c_{(dr)k}(^{(\alpha)}\boldsymbol{\pi}) = \pi_k^2 \boldsymbol{\delta}_k' [\mathbf{A}(^{(\alpha)}\boldsymbol{\pi})]^{-1} \left[\sum_{j \in U} \boldsymbol{\delta}_j \boldsymbol{\delta}_j' \sigma_{rj}^2 \gamma_{dj} \left(1 - {}^{(\alpha)}\pi_j\right)^2\right] [\mathbf{A}(^{(\alpha)}\boldsymbol{\pi})]^{-1} \boldsymbol{\delta}_k. \tag{5.5}$$

- ***Launch of the Inner Loop***. The Inner Loop is executed until convergence.
- ***Updating or exiting.*** If the vector $^{(\alpha+1)}\boldsymbol{\pi}$ is such that $\left|^{(\alpha+1)}\boldsymbol{\pi} - {}^{(\alpha)}\boldsymbol{\pi}\right| > \boldsymbol{\varepsilon}$, then the Outer Loop is iterated by updating the vector $^{(\alpha)}\boldsymbol{\pi}$ with $^{(\alpha+1)}\boldsymbol{\pi}$. If $\left|^{(\alpha+1)}\boldsymbol{\pi} - {}^{(\alpha)}\boldsymbol{\pi}\right| \leq \boldsymbol{\varepsilon}$, then the Outer Loop closes and $^{(\alpha)}\boldsymbol{\pi}$ represents the optimal values solution to the problem of the system (5.1).

Inner Loop

- ***Fixing the values for the LCSP***. The following values are computed:

$$\begin{aligned} {}^{(\alpha\tau)}\mathrm{AAV}_{3(dr)} &= \sum_{k\in U}\left(1-{}^{(\alpha\tau)}\pi_k\right)a_{(dr)k}\left({}^{(\alpha)}\boldsymbol{\pi}\right)\left[2\tilde{y}_{rk}\gamma_{dk}-{}^{(\alpha\tau)}\pi_k a_{(dr)k}\left({}^{(\alpha)}\boldsymbol{\pi}\right)\right] \\ &+ \sum_{k\in U}\left(1-{}^{(\alpha\tau)}\pi_k\right)\left[2b_{(dr)k}\left({}^{(\alpha)}\boldsymbol{\pi}\right)-{}^{(\alpha\tau)}\pi_k c_{(dr)k}\left({}^{(\alpha)}\boldsymbol{\pi}\right)\right]. \end{aligned} \tag{5.6}$$

  in accordance with expression (A1.7) in Appendix A1.

- ***Solving the LCSP***. Considering the ${}^{(a\tau)}\mathrm{AAV}_{3(dr)}$ values as fixed, the ${}^{(\alpha\tau+1)}\boldsymbol{\pi}$ is obtained by solving, by a standard algorithm for a classical LCSP, the following optimization problem:

$$\begin{cases} \mathrm{Min}\left(\sum_{k\in U}{}^{(\alpha\tau+1)}\pi_k c_k\right) \\ \sum_{k\in U}\dfrac{\left(\tilde{y}_{rk}^2+\sigma_{rk}^2\right)\gamma_{dk}}{{}^{(\alpha\tau+1)}\pi_k}\le\dfrac{N-H}{N}\bar{V}_{(dr)}+\sum_{k\in U}\left(\tilde{y}_{rk}^2+\sigma_{rk}^2\right)\gamma_{dk}+{}^{(\alpha\tau)}\mathrm{AAV}_{3(dr)}. \\ 0<{}^{(\alpha\tau+1)}\pi_k\le 1 \qquad (k=1,\ldots,N) \end{cases} \tag{5.7}$$

- ***Updating or exiting.*** If the vector ${}^{(\alpha\tau+1)}\boldsymbol{\pi}$ is such that $\left|{}^{(\alpha\tau+1)}\boldsymbol{\pi}-{}^{(\alpha\tau)}\boldsymbol{\pi}\right|>\boldsymbol{\varepsilon}$, then the Inner Loop is iterated by updating the vector ${}^{(\alpha\tau)}\boldsymbol{\pi}$ with ${}^{(\alpha\tau+1)}\boldsymbol{\pi}$. If $\left|{}^{(\alpha\tau+1)}\boldsymbol{\pi}-{}^{(\alpha\tau)}\boldsymbol{\pi}\right|\le\boldsymbol{\varepsilon}$ then the Inner Loop closes and the updated vector ${}^{(\alpha+1)}\boldsymbol{\pi}$ for the Outer Loop is given by ${}^{(\alpha\tau+1)}\boldsymbol{\pi}$.

Remark 5.1. The problem of the system (5.7) can be solved by the algorithm proposed in Falorsi and Righi (2008, Section 3.1) which represents a slight modification of Chromy's algorithm (1987), originally developed for multivariate optimal allocation in SSRSWOR designs and implemented in standard software tools (see for example the Mauss-R software available at: http://www3.istat.it/strumenti/metodi/software/campione/mauss_r/). Alternatively, the LCSP can be dealt with by the SAS procedure NLP as suggested by Choudhry et al. (2012).

Remark 5.2. The algorithm distinguishes the ${}^{(\alpha)}\pi_k$ (updated in the Outer loop) from the ${}^{(\alpha\tau)}\pi_k$ (updated in the Inner loop). The innovation of the proposed algorithm lies precisely in this peculiarity. If this distinction between the inclusion probabilities is not made, i.e., ${}^{(\alpha\tau)}\boldsymbol{\pi}={}^{(\alpha)}\boldsymbol{\pi}$, we have observed in several experiments that the iterate solutions of the LCSP for each Outer Loop do not converge to a stationary point.

Remark 5.3. After the optimization phase, in which the $\boldsymbol{\pi}$ vector is defined as solution to problem of system (5.1), a *calibration phase* is performed (Falorsi and Righi 2008) to obtain calibrated inclusion probabilities, ${}_{\mathrm{cal}}\pi_k$, which modifies the *optimal* $\boldsymbol{\pi}$ vector marginally in order to satisfy $\sum_{k\in U}{}_{\mathrm{cal}}\pi_k\boldsymbol{\delta}_k=\mathbf{n}$, where $\mathbf{n}$ is a vector of integer numbers. The use of the Generalized Iterative Proportional Fitting algorithm (Dykstra and Wollan 1987) ensures that all resulting calibrated inclusion probabilities are in the $(0,1]$ interval.

6 Empirical evaluations

Several simulations were carried out on real and simulated data sets to investigate the empirical properties of the proposed sampling strategy. Here, we show the results obtained for a single real data exercise, referred to the 1999 population of enterprises having a number of employed persons between 1 and 99, and belonging to Computer and related economic activities (2-digits of the *Statistical classification of economic activities in the European Community rev.1*, abbreviated as NACE). Three experiments were performed. Experiment (a) checked whether the allocation obtained by the proposed algorithm converged towards the solution of the standard Chromy's algorithm for the SSRSWOR design. Experiment (b) compared the sample sizes of the standard SSRSWOR design with the Incomplete Stratified Sampling (ISS) design, in which the cross-classified strata were unplanned subpopulations; this experiment studied the risk of statistical burden due to repeated selection on different survey occasions. Finally, Experiment (c) measured the discrepancies between the expected Coefficients of Variation (CV) computed by the algorithm and the empirical CV obtained by a Monte Carlo simulation.

The c_k values were, in all three experiments, uniformly set equal to 1. The Anticipated Variance according to the approximation proposed in Remark 4.1 was also calculated.

The population chosen for the experiments had a size of $N = 10{,}392$ enterprises. The domains of interest identify two partitions of the target population: the *geographical region*, with 20 marginal domains (DOM1), and the *economic activity group* (3-digits of the NACE with 6 different groups) *by size class* (defined in terms of number of employed persons: $1 = 1-4; 2 = 5-9; 3 = 10-19;$ $4 = 20-99$), with 24 marginal domains (DOM2). The overall number of marginal domains was 44, while the number of cross-classified or multi-way strata with a not-zero population size was 360. The modal value of the population size distribution is 1, and 29.17% of the cross-classified strata have at most 2 units. This type of strata represents a critical issue in the context of standard stratified approaches. Indeed, for calculating unbiased variance estimates, these strata must be take-all strata (so that they do not contribute to the variance of the estimates), although the allocation rule would require fewer units and, in general, a non-integer number of sample units. The variables of interest were the *labour cost* and the *value added*, which are available for each population unit from an administrative data source. Typically both variables have highly skewed distributions.

The target estimates for all the empirical studies are the 88 totals at the domain level (2 variables by 44 marginal domains). In each experiment, the inclusion probabilities were determined by fixing the $\bar{V}_{(dr)} = (0.1 t_{(dr)})^2$ in (5.1), which is equivalent to fixing the maximum accepted level of the percent CV of the domain level estimates at 10%.

Empirical study (a). The first experiment took into account the partition DOM1. These domains represented both *planned* domains and *estimation* domains. Since the planned domains defined a partition of the population of interest, they could also be considered as strata in the standard sampling designs. The predictive working model was given by

$$\begin{cases} y_{rk} = \alpha_d + u_{rk} \;\; \forall k \in U_d \; (d = 1, \ldots, 20) \\ \\ E_M(u_{rk}) = 0, \; E_M(u_{rk}^2) = \sigma_{rd}^2 \;\; \forall k \in U_d; \;\; E_M(u_{rk}, u_{rl}) = 0 \;\; \forall k \neq l \end{cases}, \qquad (6.1)$$

where α_d is a fixed effect and the superpopulation variances σ^2_{rd} were estimated by means of the residual variance of the predictive model in each region. The algorithm proposed in Section 5 was performed using three different initial values of the inclusion probabilities $\bar{\pi}$, equal to 0.01, 0.50 and 0.99 respectively. The initial inclusion probability values had no impact on the final solution, although it was achieved with a different number of iterations. We note that the overall number of inner loops was 17 for $\bar{\pi} = 0.01$. The convergence was achieved with 13 inner loops for $\bar{\pi} = 0.50$; 14 inner loops were needed for $\bar{\pi} = 0.99$. However, after the ninth iteration, the three sampling sizes were quite similar (Figure 6.1). In the experiment, the overall sample sizes were 3,105 for the benchmark Chromy allocation and 3,110 for the method proposed here. However, the differences between the two sampling sizes at the domain level were fractional numbers that were always lower than 1, and with the absolute largest relative difference lower than 0.3%. This highlights that the proposed algorithm actually defines the same domain sampling sizes of those calculated by the benchmark allocation. With regards to convergence, the initial inclusion probability values have no impact on the final solution, although this is achieved with a different number of iterations.

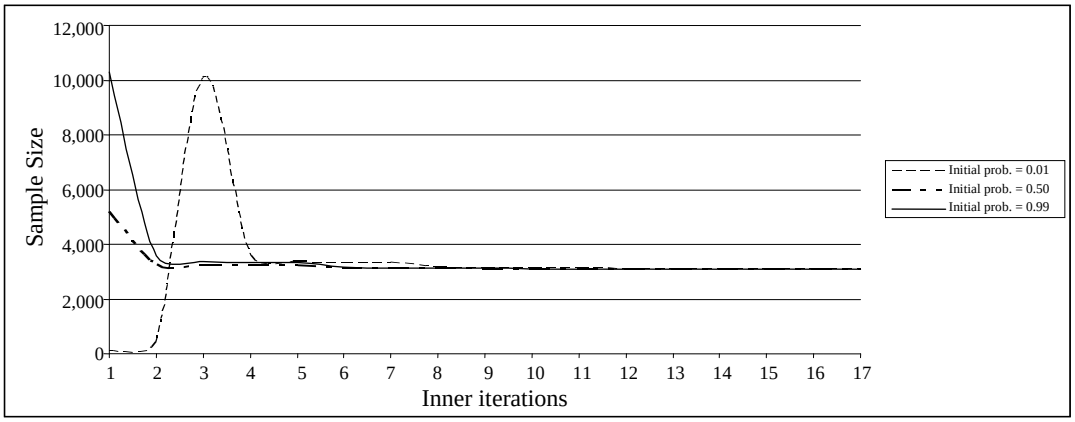

Figure 6.1 Convergence of the algorithm with different initial inclusion probabilities in the empirical study (a)

Similar results were obtained if the domains of interests were identified by the partition DOM2.

Empirical study (b). Let U_{d_1} be a specific region $(d_1 = 1, \ldots, 20)$ of DOM1, and let U_{d_2} (with $d_2 = 1, \ldots, 24$) be a specific economic activity group by the enterprise size class of the partition DOM2. Two prediction models, M_1 and M_2, were used. Referring to the notation of the ANOVA models, M_1 is the saturated model given by

$$\begin{cases} y_{rk} = \alpha_{d_1} + \lambda_{d_2} + (\alpha\lambda)_{d_1 d_2} + u_{rk} \;\; \forall k \in U_{d_1} \cap U_{d_2} \\ E_M(u_{rk}) = 0,\; E_M(u^2_{rk}) = \sigma^2_{r(d_1 d_2)} \;\; \forall k \in U_{d_1} \cap U_{d_2};\;\; E_M(u_{rk}, u_{rl}) = 0 \;\; \forall k \neq l \end{cases}, \qquad (6.2)$$

in which α_{d_1} and λ_{d_2} are the main effects, related to the domains U_{d_1} and U_{d_2} respectively and with $(\alpha\lambda)_{d_1 d_2}$ as the interaction effect. The model variances $\sigma^2_{r(d_1 d_2)}$ were estimated by means of the ordinary

least square method, by computing the variances of the residual terms at the $U_{d_1} \cap U_{d_2}$ level. Model M_2 is identical to model M_1 without the interaction factor. Table 6.1 shows the goodness of fit of the two models.

Table 6.1
Goodness of fit of the models used for the prediction

Model	Goodness of fit R^2%	
	Labour cost	Value added
Model M_1 (Expression 6.2)	68.1	64.1
Model M_2 (Expression 6.2 without interactions)	65.1	61.0

Three different allocations were considered for the SSRSWOR in the case of model M_1 : (*i*) no stratum sample size constraint is given; (*ii*) at least 1 sample unit per stratum is required (to obtain unbiased point estimates); (*iii*) at least 2 sample units per stratum are required (to achieve unbiased variance estimates) for all strata having a population size of 2 or more enterprises. The first two allocations were rather theoretical since in all the business surveys conducted by the Italian National Statistical Institute, the selection of at least two units per stratum is required. The results of the experiment are shown in Table 6.2 below. Only the results for the case in which the initial inclusion probabilities were equal to $\bar{\pi} = 0.50$ are investigated herein; identical sample sizes were obtained with the other initial values of the inclusion probabilities, with a slightly slower convergence process. The three SSRSWOR designs have 716.6, 944 and 1,042 sample units respectively. The Incomplete stratified Sampling (ISS) design with model M_1 led to 936 units; while model M_2 led to 991 units. The better result obtained by model M_1 with respect to model M_2 was due to the fact that model M_1 had a better fit. Finally, the ISS designs helped tackling the statistical burden of respondent enterprises. Indeed, assuming that the inclusion probabilities remain fixed for the different survey occasions, their distributions may be used to assess the statistical burden in repeated surveys. Table 6.2 shows that the number of enterprises drawn with certainty in each survey occasion was 175 for the third SSRSWOR designs, while 30 and 40 enterprises were selected with certainty in the first and second ISS designs, respectively. Analysing the sizes (in terms of employed persons) of the enterprises included in the sample with certainty, the third SSRSWOR design had an average size equal to 20.6. In some cases, enterprises with 2 employed persons were included in the sample with certainty. Conversely, we observe that in the first and second ISS designs, the enterprises with minimum size had 17 and 16 employed persons respectively, and an average size larger than 40 units.

Table 6.2
Sample sizes and distribution of the enterprises included in the sample with certainty, for different sampling designs

Sampling design		Sample size	Enterprises selected with certainty		
			Number	Number of employed	
				Average	Minimum
Standard Stratified with M_1 model	No stratum sample size constraint	716.6	10	47.0	23.0
	At least 1 sample unit per stratum	944.0	119	24.0	2.0
	At least 2 sample units per stratum	1,042.0	175	20.6	2.0
Incomplete Stratified Sampling with M_1 model		936.0	30	50.1	17.0
Incomplete Stratified Sampling with M_2 model without interactions		991.0	40	42.9	16.0

Finally, to assess the solution's sensitivity, the experiment was repeated artificially and the prediction values of $\tilde{y}_{rk}$ and $\tilde{\sigma}^2_{rk}$ in the optimization problem (5.1) were changed. In particular, we increased the prediction values of $\tilde{\sigma}^2_{rk}$ by 20% and 120% respectively, and decreased by 20% the $\tilde{y}_{rk}$ values predicted by model M_1. As expected, the sample sizes increased, but the SSRSWOR design with at least 1 sample unit per stratum and the first ISS design roughly defined the same sample sizes (Table 6.3).

Table 6.3
Sample sizes with modified expected values of the predictions of model (4.1)

Sampling design		Sample size		
		$\tilde{\sigma}^2_{rk}$ increased by 20%	$\tilde{\sigma}^2_{rk}$ increased by 120%	$\tilde{y}_{rk}$ decreased by 20%
SSRSWOR with M_1 model	No stratum sample size constraint	821.0	1,269.0	993.8
	At least 1 sample unit per stratum	1,035.0	1,472.0	1,206.0
	At least 2 sample units per stratum	1,125.0	1,536.0	1,283.0
ISS design with M_1 model		1,039.7	1,460.9	1,207.5

Empirical study (c). The heteroschedastic linear prediction model M_3 was used:

$$\begin{cases} y_{rk} = \alpha_r + \varphi_r x_k + u_{rk} \\ E_M(u_{rk}) = 0, \;\; E_M(u^2_{rk}) = \sigma^2_{rk} = \sigma^2_r x_k \;\; \forall k \in U; \; E_M(\varepsilon_{rk}, \varepsilon_{rl}) = 0 \;\; \forall k \neq l \end{cases}, \tag{6.3}$$

where x_k is the *number of employed persons* in the k^{th} enterprise, and α_r and φ_r are the regression parameters. Note that the number of employed persons is available in the sampling frame in Italy.

Two different model variance estimates were carried out:
(a) $\tilde{\sigma}^2_{rk} = 1/N_{(X=x_k)} \sum_{k \in U_{(X=x_k)}} (y_{rk} - \mathrm{A}_r - \mathrm{F}_r x_k)^2$ and (b) $\tilde{\sigma}^2_{rk} = \tilde{\sigma}^2_r x_k$, in which $\tilde{\sigma}^2_r = 1/(N-2) \sum_{k \in U} [(y_{rk} - \mathrm{A}_r - \mathrm{F}_r x_k)/x_k]^2$, where $U_{(X=x)}$ is the population of enterprises, of size $N_{(X=x)}$, for which the variable X assumes the value x; A_r and F_r are the weighted least square estimates for the complete enumerated population of α_r and φ_r respectively. The sum of the estimated model variances obtained with method (a) is smaller than that obtained with method (b). This was reflected in the computed sample sizes. The first allocation defined an overall sample size of 927 units, while the sample size of the second allocation was 951. Successively, 1,000 samples were drawn for both allocations and the ratios $\mathrm{RCV}(\hat{t}_{(dr)}) = \mathrm{ECV}(\hat{t}_{(dr)})/\mathrm{SCV}(\hat{t}_{(dr)})$ were calculated, with $\mathrm{ECV}(\hat{t}_{(dr)}) = \left[\sqrt{\mathrm{AAV}(\hat{t}_{(dr)})}/\hat{t}_{(dr)}\right]100$ as the expected CV (%) and

$$\mathrm{SCV}(\hat{t}_{(dr)}) = 100 \sqrt{(1/I)\left[\sum_{i=1}^{I} \hat{t}^i_{(dr)} - (1/I)\sum_{i=1}^{I} \hat{t}^i_{(dr)}\right]^2} \Big/ (1/I)\sum_{i=1}^{I} \hat{t}^i_{(dr)}$$

as the simulated (or empirical) CV, obtained as a result of the simulation, having denoted with $\hat{t}^i_{(dr)}$ the HT estimate in the i^{th} iteration and $I = 1{,}000$. For the sake of brevity, only the the main results of allocation (b) are shown in Figure 6.2, for DOM1 and DOM2 respectively, and both variables of interest. Examining the figure on the left, we emphasize that the simulation generally produces a simulated CV that is smaller

than expected, with an RCV ratio larger than 1 for both variables. One exception occurs, for the value added in one domain of DOM1.

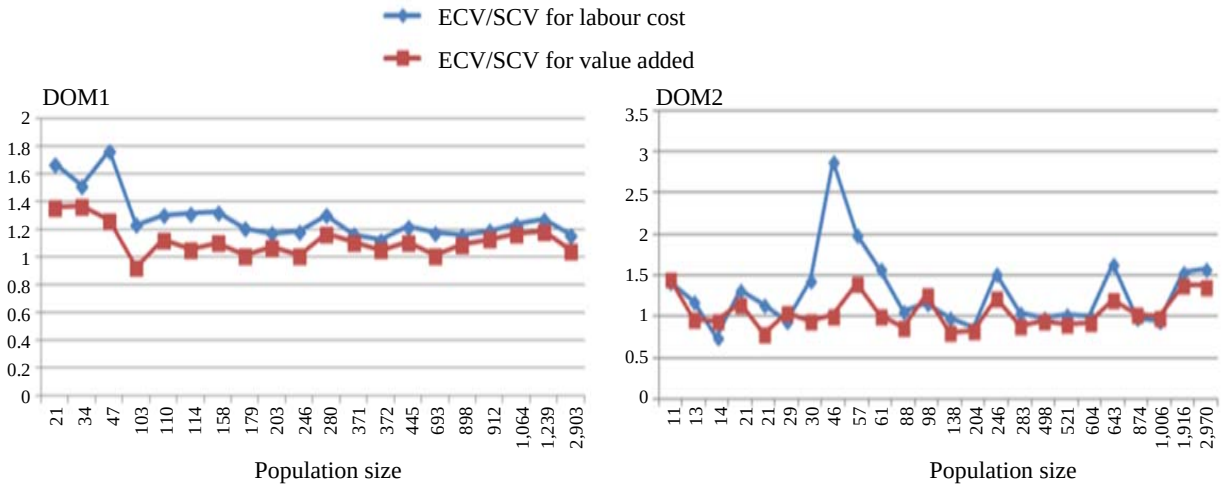

Figure 6.2 RCVs by population size for labour cost and value added

RCV lower than 1 may be explained by the increase of the domain sample sizes, due to the calibration step. We note that in general, these discrepancies are observed in domains with a small population size; thus, the calibration step may have a non-negligible impact. The figure on the right shows more articulated and conflicting empirical evidence. First, we note that the RCV are often larger or very close to 1. Nevertheless, in three domains, the value added variable has simulated CV's equal to 11.5%, 12.0% and 12.3%. In these rare cases, and in some others (labour cost in two domains), the discrepancies are coherent with the findings of Deville and Tillé (2005) on the empirical properties of variance approximation for balanced sampling.

7 Conclusions

The paper proposes a new approach for defining the optimal inclusion probabilities in various survey contexts, which are characterized by the need to disseminate survey estimates of prefixed accuracy, for a multiplicity of both variables and domains of interest.

This paper's main contribution is the practical computation of these probabilities by means of a new algorithm, which is suitable for a general multi-way sampling design in which the standard stratified sampling represents a special case. The proposed approach, the algorithm and the final computation are domain- and variable-driven.

In our framework, the domain membership indicator variables are assumed to be known, while the variables of interest are not known. The procedure is, then, applied on the predicted values of the characteristics of interest via a superpopulation model, and the algorithm enables taking into account

model uncertainty; this reflects the non-knowledge of the values of variables of interest. Using the Anticipated Variance as the measure of the estimators' precision, this approach overcomes the limits of the standard algorithms for the sample allocation, in which the variables of interest driving the solution are assumed to be known.

The proposed algorithm exploits standard procedure, but does present some computational innovations which may be useful for dealing with the complexity deriving from the fact that the Anticipated Variances are implicit functions of the inclusion probabilities. The algorithm was tested on simulated and real survey data, to evaluate its performance and properties. The results of a small set of experiments are presented here. They confirm an improvement, in terms of efficiency, of the sampling strategy. A natural generalization of the case examined here may be developed by considering, as known during the design planning stage, the indicators of the domains and other quantitative independent variables. We note that the Anticipated Variance considering only the domain indicators is larger than the Anticipated Variance of this more general case. Thus, our solution represents an upper (and somehow robust) boundary solution in the design phase. Furthermore, the algorithmic solution can be easily adapted to this more general situation.

Acknowledgements

This research was funded by the partnership of the Global Strategy to improve Agricultural and Rural Statistics: http://www.fao.org/economic/ess/ess-capacity/ess-strategy/en/.

Appendix A1

AV of the HT estimator

Let us consider the residual $\eta_{(dr)k}$ as expressed by equation (3.5), and replace the term y_{rk} with $\tilde{y}_{rk} + u_{rk}$, thus obtaining

$$\eta_{(dr)k} = (\tilde{y}_{rk} + u_{rk})\gamma_{dk} - \pi_k \boldsymbol{\delta}'_k [\mathbf{A}(\boldsymbol{\pi})]^{-1} \sum_{j \in U} \pi_j \boldsymbol{\delta}_j (\tilde{y}_{rj} + u_{rj})\gamma_{dj} (1/\pi_j - 1). \quad \text{(A1.1)}$$

The weighted least predictions of $\tilde{y}_{rk}\gamma_{dk}$ and $u_{rk}\gamma_{dk}$, with predictors $\pi_k \boldsymbol{\delta}_k$ and weights $1/\pi_k - 1$, are

$$\hat{\tilde{y}}_{(dr)k} = \pi_k a_{(dr)k} \quad \text{(A1.2)}$$

and

$$\hat{u}_{(dr)k} = \pi_k \boldsymbol{\delta}'_k [\mathbf{A}(\boldsymbol{\pi})]^{-1} \sum_{j \in U} \pi_j \boldsymbol{\delta}_j u_{rj}\gamma_{dj} (1/\pi_j - 1), \quad \text{(A1.3)}$$

with

$$a_{(dr)k}(\boldsymbol{\pi}) = \boldsymbol{\delta}'_k [\mathbf{A}(\boldsymbol{\pi})]^{-1} \sum_{j \in U} \pi_j \boldsymbol{\delta}_j \tilde{y}_{rj}\gamma_{dk} (1/\pi_j - 1). \quad \text{(A1.4)}$$

Using the formulae (A1.2) and (A1.3), the expression (A1.1) may be reformulated as $\eta_{(dr)k} = (\tilde{y}_{rk} + u_{rk})\gamma_{dk} - [\hat{\tilde{y}}_{(dr)k} + \hat{u}_{(dr)k}]$. Therefore, the model expectation of $\eta^2_{(dr)k}$ is

$$E_M(\eta^2_{(dr)k}) = (\tilde{y}_{rk}\gamma_{dk} - \hat{\tilde{y}}_{(dr)k})^2 + E_M[(u_{rk}\gamma_{dk} - \hat{u}_{(dr)k})^2] + \text{Mean zero terms}, \quad \text{(A1.5)}$$

because $E_M(u_{rk}) = 0$. Furthermore,

$$E_M[(u_{rk}\gamma_{dk} - \hat{u}_{(dr)k})^2] = \sigma^2_{rk}\gamma_{dk} + E_M(\hat{u}_{(dr)k})^2 - 2E_M(u_{rk}\gamma_{dk}, \hat{u}_{(dr)k}), \quad \text{(A1.6)}$$

where $E_M(u_{rk}\gamma_{dk}\hat{u}_{(dr)k}) = \pi_k b_{(dr)k}(\boldsymbol{\pi})$ and $E_M(\hat{u}_{(dr)k})^2 = \pi^2_k c_{(dr)k}(\boldsymbol{\pi})$, with

$$b_{(dr)k}(\boldsymbol{\pi}) = \boldsymbol{\delta}'_k[\mathbf{A}(\boldsymbol{\pi})]^{-1}\boldsymbol{\delta}_k\sigma^2_{rk}\gamma_{dk}(1-\pi_k) \quad \text{(A1.7)}$$

and

$$c_{(dr)k}(\boldsymbol{\pi}) = \boldsymbol{\delta}'_k[\mathbf{A}(\boldsymbol{\pi})]^{-1}\left[\sum_{j\in U}\boldsymbol{\delta}_j\boldsymbol{\delta}'_j\sigma^2_{rj}\gamma_{dj}(1-\pi_j)^2\right][\mathbf{A}(\boldsymbol{\pi})]^{-1}\boldsymbol{\delta}_k. \quad \text{(A1.8)}$$

Expression (4.5) is easily derived by plugging expressions from (A1.2) to (A1.8) into equation (4.3).

Appendix A2

Convergence of the algorithm

The optimization problem (5.1) is solved by two nested *fixed point iterations*. Given an unknown vector $\mathbf{x}$ of dimension q, the fixed point iteration chooses an initial guess ${}^{(0)}\mathbf{x}$. Then, it computes subsequent iterates by ${}^{(\tau+1)}\mathbf{x} = \mathbf{g}({}^{(\tau)}\mathbf{x})$, with $\tau = 1, 2, \ldots$, with $\mathbf{g}(\cdot)$ being a system of q updating equations. The multivariate function $\mathbf{g}$ has a fixed point in a domain $Q \subseteq \Re^q$ if $\mathbf{g}$ maps Q in Q. Let $J_{\mathbf{g}}(\mathbf{x})$ be the Jacobian matrix of first partial derivate of $\mathbf{g}$ evaluated at $\mathbf{x}$, if there exists a constant $\rho < 1$ such that, in some natural matrix norm, $\|J_{\mathbf{g}}(\mathbf{x})\| \leq \rho, \mathbf{x} \in Q$, $\mathbf{g}$ has a unique fixed point $\mathbf{x}^* \in Q$, and the fixed point iteration is guaranteed to converge to $\mathbf{x}^*$ for any initial guess chosen in Q. As regards the proposed algorithm, the convergence of the IL and OL is obtained when the terms ${}^{(\alpha\tau)}\mathrm{AAV}_{3(dr)}$ converge to the fixed point. This means that the vectors ${}^{(\alpha)}\boldsymbol{\pi}$ and ${}^{(\alpha\tau)}\boldsymbol{\pi}$ do not change in the OL and IL iterations. The demonstration below considers the method proposed by Chromy (1987) to solve the LCSP of system (5.7), and makes use of some reasonable assumptions: (1) $\hat{u}_{(dr)k} \cong 0$; (2) $[N/(N-H)] \cong 1$; (3) $\hat{\tilde{y}}_{rk} \cong \tilde{y}_{rk}$; (4) ${}^{(\alpha)}\pi_k \cong {}^{(\alpha\tau)}\Delta\,{}^{(\alpha\tau)}\pi_k$ with $0 < {}^{(\alpha\tau)}\Delta \leq 1$; (5) $c_k \cong \bar{c}$. Assumption (1) corresponds to the upward approximation of the Anticipated Variance, given in Remark 4.1, and implies that $b_{(dr)k}({}^{(\alpha)}\boldsymbol{\pi}) = c_{(dr)k}({}^{(\alpha)}\boldsymbol{\pi}) = 0$. Assumption (3) implies that $a_{(dr)k}({}^{(\alpha)}\boldsymbol{\pi})\tilde{y}_{rk}\gamma_{dk} \cong \tilde{y}^2_{rk}\gamma_{dk}/{}^{(\alpha)}\pi_k$. Assumption (4) states that the structure of the inclusion probabilities remains roughly constant in the different IL iterations. The assumption becomes reasonable considering that the updating equation A2.2 below (of a given inclusion probability) is essentially determined by the variance threshold that requires the largest sample size. It is plausible to hypothesize that this threshold remains more or less the same in the subsequent IL iterations of a given OL.

Proof of convergence of the Inner Loop. By reformulating expression (4.6) in accordance with the assumptions from (1) to (4),

$$ {}^{(\alpha\tau+1)}\mathrm{AAV}_{3(dr)} = \sum_{k\in U}\left[\left(\frac{1}{{}^{(\alpha\tau+1)}\pi_k}-1\right)\left(2\frac{\tilde{y}_{rk}^2\gamma_{dk}}{{}^{(\alpha\tau+1)}\Delta}-\frac{\tilde{y}_{rk}^2\gamma_{dk}}{{}^{(\alpha\tau+1)}\Delta^2}\right)\right]. \tag{A2.1} $$

Considering in problem (5.7) that the ${}^{(a\tau)}\mathrm{AAV}_{3(dr)}$ values are fixed, each value of the vector ${}^{(\alpha\tau+1)}\boldsymbol{\pi}$ is obtained as a solution of the LCSP with the Chromy algorithm. Denote with $\alpha\tau v^*$ the iteration of the Chromy algorithm into which it converges, where ${}^{(\alpha\tau v^*+1)}\boldsymbol{\pi} \cong {}^{(\alpha\tau v^*)}\boldsymbol{\pi}$. Then, the IL updates the generic probability in accordance with the expression

$$ {}^{(\alpha\tau+1)}\pi_k = \left[\sum_{(dr)} {}^{(\alpha\tau v^*+1)}\phi_{(dr)}\frac{\left(\tilde{y}_{rk}^2+\sigma_{rk}^2\right)\gamma_{dk}}{\overline{c}}\right]^{1/2}, \tag{A2.2} $$

where the right-hand term represents the updating formula of the Chromy algorithm, and $\sum_{(dr)}$ stands for $\sum_{d=1}^{D}\sum_{r=1}^{R}$, and ${}^{(\alpha\tau v^*+1)}\phi_{(dr)}$ is the generalized Lagrange multiplier, where

$$ \begin{aligned} {}^{(\alpha\tau v^*+1)}\phi_{(dr)} &= {}^{(\alpha\tau v^*)}\phi_{(dr)}\left[\frac{{}^{(\alpha\tau v^*)}V_{(dr)}}{\ddot{V}_{(dr)}+{}^{(\alpha\tau)}\mathrm{AAV}_{3(dr)}}\right]^2, \\ {}^{(\alpha\tau v^*)}V_{(dr)} &= \sum_{k\in U}\frac{\left(\tilde{y}_{rk}^2+\sigma_{rk}^2\right)\gamma_{dk}}{{}^{(\alpha\tau v^*)}\pi_k} \end{aligned} \tag{A2.3} $$

and

$$ \ddot{V}_{(dr)} = \overline{V}_{(dr)} + \sum_{k\in U}\left(\tilde{y}_{rk}^2+\sigma_{rk}^2\right)\gamma_{dk}. $$

The Kuhn-Tucker theory states that ${}^{(\alpha\tau v^*)}\phi_{(dr)}\left[{}^{(\alpha\tau v^*)}V_{(dr)} - \left(\ddot{V}_{(dr)} + {}^{(\alpha\tau)}\mathrm{AV}_{3(dr)}\right)\right] = 0$; therefore, ${}^{(\alpha\tau v^*+1)}\phi_{(dr)} = {}^{(\alpha\tau v^*)}\phi_{(dr)}$ and ${}^{(\alpha\tau v^*+1)}\phi_{(dr)} > 0$ iff ${}^{(\alpha\tau v^*)}V_{(dr)}\big/\left(\ddot{V}_{(dr)} + {}^{(\alpha\tau)}\mathrm{AV}_{3(dr)}\right) = 1$. Chromy asserts that few ${}^{(\alpha\tau v^*)}\phi_{(dr)}$ (for $r = 1,\ldots,R; d = 1,\ldots,D$) are larger than zero, and that in most cases, only one value is strictly positive. Denoting with ${}^{(\alpha\tau)}\mathbf{AAV}_3 = \left({}^{(\alpha\tau)}\mathrm{AAV}_{3(11)},\ldots,{}^{(\alpha\tau)}\mathrm{AAV}_{3(1R)},\ldots,{}^{(\alpha\tau)}\mathrm{AAV}_{3(DR)}\right)'$, we define ${}^{(\alpha\tau+1)}\mathbf{AAV}_3 = \mathbf{g}\left({}^{(\alpha\tau)}\mathbf{AAV}_3\right)$ as the system of $D\times R$ updating equations where the generic $(\overline{dr})$ equation of the system

$$ \begin{aligned} g_{(\overline{dr})}\left({}^{(\alpha\tau)}\mathbf{AAV_3}\right) &\cong \sum_{k\in U}\left(2\frac{\tilde{y}_{\overline{r}k}^2\gamma_{\overline{d}k}}{{}^{(\alpha\tau+1)}\Delta}-\frac{\tilde{y}_{\overline{r}k}^2\gamma_{\overline{d}k}}{{}^{(\alpha\tau+1)}\Delta^2}\right) \\ &\times \left\{\left[\sum_{(dr)} {}^{(\alpha\tau v^*)}\phi_{(dr)}\left[\frac{{}^{(\alpha\tau v^*)}V_{(dr)}}{\ddot{V}_{(dr)}+{}^{(\alpha\tau)}\mathrm{AAV}_{3(dr)}}\right]^2\frac{\left(\tilde{y}_{rk}^2+\sigma_{rk}^2\right)\gamma_{dk}}{\overline{c}}\right]^{-1/2}-1\right\}, \end{aligned} \tag{A2.4} $$

is obtained by plugging expression (A2.2) into (A2.1). If the convergence is obtained, then in the last iteration, ${}^{(\alpha\tau+1)}\mathbf{AAV}_3 \cong {}^{(\alpha\tau)}\mathbf{AAV}_3$. The function of equation (A2.4) is continuous and differentiable. Moreover, it maps onto the interval of the possible values of $\mathrm{AAV}_{3(dr)}$. Then, the IL converges if the following condition is fulfilled:

$$\left\| J_{\mathbf{g}}\left(\mathbf{AAV}_3\right)\right\| \leq 1. \tag{A2.5}$$

The Jacobian matrix is positive semi-defined, and a well-known result states that trace $\left(J_{\mathbf{g}}J'_{\mathbf{g}}\right) \leq$ trace $\left(J_{\mathbf{g}}\right)^2$. By considering the Frobenius norm $\left\|J_{\mathbf{g}}\right\|_F = \sqrt{\text{trace}\left(J_{\mathbf{g}}J'_{\mathbf{g}}\right)}$, it is $\left\|J_{\mathbf{g}}\right\|_F \leq$ trace $\left(J_{\mathbf{g}}\right)$. Thus we can take into account the trace of the Jacobian matrix to verify condition (A2.5). Let $g'_{(\overline{dr})} = \partial g_{(\overline{dr})}\left({}^{(\alpha\tau-1)}\mathbf{AAV}_{3(dr)} \big/ \partial\, {}^{(\alpha\tau-1)}\mathrm{AAV}_{3(\overline{dr})}\right)$ be the $(\overline{dr})$ element of the diagonal of $J_{\mathbf{g}}\left(\mathbf{AAV}_3\right)$. Using the Kuhn-Tucker condition ${}^{(\alpha\tau v^*)}V_{(dr)} \big/ \left(\ddot{V}_{(dr)} + {}^{(\alpha\tau)}\mathrm{AV}_{3(dr)}\right) = 1$,

$$g'_{(\overline{dr})} = \sum_{k\in U}\left(2\frac{\tilde{y}^2_{\overline{r}k}\gamma_{\overline{d}k}}{{}^{(\alpha\tau+1)}\Delta} - \frac{\tilde{y}^2_{\overline{r}k}\gamma_{\overline{d}k}}{{}^{(\alpha\tau+1)}\Delta^2}\right)\left[\sum_{(dr)} {}^{(\alpha\tau v^*)}\phi_{(dr)}\frac{\left(\tilde{y}^2_{rk} + \sigma^2_{rk}\right)\gamma_{dk}}{\overline{c}}\right]^{-3/2}$$

$$\times \; {}^{(\alpha\tau v^*)}\phi_{(\overline{dr})}\frac{1}{{}^{(\alpha\tau v^*)}V_{(\overline{dr})}}\frac{\left(\tilde{y}^2_{\overline{r}k} + \sigma^2_{\overline{r}k}\right)\gamma_{\overline{d}k}}{\overline{c}}.$$

Since many ${}^{(\alpha\tau v^*)}\phi_{(\overline{dr})} = 0$ (Chromy 1987), the respective $g'_{(\overline{dr})}$ is null. When ${}^{(\alpha\tau v^*)}\phi_{(\overline{dr})} > 0$, then

$$g'_{(\overline{dr})} \leq \sum_{k\in U}\left(2\frac{\tilde{y}^2_{\overline{r}k}\gamma_{\overline{d}k}}{{}^{(\alpha\tau+1)}\Delta} - \frac{\tilde{y}^2_{\overline{r}k}\gamma_{\overline{d}k}}{{}^{(\alpha\tau+1)}\Delta^2}\right)\left[{}^{(\alpha\tau v^*)}\phi_{(\overline{dr})}\frac{\left(\tilde{y}^2_{\overline{r}k} + \sigma^2_{\overline{r}k}\right)\gamma_{\overline{d}k}}{\overline{c}}\right]^{-3/2} \times {}^{(\alpha\tau v^*)}\phi_{(\overline{dr})}\frac{1}{{}^{(\alpha\tau v^*)}V_{(\overline{dr})}}\frac{\left(\tilde{y}^2_{\overline{r}k} + \sigma^2_{\overline{r}k}\right)\gamma_{\overline{d}k}}{\overline{c}}$$

$$= \sum_{k\in U}\left(2\frac{\tilde{y}^2_{\overline{r}k}\gamma_{\overline{d}k}}{{}^{(\alpha\tau+1)}\Delta} - \frac{\tilde{y}^2_{\overline{r}k}\gamma_{\overline{d}k}}{{}^{(\alpha\tau+1)}\Delta^2}\right)\frac{1}{\sqrt{{}^{(\alpha\tau v^*)}\phi_{(\overline{dr})}\frac{\left(\tilde{y}^2_{\overline{r}k} + \sigma^2_{\overline{r}k}\right)\gamma_{\overline{d}k}}{\overline{c}}\,{}^{(\alpha\tau v^*)}V_{(\overline{dr})}}}$$

$$\leq \sum_{k\in U}\frac{\frac{\tilde{y}_{\overline{r}k}\gamma_{\overline{d}k}}{{}^{(\alpha\tau+1)}\Delta}\left(2 - \frac{1}{{}^{(\alpha\tau+1)}\Delta}\right)}{\sqrt{\overline{c}\,{}^{(\alpha\tau v^*)}\phi_{(\overline{dr})}\gamma_{\overline{d}k}}\,{}^{(\alpha\tau v^*)}V_{(\overline{dr})}} << 1.$$

Therefore, the trace $\left(J_{\mathbf{g}}\right)$ should be less than 1.

Proof of convergence of the Outer Loop. Let ${}^{(\alpha\tau+1)}\boldsymbol{\pi}$ be the fixed point solution of the IL; then, the OL updates the vector ${}^{(\alpha)}\boldsymbol{\pi}$ with ${}^{(\alpha+1)}\boldsymbol{\pi} = {}^{(\alpha\tau+1)}\boldsymbol{\pi}$. Under conditions (1), (2) and (3),

$${}^{(\alpha+1)}\mathrm{AAV}_{3(dr)} = \sum_{k\in U}\left(\frac{1}{{}^{(\alpha\tau+1)}\pi_k} - 1\right)\tilde{y}^2_{rk}\gamma_{dk}. \tag{A2.6}$$

Plugging expression (A2.2) into formula (A2.6) when the IL converges, the system of $D \times R$ updating equations of ${}^{(\alpha+1)}\mathbf{AAV}_3$ is given by ${}^{(\alpha+1)}\mathbf{AAV}_3 = \mathbf{j}\left({}^{(\alpha\tau)}\mathbf{AAV}_3\right)$, where the generic equation of $\mathbf{j}$ is

$$\begin{aligned} {}^{(\alpha+1)}\mathrm{AAV}_{3(dr)} &= j_{(\overline{dr})}\left({}^{(\alpha\tau)}\mathbf{AAV}_3\right) \\ &= \sum\nolimits_{k\in U} \tilde{y}_{\overline{r}k}^2 \gamma_{\overline{d}k} \left(\left[\sum\nolimits_{(dr)} {}^{(\alpha\tau\nu^*)}\phi_{(dr)} \left[\frac{{}^{(\alpha\tau\nu^*)}V_{(dr)}}{\ddot{V}_{(\overline{dr})} + {}^{(\alpha\tau)}\mathrm{AAV}_{3(\overline{dr})}} \right]^2 \frac{\left(\tilde{y}_{rk}^2 + \sigma_{rk}^2\right)\gamma_{dk}}{\overline{c}} \right]^{-1/2} - 1 \right). \end{aligned} \tag{A2.7}$$

Denoting with ${}^{(\alpha)}\mathbf{AAV}_3 = {}^{(\alpha\tau=0)}\mathbf{AAV}_3$, the system $\mathbf{j}$ may be expressed in a recursive form

$${}^{(\alpha+1)}\mathbf{AAV}_3 \cong \mathbf{j}\left(\mathbf{g}\left({}^{(\alpha\tau-1)}\mathbf{AAV}_3\right)\right) = \mathbf{j}\left(\mathbf{g}\left(\mathbf{g}\left(.....\mathbf{g}\left({}^{(\alpha\tau=0)}\mathbf{AAV}_3\right)\right)\right)\right) = \mathbf{f}\left({}^{(\alpha)}\mathbf{AAV}_3\right),$$

with $\mathbf{f}(\cdot) = \mathbf{j}(\mathbf{g}(\mathbf{g}(.....\mathbf{g}(\cdot))))$ as the system of $D \times R$ updating equations of ${}^{(\alpha+1)}\mathbf{AAV}_3$, with respect to the previous values of the OL, ${}^{(\alpha)}\mathbf{AAV}_3$. To demonstrate the convergence of OL, it is necessary to demonstrate that the Jacobian norm $\left\|J_{\mathbf{f}}\left(\mathbf{AAV}_3\right)\right\|$ is lower than 1. Using standard results of matrix algebra,

$$\left\|J_{\mathbf{f}}\left(\mathbf{AAV}_3\right)\right\| \le \left\|J_{\mathbf{j}}\left({}^{(\alpha\tau)}\mathbf{AAV}_3\right)\right\| \times \left\|J_{\mathbf{g}}\left({}^{(\alpha\tau-1)}\mathbf{AAV}_3\right)\right\| \times \ldots \times \left\|J_{\mathbf{g}}\left({}^{(\alpha\tau=0)}\mathbf{AAV}_3\right)\right\|,$$

in which the generic norm $\left\|J_{\mathbf{g}}(\cdot)\right\|$ is lesser than 1 (see the IL proof of convergence). Let $j'_{(\overline{dr})}$ be the $(\overline{dr})$ element on the diagonal of $J_{\mathbf{j}}\left({}^{(\alpha\tau)}\mathbf{AAV}_3\right)$. It is

$$\begin{aligned} j'_{(\overline{dr})} &= \sum\nolimits_{k\in U} \tilde{y}_{\overline{r}k}^2 \gamma_{\overline{d}k} \left[\sum\nolimits_{(dr)} {}^{(\alpha\tau\nu^*)}\phi_{(dr)} \frac{\left(\tilde{y}_{rk}^2 + \sigma_{rk}^2\right)\gamma_{dk}}{\overline{c}} \right]^{-3/2} \\ &\times {}^{(\alpha\tau\nu^*)}\phi_{(\overline{dr})} \frac{1}{{}^{(\alpha\tau\nu^*)}V_{(\overline{dr})}} \frac{\left(\tilde{y}_{\overline{r}k}^2 + \sigma_{\overline{r}k}^2\right)\gamma_{\overline{d}k}}{\overline{c}}. \end{aligned} \tag{A2.8}$$

Therefore, we have

$$\begin{aligned} j'_{(\overline{dr})} &\le \sum\nolimits_{k\in U} \tilde{y}_{\overline{r}k}^2 \gamma_{\overline{d}k} \left[{}^{(\alpha\tau\nu^*)}\phi_{(\overline{dr})} \frac{\left(\tilde{y}_{\overline{r}k}^2 + \sigma_{\overline{r}k}^2\right)\gamma_{\overline{d}k}}{\overline{c}} \right]^{-3/2} {}^{(\alpha\tau\nu^*)}\phi_{(\overline{dr})} \frac{1}{{}^{(\alpha\tau\nu^*)}V_{(\overline{dr})}} \frac{\left(\tilde{y}_{\overline{r}k}^2 + \sigma_{\overline{r}k}^2\right)\gamma_{\overline{d}k}}{\overline{c}} \\ &= \frac{1}{{}^{(\alpha\tau\nu^*)}V_{(\overline{dr})}} \sum\nolimits_{k\in U} \tilde{y}_{\overline{r}k}^2 \gamma_{\overline{d}k} \left[{}^{(\alpha\tau\nu^*)}\phi_{(\overline{dr})} \frac{\left(\tilde{y}_{\overline{r}k}^2 + \sigma_{\overline{r}k}^2\right)\gamma_{\overline{d}k}}{\overline{c}} \right]^{-1/2}. \end{aligned}$$

The following inequality holds

$$j'_{(\overline{dr})} < \frac{\sum_{k\in U} \tilde{y}_{\overline{r}k}\gamma_{\overline{d}k}}{\sqrt{\overline{c}\,{}^{(\alpha\tau\nu^*)}\phi_{(\overline{dr})}}\,{}^{(\alpha\tau\nu^*)}V_{(\overline{dr})}} << 1.$$

Consequently, the norm $\left\|J_{\mathbf{j}}\left({}^{(\alpha\tau)}\mathbf{AAV}_3\right)\right\| < 1$, and therefore the OL converges.

Appendix A3

Proof that the approximation of Remark 4.1 is upward

Since $\hat{u}_{(dr)k}$ is the weighted least square prediction of $u_{rk}\gamma_{dk}$, by using a different value of the $\hat{u}_{(dr)k}$, such as $\hat{u}_{(dr)k} = 0$, we obtain

$$\sum_{k\in U}(1/\pi_k - 1)\, E_M\left[(u_{rk}\gamma_{dk} - \hat{u}_{(dr)k})^2\right] \leq \sum_{k\in U}(1/\pi_k - 1)\, E_M\left[(u_{rk}\gamma_{dk} - 0)^2\right],$$

where $E_M\left[(u_{rk}\gamma_{dk} - 0)^2\right] = \sigma^2_{rk}\gamma_{dk}$. Replacing the terms $E_M\left[(u_{rk}\gamma_{dk} - \hat{u}_{(dr)k})^2\right]$ with $\sigma^2_{rk}\gamma_{dk}$ in expression (A1.5), the AAV (4.3) is inflated. The approximation $\hat{u}_{(dr)k} = 0$ implies that $b_{(dr)k}(\boldsymbol{\pi}) = c_{(dr)k}(\boldsymbol{\pi}) = 0$. Finally, we emphasize that in most cases, the upward is slight, since the $\hat{u}_{(dr)k}$ are obtained by the $\mathbf{z}_k$ variables that generally have a very low predictive power for the $u_{rk}\gamma_{dk}$ values (see Section 4). In these situations $\hat{u}_{(dr)k} \cong (1/N)\sum_{k\in U} u_{rk}\gamma_{dk} \cong 0$. So $E_M(u_{rk}\gamma_{dk}\hat{u}_{(dr)k}) \cong 0$ and $E_M(\hat{u}_{(dr)k})^2 \cong 0$.

Appendix A4

Proof of expression (4.7)

In this case, each $\boldsymbol{\delta}_k$ vector has $H - 1$ zero elements and 1 element equal to 1 (corresponding to the planned population to which the unit k belongs). Given the input values, the optimization procedure $\pi_k = \pi_h$ for $k \in U_h$. Under the above assumption, $[\mathbf{A}(\boldsymbol{\pi})]^{-1}$ is a diagonal matrix with the hh^{th} element given by $[\mathbf{A}_{hh}(\boldsymbol{\pi})]^{-1} = \left[N_h\pi_h^2(1/\pi_h - 1)\right]^{-1}$. Considering $\tilde{y}_{rk} = \bar{Y}_{rh}$, expressions (A1.2) and (A1.3) can be reformulated as, respectively,

$$\hat{\bar{y}}_{(dr)k} = \pi_h\boldsymbol{\delta}_k'[\mathbf{A}(\boldsymbol{\pi})]^{-1} N_h\pi_h(1/\pi_h - 1)\bar{Y}_{rh} = \bar{Y}_{rh}. \tag{A4.1}$$

$$\hat{u}_{(dr)k} = \pi_h\boldsymbol{\delta}_k'[\mathbf{A}(\boldsymbol{\pi})]^{-1}\pi_h(1/\pi_h - 1)\sum_{j\in U}u_{rj} = (\pi_h N_h)^{-1}\sum_{j\in U_h}u_{rj}, \tag{A4.2}$$

but $\sum_{j\in U_h}u_{rj} = 0$ as the sum of the residual of a regression model.

Using the formulae (A4.1) and (A4.2), expression (4.5) is given by

$$\begin{aligned}\text{AAV}(\hat{t}_{(dr)}) &= [N/(N-H)]\sum_h\left(\frac{1}{\pi_h} - 1\right)\sum_{k\in U_h}E_M(u_{rk}\gamma_{dk})^2 \\ &= [N/(N-H)]\sum_{d=1}^{D}\sum_{h\in H_d}\sigma^2_{rh}N_h(N_h/n_h - 1),\end{aligned}$$

since $\pi_h = n_h/N_h$ and expression (4.7) may be obtained.

References

Bethel, J. (1989). Sample allocation in multivariate surveys. *Survey Methodology*, 15, 1, 47-57.

Boyd, S., and Vanderberg, L. (2004). *Convex Optimization*. Cambridge University Press.

Breidt, F.J., and Chauvet, G. (2011). Improved variance estimation for balanced samples drawn via the cube method. *Journal of Statistical Planning and Inference*, 141, 479-487.

Chauvet, G., Bonnéry, D. and Deville, J.-C. (2011). Optimal inclusion probabilities for balanced sampling. *Journal of Statistical Planning and Inference*, 141, 984-994.

Choudhry, G.H., Rao, J.N.K. and Hidiroglou, M.A. (2012). On sample allocation for efficient domain estimation. *Survey Methodology*, 18, 1, 23-29.

Chromy, J. (1987). Design optimization with multiple objectives. *Proceedings of the Survey Research Methods Section,* American Statistical Association, 194-199.

Cochran, W.G. (1977). *Sampling Techniques*. New York: John Wiley & Sons, Inc.

Deville, J.-C., and Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91, 893-912.

Deville, J.-C., and Tillé, Y. (2005). Variance approximation under balanced sampling, *Journal of Statistical Planning and Inference*, 128, 569-591.

Dykstra R. and Wollan P. (1987). Finding I-projections subject to a finite set of linear inequality constraints, *Applied Statistics*, 36, 377-383.

Ernst, L.R. (1989). Further applications of linear programming to sampling problems. *Proceedings of the Survey Research Methods Section,* American Statistical Association, 625-631.

Falorsi, P.D., and Righi, P. (2008). A balanced sampling approach for multi-way stratification designs for small area estimation. *Survey Methodology*, 34, 2, 223-234.

Falorsi, P.D., Orsini, D. and Righi, P. (2006). Balanced and coordinated sampling designs for small domain estimation. *Statistics in Transition*, 7, 1173-1198.

Gonzalez, J.M., and Eltinge, J.L. (2010). Optimal survey design: A review. *Section on Survey Research Methods – JSM 2010*, October.

Isaki, C.T., and Fuller, W.A. (1982). Survey design under a regression superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

Khan, M.G.M., Mati, T. and Ahsan, M.J. (2010). An optimal multivariate stratified sampling design using auxiliary information: An integer solution using goal programming approach. *Journal of Official Statistics*, 26, 695-708.

Kokan, A., and Khan, S. (1967). Optimum allocation in multivariate surveys: An analytical solution. *Journal of the Royal Statistical Society, Series B*, 29, 115-125.

Lu, W., and Sitter, R.R. (2002). Multi-way stratification by linear programming made practical. *Survey Methodology*, 28, 2, 199-207.

Nedyalkova, D., and Tillé, Y. (2008). Optimal sampling and estimation strategies under the linear model. *Biometrika*, 95, 521-537.

Tillé, Y. (2006). *Sampling Algorithms*. Springer-Verlag, New York.

Tillé, Y., and Favre, A.-C. (2005). Optimal allocation in balanced sampling. *Statistics and Probability Letters*, 74, 31-37.

Winkler, W.E. (2001). Multi-way survey stratification and sampling. *Research Report Series*, Statistics #2001-01. Statistical Research Division U.S. Bureau of the Census Washington D.C. 20233.