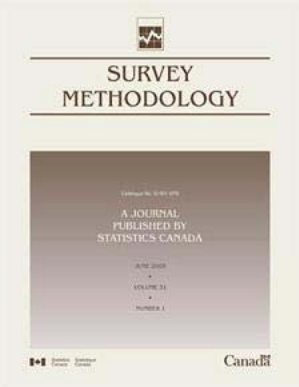
Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Calibration and restricted weights



by Alain Théberge

June 2000

Statistics Canada Statistique Canada

Canada

Survey Methodology, June 2000
Vol. 26, No. 1, pp. 99-107
Statistics Canada, Catalogue No. 12-001

Calibration and restricted weights

Alain Théberge [1]

Abstract

To better understand the impact of imposing a restricted region on calibration weights, the author reviews the latter's asymptotic behaviour. Necessary and sufficient conditions are provided for the existence of a solution to the calibration equation with weights within given intervals. A more general formulation of the calibration problem leads to a compromise between the need to satisfy the calibration equation and the attempt to obtain weights that are close to Horvitz-Thompson weights. If the requirements for the calibration equation are relaxed, then various estimation methods with restricted weights can be used. The estimators that are introduced usually have the same asymptotic properties as the calibration estimator with no weight restrictions, and some have weights which can be calculated explicitly, without any iterative process. The author shows how these estimators can be adapted to take advantage of a synthetic estimator. An approach similar to that used to restrict weights is applied to outliers.

Key Words: Small domains; Moore-Penrose inverse; Inequality solutions; Asymptotic properties; Outliers.

1. Introduction

The calibration estimator has good asymptotic properties. However, for samples of small size, or if calibration is done at the domain level and some of the domains involve few observations, the weights of such an estimator can include extreme values. One way of overcoming this problem consists in using the calibration method with distance measurements which restrict the weights of observations to certain intervals about the sampling weights. This approach was developed by Deville and Särndal (1992). Other methods aimed at providing robust estimates satisfying the calibration equation can be found in Duchesne (1999). That paper contains an extensive bibliography on robust estimators. However, there is no guaranteed solution to the calibration equation with restricted weights. Even when such weights exist, the statistician might prefer solving the problem of extreme weights by relaxing somewhat the requirements for the calibration equation, instead of tightening the constraints on the weights by using a distance measurement that is more "restrictive". This paper provides a formulation of the calibration problem which offers more flexibility to the statistician. The problem in fact is one of minimization similar to that encountered in ridge regression. Bardsley and Chambers (1984) encountered this same minimization problem in their search for model-based estimators. This formulation of the calibration problem can be used to restrict weights without the use of special distances between calibrated weights and Horvitz-Thompson weights. Rao and Singh (1997) combined this approach with iterative methods using distance measurements. Other ways of restricting weights will also be reviewed.

In the next section, the calibration method is outlined without applying limits to the values of weights. The calibration problem thus outlined does not assume there is a solution to the calibration equation. The asymptotic properties of calibrated weights are discussed. These properties have a bearing on the asymptotic behaviour of the estimators whose weights are restricted. In section 3, necessary and sufficient conditions are provided for the existence of restricted weights which satisfy the calibration equation. Section 4 discusses how the estimation problem can be formulated by varying the importance attributed to the calibration equation. Section 5 provides various means of restricting weights without recourse to a specific distance. Section 6 introduces an estimator with restricted weights which is useful for small domains. Finally, in section 7, outliers are discussed in terms of a method similar to that used to deal with extreme weights.

2. Calibration

Let $\boldsymbol{Y} \in \mathbb{R}^{N\times d}$ denote a matrix of d variables of interest for a population of size N, and let $\boldsymbol{c} \in \mathbb{R}^{N}$ denote a vector of known constants; a sample s of size n is drawn, and the sub-script s is used to designate the sub-vectors or sub-matrices corresponding to the sample. We wish to estimate $\boldsymbol{Y}'\boldsymbol{c}$ using $\boldsymbol{Y}_s'\boldsymbol{w}_s$, where $\boldsymbol{w}_s \in \mathbb{R}^{n}$ is a weight vector for the sampled units. For a vector $\boldsymbol{v}$ and a positive diagonal matrix $\boldsymbol{F}$ of identical dimension, we define $\|\boldsymbol{v}\|_{\boldsymbol{F}}^2 = \boldsymbol{v}'\boldsymbol{F}\boldsymbol{v}$. For an auxiliary information matrix $\boldsymbol{X} \in \mathbb{R}^{N\times p}$, $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ the diagonal matrix of sampling weights, given positive diagonal matrices $\boldsymbol{U}_s \in \mathbb{R}^{n\times n}$ and $\boldsymbol{T} \in \mathbb{R}^{p\times p}$, we seek, among the weight vectors $\boldsymbol{W}_s \in \mathbb{R}^{n}$ which minimize $\|\boldsymbol{X}_s'\boldsymbol{w}_s - \boldsymbol{X}'\boldsymbol{c}\|_{\boldsymbol{T}}^2$, the one which minimizes $D_s(\boldsymbol{w}_s) = \|\boldsymbol{w}_s - \boldsymbol{A}_s\boldsymbol{c}_s\|_{\boldsymbol{U}_s}^2$. This formulation of the problem, which does not assume the existence of weights satisfying the calibration equation, $\boldsymbol{X}_s'\boldsymbol{w}_s = \boldsymbol{X}'\boldsymbol{c}$, can be found in Théberge (1999). The

1. Alain Théberge, Social Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6, Canada.

solution represents the vector of calibrated weights $\boldsymbol{w}_{\text{cal}}$. We have

$$\boldsymbol{w}_{\text{cal}} = \boldsymbol{A}_s\boldsymbol{c}_s + \boldsymbol{U}_s^{-1}\boldsymbol{X}_s\boldsymbol{T}^{1/2}(\boldsymbol{T}^{1/2}\boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s\boldsymbol{T}^{1/2})^{\dagger} \boldsymbol{T}^{1/2}(\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s), \quad (1)$$

where $\boldsymbol{F}^{\dagger}$ denotes the Moore-Penrose inverse of the matrix $\boldsymbol{F}$.

To better review the asymptotic properties of calibration estimators with restricted weights, let us now examine the behaviour of $\boldsymbol{w}_{\text{cal}}$ when $n \to \infty$. We assume there exists an asymptotic setup in which the size of the population and the size of the sample tend towards infinity (see for example Isaki and Fuller (1982)), and for which we have

$$\begin{aligned} \boldsymbol{Y}'\boldsymbol{c} &= O_p(N^{\gamma})(\gamma \geq 0) \\ \boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s &= O_p(n^{-1/2}N^{\gamma}) \\ \boldsymbol{T}^{1/2}\boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s\boldsymbol{T}^{1/2} &= O_p(n). \end{aligned} \quad (2)$$

It follows that $(\boldsymbol{T}^{1/2}\boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s\boldsymbol{T}^{1/2})^{\dagger} = O_p(n^{-1})$, since one of the properties of the Moore-Penrose inverse of a matrix $\boldsymbol{F}$ is $\boldsymbol{F}^{\dagger}\boldsymbol{F}\boldsymbol{F}^{\dagger} = \boldsymbol{F}^{\dagger}$. Usually, we can expect to have $\gamma = 1$ when each element of the vector $\boldsymbol{c}$ has a value of 1 (estimate of a total), and $\gamma = 0$ when each element of $\boldsymbol{c}$ has a value of $1/N$ (estimate of a mean). For conditions (2) we therefore have,

$$\begin{aligned} \boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s\boldsymbol{c}_s &= \boldsymbol{U}_s^{-1}\boldsymbol{X}_s\boldsymbol{T}^{1/2}(\boldsymbol{T}^{1/2}\boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s\boldsymbol{T}^{1/2})^{\dagger} \\ &\quad \boldsymbol{T}^{1/2}(\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s) \\ &= O_p(n^{-1})O_p(n^{-1/2}N^{\gamma}) \\ &= O_p(n^{-3/2}N^{\gamma}). \end{aligned} \quad (3)$$

Thus $\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s\boldsymbol{c}_s$ converges in probability to $\boldsymbol{0}$, if

$$\lim_{n,N\to\infty} n^{-3/2}N^{\gamma} = 0.$$

For an asymptotic setup such as that of Brewer (1979) in which the sampling fraction n/N is constant, or any setup for which the sampling fraction converges to a positive number, this condition is verified if $\gamma < 3/2$.

Writing $\boldsymbol{w}_{\text{cal}} = \boldsymbol{A}_s\boldsymbol{c}_s + \boldsymbol{U}_s^{-1}\boldsymbol{X}_s\boldsymbol{T}^{1/2}\boldsymbol{H}_s^{\dagger}\boldsymbol{T}^{1/2}(\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s)$, where $\boldsymbol{H}_s = \boldsymbol{T}^{1/2}\boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s\boldsymbol{T}^{1/2}$, we have

$$\begin{aligned} D_s(\boldsymbol{w}_{\text{cal}}) &= (\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s)'\boldsymbol{T}^{1/2}\boldsymbol{H}_s^{\dagger}\boldsymbol{H}_s\boldsymbol{H}_s^{\dagger} \\ &\quad \boldsymbol{T}^{1/2}(\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s) \\ &= (\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s)'\boldsymbol{T}^{1/2}\boldsymbol{H}_s^{\dagger}\boldsymbol{T}^{1/2}(\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s) \\ &= O_p(n^{-1/2}N^{\gamma})O_p(n^{-1})O_p(n^{-1/2}N^{\gamma}) \\ &= O_p(n^{-2}N^{2\gamma}). \end{aligned} \quad (4)$$

Again for an asymptotic setup in which the sampling fraction converges to a positive number, we have $D_s(\boldsymbol{w}_{\text{cal}})$ converging in probability to 0, if $\gamma < 1$. Thus there are cases, *e.g.*, for the estimate of a total, where $\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s\boldsymbol{c}_s$ converges in probability to $\boldsymbol{0}$, but where $D_s(\boldsymbol{w}_{\text{cal}}) = \|\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s\boldsymbol{c}_s\|_{U_s}^2$ does not converge to 0.

3. Calibration equation solutions and restricted weights

Even in the absence of weight restrictions, there might not be a solution to the calibration equation. By applying Graybill (1983, 113) to the calibration problem, we find that the calibration equation $\boldsymbol{X}_s'\boldsymbol{w}_s = \boldsymbol{X}'\boldsymbol{c}$ can be solved if and only if $(\boldsymbol{X}_s^{\dagger}\boldsymbol{X}_s)'\boldsymbol{X}'\boldsymbol{c} = \boldsymbol{X}'\boldsymbol{c}$. If there is a solution, the calibrated weights might be negative or exceptionally large. Deville and Särndal (1992) proposed using various distance measures other than a weighted sum of squares to measure the distance between Horvitz-Thompson weights and calibrated weights, so as to restrict the weights to certain intervals while satisfying the calibration equation. This approach can only work if there are in these intervals weights which satisfy the calibration equation. The main goal of this section is to find necessary and sufficient conditions for the existence of a weight vector $\boldsymbol{w}_s$ within given bounds, such that the estimates of totals for auxiliary variables are also bounded. In other words, we are seeking conditions for the existence of a vector $\boldsymbol{w}_s$ such that $\boldsymbol{w}^{(L)} \leq \boldsymbol{w}_s \leq \boldsymbol{w}^{(H)}$ and $\boldsymbol{t}^{(L)} \leq \boldsymbol{X}_s'\boldsymbol{w}_s \leq \boldsymbol{t}^{(H)}$, where $\boldsymbol{w}^{(L)}$, $\boldsymbol{w}^{(H)}$, $\boldsymbol{t}^{(L)}$ and $\boldsymbol{t}^{(H)}$ are given. In particular, by assuming $\boldsymbol{t}^{(L)} = \boldsymbol{t}^{(H)} = \boldsymbol{X}'\boldsymbol{c}$, we would obtain conditions for the existence of weights restricted to the intervals $\boldsymbol{w}^{(L)} \leq \boldsymbol{w}_s \leq \boldsymbol{w}^{(H)}$, satisfying the calibration equation.

A first step is provided by the following Fan (1956) theorem. It is formulated here for a matrix $\boldsymbol{M}$ of finite dimension, although the proof provided by Fan also applies to a matrix of infinite dimension. The theorem uses the kernel of $\boldsymbol{M}'$, $N(\boldsymbol{M}')$, defined as the set of vectors $\boldsymbol{\alpha}$ such that $\boldsymbol{M}'\boldsymbol{\alpha} = \boldsymbol{0}$.

Theorem: Let $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{l} \in \mathbb{R}^m$, $\exists \boldsymbol{w} \in \mathbb{R}^n$ such that $\boldsymbol{M}\boldsymbol{w} \geq \boldsymbol{l}$ if and only if for any $\boldsymbol{\lambda} \geq \boldsymbol{0}$ in $\boldsymbol{N}(\boldsymbol{M}')$, we have $\boldsymbol{l}'\boldsymbol{\lambda} \leq 0$.

Corollary: Let $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{l}, \boldsymbol{h} \in \mathbb{R}^m$, $\exists \boldsymbol{w} \in \mathbb{R}^n$ such that $\boldsymbol{l} \leq \boldsymbol{M}\boldsymbol{w} \leq \boldsymbol{h}$ if and only if first $\boldsymbol{l} \leq \boldsymbol{h}$ and secondly $\boldsymbol{\lambda} \in N(\boldsymbol{M}') \Rightarrow -\boldsymbol{l}'\boldsymbol{\lambda}_- \leq \boldsymbol{h}'\boldsymbol{\lambda}_+$, where $\boldsymbol{\lambda}_+ = \max(\boldsymbol{\lambda}, \boldsymbol{0})$ and $\boldsymbol{\lambda}_- = \min(\boldsymbol{\lambda}, \boldsymbol{0})$ with the extrema taken elementwise.

The corollary is obtained by using the theorem with

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{M} \\ -\boldsymbol{M} \end{pmatrix}, \ \boldsymbol{l} = \begin{pmatrix} \boldsymbol{l} \\ -\boldsymbol{h} \end{pmatrix} \text{ and } \boldsymbol{\lambda} = \begin{pmatrix} -\boldsymbol{\lambda}_- \\ \boldsymbol{\lambda}_+ \end{pmatrix}$$

Let p denote the dimension of $N(\boldsymbol{M}')$. If p is equal to zero, then $\boldsymbol{\lambda} \in N(\boldsymbol{M}')$ implies $\boldsymbol{\lambda} = \boldsymbol{0}$, and the condition of the theorem (or the similar condition of the corollary) is obviously met. If p is equal to one, then $\boldsymbol{\lambda} \in N(\boldsymbol{M}')$ implies that $\boldsymbol{\lambda}$ is a multiple of a vector $\boldsymbol{z}$, and it is sufficient to

check the condition for $\boldsymbol{\lambda} = \boldsymbol{z}$ and $\boldsymbol{\lambda} = -\boldsymbol{z}$. If we use the property $(-\boldsymbol{\lambda})_- = -(\boldsymbol{\lambda})_+$, the problem outlined at the beginning of the section can now be resolved if $\boldsymbol{X}_s$ is a vector. The corollary with

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{I}_n \\ \boldsymbol{X}_s' \end{pmatrix}, \ \boldsymbol{l} = \begin{pmatrix} \boldsymbol{w}^{(L)} \\ \boldsymbol{t}^{(L)} \end{pmatrix}, \ \boldsymbol{h} = \begin{pmatrix} \boldsymbol{w}^{(H)} \\ \boldsymbol{t}^{(H)} \end{pmatrix},$$

and the fact that

$$\boldsymbol{z} = \begin{pmatrix} -\boldsymbol{X}_s \\ 1 \end{pmatrix}$$

spans $N(\boldsymbol{M}')$, provide the necessary and sufficient conditions

$$\begin{aligned} \boldsymbol{w}^{(L)} &\le \boldsymbol{w}^{(H)} \\ \boldsymbol{t}^{(L)} &\le \boldsymbol{t}^{(H)} \\ (\boldsymbol{X}_s)_+' \boldsymbol{w}^{(L)} + (\boldsymbol{X}_s)_-' \boldsymbol{w}^{(H)} &\le \boldsymbol{t}^{(H)} \\ \boldsymbol{t}^{(L)} &\le (\boldsymbol{X}_s)_+' \boldsymbol{w}^{(H)} + (\boldsymbol{X}_s)_-' \boldsymbol{w}^{(L)}. \end{aligned} \tag{5}$$

The third inequality in (5) states that the weighted total of the auxiliary variable must not be greater than $\boldsymbol{t}^{(H)}$, when the smallest possible weight $\boldsymbol{w}^{(L)}$ is given to units for which the auxiliary variable is positive, and when the greatest possible weight $\boldsymbol{w}^{(H)}$ is given to units for which the auxiliary variable is negative. The fourth inequality in (5) states that the weighted total of the auxiliary variable must not be less than $\boldsymbol{t}^{(L)}$, when the largest possible weight is given to units for which the auxiliary variable is positive, and when the smallest possible weight is given to units for which the auxiliary variable is negative.

Even for $p > 1$, it is sufficient to check the condition of the corollary for a finite number of values of $\boldsymbol{\lambda}$. Let $\boldsymbol{V} \in \mathbb{R}^{m \times p}$, $2 \le p \le m$ denote a matrix whose columns form a basis for $N(\boldsymbol{M}')$. It is always possible to construct $\boldsymbol{V}$ such that p of its rows, $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_m$, are the unit vectors of $\mathbb{R}^p$, and we will assume that $\boldsymbol{V}$ is of this form. It will be shown in Appendix A that it is sufficient to check the condition of the corollary for vectors $\boldsymbol{\lambda} = \boldsymbol{V}\boldsymbol{\varphi}$ and $\boldsymbol{\lambda} = -\boldsymbol{V}\boldsymbol{\varphi}$, where $\boldsymbol{\varphi} = (\varphi_1, ..., \varphi_p)'$ is a non-zero vector satisfying $\boldsymbol{v}_i'\boldsymbol{\varphi} = 0$ for a subset of $(p-1)$ linearly independent vectors $\boldsymbol{v}_i$. We must therefore check the condition at the most for $\binom{m}{p-1}$ vectors $\boldsymbol{\varphi}$, *i.e.*, at the most $2\binom{m}{p-1}$ values of $\boldsymbol{\lambda}$.

Using the corollary with

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{I}_n \\ \boldsymbol{X}_s' \end{pmatrix}, \ \boldsymbol{l} = \begin{pmatrix} \boldsymbol{w}^{(L)} \\ \boldsymbol{t}^{(L)} \end{pmatrix}, \ \boldsymbol{h} = \begin{pmatrix} \boldsymbol{w}^{(H)} \\ \boldsymbol{t}^{(H)} \end{pmatrix},$$

and noting that the columns of

$$\boldsymbol{V} = \begin{pmatrix} -\boldsymbol{X}_s \\ \boldsymbol{I}_p \end{pmatrix}$$

form a basis for $N(\boldsymbol{M}')$, we obtain the following necessary and sufficient conditions for the existence of a solution to the problem mentioned at the beginning of this section when ever $\boldsymbol{X}_s \in \mathbb{R}^{n \times p}$ with $p > 1$. We must have $\boldsymbol{w}^{(L)} \le \boldsymbol{w}^{(H)}$, $\boldsymbol{t}^{(L)} \le \boldsymbol{t}^{(H)}$, and for each subset of $(p-1)$ linearly independent rows of

$$\boldsymbol{V} = \begin{pmatrix} -\boldsymbol{X}_s \\ \boldsymbol{I}_p \end{pmatrix}$$

it is necessary that

$$(\boldsymbol{X}_s\boldsymbol{\varphi})_+' \boldsymbol{w}^{(L)} - \boldsymbol{\varphi}_-' \boldsymbol{t}^{(L)} \le -(\boldsymbol{X}_s\boldsymbol{\varphi})_-' \boldsymbol{w}^{(H)} + \boldsymbol{\varphi}_+' \boldsymbol{t}^{(H)}$$

$$-(\boldsymbol{X}_s\boldsymbol{\varphi})_-' \boldsymbol{w}^{(L)} + \boldsymbol{\varphi}_-' \boldsymbol{t}^{(L)} \le (\boldsymbol{X}_s\boldsymbol{\varphi})_+' \boldsymbol{w}^{(H)} - \boldsymbol{\varphi}_-' \boldsymbol{t}^{(H)} \tag{6}$$

for a non-zero vector $\boldsymbol{\varphi} \in \mathbb{R}^p$ orthogonal to each row of the subset. The second inequality in (6) is obtained from the first by changing the sign of $\boldsymbol{\varphi}$.

If $\boldsymbol{V}_{\text{sub}} \in \mathbb{R}^{p \times p}$ is a non-singular matrix whose rows are rows of $\boldsymbol{V}$, then each column of $\boldsymbol{V}_{\text{sub}}^{-1}$ is a vector perpendicular to $(p-1)$ linearly independent rows of $\boldsymbol{V}$. Hence the following result:

There exists a weight vector $\boldsymbol{w}_s$ such that $\boldsymbol{w}^{(L)} \le \boldsymbol{w}_s \le \boldsymbol{w}^{(H)}$ and $\boldsymbol{t}^{(L)} \le \boldsymbol{X}_s'\boldsymbol{w}_s \le \boldsymbol{t}^{(H)}$ if and only if $\boldsymbol{w}^{(L)} \le \boldsymbol{w}^{(H)}$, $\boldsymbol{t}^{(L)} \le \boldsymbol{t}^{(H)}$ and

$$\begin{aligned} &(\boldsymbol{X}_s\boldsymbol{V}_{\text{sub}}^{-1})_+' \boldsymbol{w}^{(L)} - (\boldsymbol{V}_{\text{sub}}^{-1})_-' \boldsymbol{t}^{(L)} \le \\ &\qquad -(\boldsymbol{X}_s\boldsymbol{V}_{\text{sub}}^{-1})_-' \boldsymbol{w}^{(H)} + (\boldsymbol{V}_{\text{sub}}^{-1})_+' \boldsymbol{t}^{(H)} \\ &-(\boldsymbol{X}_s\boldsymbol{V}_{\text{sub}}^{-1})_-' \boldsymbol{w}^{(L)} + (\boldsymbol{V}_{\text{sub}}^{-1})_+' \boldsymbol{t}^{(L)} \le \\ &\qquad (\boldsymbol{X}_s\boldsymbol{V}_{\text{sub}}^{-1})_+' \boldsymbol{w}^{(H)} - (\boldsymbol{V}_{\text{sub}}^{-1})_-' \boldsymbol{t}^{(H)} \end{aligned} \tag{7}$$

for all non-singular matrixes $\boldsymbol{V}_{\text{sub}} \in \mathbb{R}^{p \times p}$ whose rows are rows of

$$\boldsymbol{V} = \begin{pmatrix} -\boldsymbol{X}_s \\ \boldsymbol{I}_p \end{pmatrix}.$$

These conditions are somewhat redundant. For example, if inequalities (7) are met for $\boldsymbol{V}_{\text{sub}} = \boldsymbol{V}_1$, then they are necessarily met for any matrix $\boldsymbol{V}_2$ obtained from $\boldsymbol{V}_1$ through a permutation of rows.

Another example is provided by weighting observations in a contingency table. Assuming $\hat{N}_{ij} = n_{ij} w_{ij}$ $(i = 1, 2, ..., R;\ j = 1, 2, ..., C)$, where n_{ij} is the number of observations in cell (i, j) of a contingency table and w_{ij} is the weight of each of these observations, we wish to know if there are weights w_{ij} such that $\hat{N}_{ij}$ satisfies certain constraints. For example, motivated by the problem of convergence of the raking ratio procedure, Bacharach (1965) provided necessary and sufficient conditions for the existence of weights w_{ij} such that $\hat{N}_{ij} \ge 0$, $\sum_{i=1}^{R} \hat{N}_{ij} = N_{.j}$ $(j = 1, ..., C)$, $\sum_{j=1}^{C} \hat{N}_{ij} = N_{i.}$ $(i = 1, ..., R)$, where the values of $N_{.j}$ and $N_{i.}$ are given. The following result, demonstrated in Appendix B, is more general.

For arbitrary constants $N_{ij}^{(L)}, N_{ij}^{(H)}, N_{.j}^{(L)}, N_{.j}^{(H)}, N_{i.}^{(L)}, N_{i.}^{(H)}, N_{..}^{(L)}$, and $N_{..}^{(H)}$, there are $\hat{N}_{ij}$ such that

$$\begin{aligned} &N_{ij}^{(L)} \le \hat{N}_{ij} \le N_{ij}^{(H)} && i = 1, ..., R;\ j = 1, ..., C; \\ &N_{.j}^{(L)} \le \sum_{i=1}^{R} \hat{N}_{ij} \le N_{.j}^{(H)} && j = 1, ..., C; \\ &N_{i.}^{(L)} \le \sum_{j=1}^{C} \hat{N}_{ij} \le N_{i.}^{(H)} && i = 1, ..., R; \\ &N_{..}^{(L)} \le \sum_{i=1}^{R}\sum_{j=1}^{C} \hat{N}_{ij} \le N_{..}^{(H)}, \end{aligned} \tag{8}$$

if and only if

$$\begin{aligned} &N_{ij}^{(L)} \le N_{ij}^{(H)} && i = 1, ..., R; j = 1, ..., C; \\ &N_{.j}^{(L)} \le N_{.j}^{(H)} && j = 1, ..., C; \\ &N_{i.}^{(L)} \le N_{i.}^{(H)} && i = 1, ..., R; \\ &N_{..}^{(L)} \le N_{..}^{(H)} \end{aligned} \tag{9}$$

$$\sum_{j\in T}\left(N_{.j}^{(L)} - \sum_{i\notin S} N_{ij}^{(H)}\right) \le \sum_{i\in S}\left(N_{i.}^{(H)} - \sum_{j\notin T} N_{ij}^{(L)}\right)$$

$$\sum_{i\in S}\left(N_{i.}^{(L)} - \sum_{j\notin T} N_{ij}^{(H)}\right) \le \sum_{j\in T}\left(N_{.j}^{(H)} - \sum_{i\notin T} N_{.j}^{(L)}\right)$$

$$N_{..}^{(L)} + \sum_{j\notin T}\left(N_{.j}^{(H)} - \sum_{i\notin S} N_{ij}^{(H)}\right) \le \sum_{i\in S}\left(N_{i.}^{(H)} - \sum_{j\notin T} N_{ij}^{(L)}\right) + \sum_{j=1}^{J} N_{.j}^{(H)}$$

$$\sum_{i=1}^{I} N_{i.}^{(L)} + \sum_{j\notin T}\left(N_{.j}^{(L)} - \sum_{i\notin S} N_{ij}^{(H)}\right) \le \sum_{i\in S}\left(N_{i.}^{(L)} - \sum_{j\notin T} N_{ij}^{(L)}\right) + N_{..}^{(H)}$$

for any $S \subseteq \{1, 2, ..., R\}$, $T \subseteq \{1, 2, ..., C\}$

The number of inequalities to be checked can be reduced. For example, instead of checking

$$\sum_{j\in T}\left(N_{.j}^{(L)} - \sum_{i\notin S} N_{ij}^{(H)}\right) \le \sum_{i\in S}\left(N_{i.}^{(H)} - \sum_{j\notin T} N_{ij}^{(L)}\right)$$

for any $S \subseteq \{1, 2, ..., R\}$, and $T \subseteq \{1, 2, ..., C\}$, it can be readily shown that an equivalent procedure would be to check that

$$\sum_{j\in T} N_{.j}^{(L)} \le \sum_{i=1}^{R} \min\left(\left(N_{i.}^{(H)} - \sum_{j\notin T} N_{ij}^{(L)}\right), \sum_{j\in T} N_{ij}^{(H)}\right)$$

for any $T \subseteq \{1, 2, ..., C\}$.

4. Mitigated calibration

There may be dissatisfaction with the two-step approach of calibration, where an attempt is first made to find weight vectors that best satisfy the calibration equation, and then from this set of vectors to find the one which comes closest to Horvitz-Thompson weights. For small samples, this method may lead to weights which the statistician will find too far from Horvitz-Thompson weights.

There may be a preference for varying the importance attributed to the calibration equation relative to the norm of $\boldsymbol{w}_s - \boldsymbol{A}_s\boldsymbol{c}_s$. Thus, there may be a desire to find a weight vector $\boldsymbol{w}_s$ which minimizes

$$\left\| \begin{pmatrix} \boldsymbol{w}_s - \boldsymbol{A}_s\boldsymbol{c}_s \\ \boldsymbol{X}_s'\boldsymbol{w}_s - \boldsymbol{X}'\boldsymbol{c} \end{pmatrix} \right\|_{\boldsymbol{V}}^2,$$

where

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{U}_s & \boldsymbol{0} \\ \boldsymbol{0} & \alpha\boldsymbol{T} \end{pmatrix}$$

and $\alpha \ge 0$. We then minimize

$$\|\boldsymbol{w}_s - \boldsymbol{A}_s\boldsymbol{c}_s\|_{\boldsymbol{U}_s}^2 + \alpha\|\boldsymbol{X}_s'\boldsymbol{w}_s - \boldsymbol{X}'\boldsymbol{c}\|_{\boldsymbol{T}}^2 = D_s(\boldsymbol{w}_s) + \alpha\|\boldsymbol{X}_s'\boldsymbol{w}_s - \boldsymbol{X}'\boldsymbol{c}\|_{\boldsymbol{T}}^2.$$

A similar minimization problem is encountered with ridge regression. For $\alpha = 0$ the solution is provided by Horvitz-Thompson weights $\boldsymbol{w}_s = \boldsymbol{A}_s\boldsymbol{c}_s$. For $\alpha > 0$, we seek $\boldsymbol{w}_s(\alpha)$ minimizing $||\boldsymbol{K}(\boldsymbol{w}_s - \boldsymbol{A}_s\boldsymbol{c}_s) - \boldsymbol{b}||_{\boldsymbol{V}}^2$, where $\boldsymbol{K} = (\boldsymbol{I}_n, \boldsymbol{X}_s)'$, $\boldsymbol{b} = (\boldsymbol{0}_{1\times n}, (\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s)')'$ and $\boldsymbol{0}_{1\times n} \in \mathbb{R}^n$ is a row vector of zeros. Ben-Israel and Greville (1980) yields

$$\boldsymbol{w}_s(\alpha) - \boldsymbol{A}_s\boldsymbol{c}_s = (\boldsymbol{K}'\boldsymbol{V}\boldsymbol{K})^{-1}\boldsymbol{K}'\boldsymbol{V}\boldsymbol{b}. \tag{10}$$

Thus by substituting the values of $\boldsymbol{K}$, $\boldsymbol{V}$, and $\boldsymbol{b}$ we obtain

$$\boldsymbol{w}_s(\alpha) = \boldsymbol{A}_s\boldsymbol{c}_s + \alpha(\boldsymbol{U}_s + \alpha\boldsymbol{X}_s\boldsymbol{T}\boldsymbol{X}_s')^{-1}\boldsymbol{X}_s\boldsymbol{T}(\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s). \tag{11}$$

It is easily shown that

$$\alpha(\boldsymbol{U}_s + \alpha\boldsymbol{X}_s\boldsymbol{T}\boldsymbol{X}_s')^{-1}\boldsymbol{X}_s\boldsymbol{T} = \boldsymbol{U}_s^{-1}\boldsymbol{X}_s(\alpha^{-1}\boldsymbol{T}^{-1} + \boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s)^{-1},$$

hence

$$\boldsymbol{w}_s(\alpha) = \boldsymbol{A}_s\boldsymbol{c}_s + \boldsymbol{U}_s^{-1}\boldsymbol{X}_s(\alpha^{-1}\boldsymbol{T}^{-1} + \boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s)^{-1}(\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s). \tag{12}$$

The estimator $\boldsymbol{Y}_s'\boldsymbol{w}_s(\alpha)$ thus becomes $\hat{\boldsymbol{Y}}'\boldsymbol{c} + (\boldsymbol{Y}_s - \hat{\boldsymbol{Y}}_s)'\boldsymbol{A}_s\boldsymbol{c}_s$, where $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_s(\alpha)$ and

$$\hat{\boldsymbol{\beta}}_s(\alpha) = (\boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s + \alpha^{-1}\boldsymbol{T}^{-1})^{-1}\boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{Y}_s.$$

The vector of regression coefficients, then, is the one obtained with ridge regression. Just as the calibration method, and the generalized regression method described by Särndal, Swensson and Wretman (1992), lead to the same estimators, a similar parallel can be drawn between mitigated calibration and ridge regression.

On the basis of equation (12), we can also use Ben-Israel and Greville (1980), and the fact that $\boldsymbol{F}^{\dagger} = \boldsymbol{F}'(\boldsymbol{F}\boldsymbol{F}')^{\dagger}$ with $\boldsymbol{F} = \boldsymbol{T}^{1/2}\boldsymbol{X}_s'\boldsymbol{U}_s^{-1/2}$, to show that

$$\lim_{\alpha\to\infty} \boldsymbol{w}_s(\alpha) = \boldsymbol{w}_{\text{cal}}.$$

This result was to be expected, since finding the vector $\boldsymbol{w}_s(\alpha)$ which minimizes $D_s(\boldsymbol{w}_s) + \alpha\|\boldsymbol{X}_s'\boldsymbol{w}_s - \boldsymbol{X}'\boldsymbol{c}\|_{\boldsymbol{T}}^2$ when $\alpha \to \infty$ is equivalent to finding the weight vector which minimizes $D_s(\boldsymbol{w}_s)$ among those which minimize $\|\boldsymbol{X}_s'\boldsymbol{w}_s - \boldsymbol{X}'\boldsymbol{c}\|_{\boldsymbol{T}}^2$.

Rao and Singh (1997) defined tolerances for each of the p constraints of the calibration equation, and they established a relationship between these tolerances and the matrix $\alpha\boldsymbol{T}$.

For $\alpha \in [0,\infty[$ the function $\boldsymbol{w}_s(\alpha)$ is represented by a curve in $\mathbb{R}^n$ which links point $\boldsymbol{A}_s\boldsymbol{c}_s$ to point $\boldsymbol{w}_{\text{cal}}$. If $p = 1$, *i.e.* if $\boldsymbol{X}$ is a vector, this curve is a line segment. In fact, in this case the matrix $(\alpha^{-1}\boldsymbol{T}^{-1} + \boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s)^{-1}$ and the vector $\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s$ are scalars, and the weights $\boldsymbol{w}_s(\alpha)$ given by (12) are therefore equal to Horvitz-Thompson weights plus a multiple of vector $\boldsymbol{U}_s^{-1}\boldsymbol{X}_s$. And again for $p = 1$, we have

$$\lim_{\alpha\to\infty} \boldsymbol{w}_s(\alpha) = \boldsymbol{w}_{\text{cal}} = \boldsymbol{A}_s\boldsymbol{c}_s + [(\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s)/(\boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s)]\,\boldsymbol{U}_s^{-1}\boldsymbol{X}_s$$

which leads to the estimator

$$\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}} = \boldsymbol{Y}_s'\boldsymbol{A}_s\boldsymbol{c}_s + [(\boldsymbol{Y}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s)/(\boldsymbol{X}_s'\boldsymbol{U}_s^{-1}\boldsymbol{X}_s)](\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s)$$

Taking $\boldsymbol{U} = \boldsymbol{A}^{-1}\,\text{diag}(\boldsymbol{X})$, we obtain the ratio estimator

$$\boldsymbol{Y}_s'\boldsymbol{A}_s\boldsymbol{c}_s + [(\boldsymbol{Y}_s'\boldsymbol{A}_s\boldsymbol{1}_{n\times 1})/(\boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{1}_{n\times 1})](\boldsymbol{X}'\boldsymbol{c} - \boldsymbol{X}_s'\boldsymbol{A}_s\boldsymbol{c}_s),$$

where $\boldsymbol{1}_{a\times b} \in \mathbb{R}^{a\times b}$ is a matrix of ones.

Ben-Israel and Greville (1980, 111, exercise 15) showed that $D_s(\boldsymbol{w}_s(\alpha))$ is an increasing monotonic function of α. Note however that for a unit $k \in s$, $|w_k(\alpha) - a_k c_k|$ is not necessarily a monotonic function of α. As α increases, the weight vector $\boldsymbol{w}_s(\alpha)$ moves away from the Horvitz-Thompson weight vector, but this does not necessarily apply to each coordinate of the vector.

In this article, mitigated calibration is used to restrict weights, *i.e.* when the size of the sample is relatively small. It can easily be shown, however, that for an asymptotic setup satisfying (2) and for which $\hat{\boldsymbol{\beta}}_s(\alpha) - \boldsymbol{\beta}(\alpha) \to \boldsymbol{0}$ in probability, with

$$\boldsymbol{\beta}(\alpha) = (\boldsymbol{X}'\boldsymbol{U}^{-1}\boldsymbol{X} + \alpha^{-1}\boldsymbol{T}^{-1})^{-1}\boldsymbol{X}'\boldsymbol{U}^{-1}\boldsymbol{Y},$$

we have $\boldsymbol{Y}_s'\boldsymbol{w}_s(\alpha)$ is an asymptotically unbiased estimator whose asymptotic variance is

$$(\boldsymbol{Y} - \boldsymbol{Y}^*)'\,\text{diag}(\boldsymbol{c})(\boldsymbol{A}\prod\boldsymbol{A} - \boldsymbol{1}_{N\times N})\,\text{diag}(\boldsymbol{c})(\boldsymbol{Y} - \boldsymbol{Y}^*),$$

where $\boldsymbol{Y}^* = \boldsymbol{X}\boldsymbol{\beta}(\alpha)$, $\prod$ is the matrix of inclusion probabilities of order 2, and $\text{diag}(\boldsymbol{c})$ is the diagonal matrix formed from vector $\boldsymbol{c}$.

5. Estimation methods with restricted weights

In order to avoid obtaining weights having extreme values, we may wish to force the weight vector to be within a given region. This restricted region will be assumed to be convex and closed, and $\boldsymbol{A}_s\boldsymbol{c}_s$ will be assumed to be a point in this region. For example, if $\boldsymbol{w}^{(L)} < \boldsymbol{A}_s\boldsymbol{c}_s < \boldsymbol{w}^{(H)}$, we may wish to restrict the weights to region $R_{\boldsymbol{w}} = \{\boldsymbol{w}_s: \boldsymbol{w}^{(L)} \le \boldsymbol{w}_s \le \boldsymbol{w}^{(H)}\}$. We will assume that

$$\lim_{n\to\infty} \boldsymbol{w}^{(L)} - \boldsymbol{A}_s\boldsymbol{c}_s < \boldsymbol{0} \text{ and } \lim_{n\to\infty} \boldsymbol{w}^{(H)} - \boldsymbol{A}_s\boldsymbol{c}_s > \boldsymbol{0}.$$

The approach described in section 3 consists in selecting a distance measure between calibrated weights and Horvitz-Thompson weights which will provide weights that satisfy the calibration equation and which lie in the restricted region, should such weights in fact exist. The approach dealt with in this section is to temperate the requirement that the calibration equation be satisfied when the vector of calibration weights $\boldsymbol{w}_{\text{cal}}$ is outside the restricted region. Various means to temperate this requirement lead to different weighting methods.

When $\boldsymbol{w}_{\text{cal}}$ lies outside the restricted region, we could for example look for those points on the curve $\boldsymbol{w}_s(\alpha)$ parametered by $\alpha \ge 0$ which are on the border of this region. One property of these points is that they solve the minimization problem described in section 4 for corresponding values of α, thus through matrix $\boldsymbol{T}$, the importance of each calibration equation can be weighted. Using the example of the restricted region provided above, if

$$\boldsymbol{w}_{\text{cal}} = \lim_{\alpha\to\infty} \boldsymbol{w}_s(\alpha)$$

lies within this region, then $\boldsymbol{w}_{\text{res1}} = \boldsymbol{w}_{\text{cal}}$ can be used as a restricted weight vector, otherwise $\boldsymbol{w}_{\text{res1}} = \boldsymbol{w}_s(\alpha)$ with

$\alpha < \infty$ can be chosen such that $\boldsymbol{w}_s(\alpha)$ is on the boundary of the restricted region. If the asymptotic setup is such that conditions (2) are met with $\gamma < 3/2$ then for n sufficiently large, the probability that $\boldsymbol{w}_{\text{cal}}$ will be within the restricted region is equal to one. In fact, we have $\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s \boldsymbol{c}_s$ converging in probability to $\boldsymbol{0}$. The asymptotic properties of the estimator using the restricted weights, $\boldsymbol{w}_{\text{res1}}$, are therefore identical to those of the calibration estimator. It is worth noting that since $|w_k - a_k c_k|$ is not necessarily a monotonic function of α, it is possible for $\boldsymbol{w}_s(\alpha)$ to be on the boundary of the restricted region for several values of α, even if the restricted region is convex. Finding all these values is not necessarily a simple matter, and a decision has to be taken as to which value to use.

Another option for restricting weights would be to use as a restricted region those weights $\boldsymbol{w}_s$ which satisfy $D_s(\boldsymbol{w}_s) \le l$ for a bound $l > 0$. Then $\boldsymbol{w}_{\text{res2}} = \boldsymbol{w}_{\text{cal}}$ is taken as a restricted weight vector if $\boldsymbol{w}_{\text{cal}}$ lies in the restricted region, otherwise we seek $\alpha > 0$ such that $D_s(\boldsymbol{w}_s(\alpha)) = l$. This value of α is unique and can be found through iteration. Next we calculate the weights $\boldsymbol{w}_{\text{res2}} = \boldsymbol{w}_s(\alpha)$ which correspond to this value of α using equation (12). If the asymptotic setup is such that conditions (2) are met with $\gamma < 1$, and if l does not vary with n, then for n sufficiently large, the probability that $\boldsymbol{w}_{\text{cal}}$ will be within the restricted region is equal to one. In fact, we have $D_s(\boldsymbol{w}_{\text{cal}})$ converging in probability to 0. The asymptotic properties of the estimator using restricted weights, $\boldsymbol{w}_{\text{res2}}$, are then identical to those of the calibration estimator. Unfortunately, when estimating a total, we must expect to have $\gamma = 1$. In order to overcome this snag, we can use $l\sqrt{n}$ as a bound, instead of l. We can justify this bound on the basis that the length of the main diagonal of a hypercube of $\mathbb{R}^n$ is equal to the diameter of the sphere which circumscribes this hypercube, whereas the diameter of the sphere inscribed in this same hypercube is smaller by a factor of $\sqrt{n}$. The fact remains that a statistician might be uncomfortable using an asymptotic setup where the bound increases with the size of the sample. Furthermore, with this approach, the weights of the observations cannot be limited individually. Only the distance between the restricted weight vector and the Horvitz-Thompson weight vector is controlled.

With the methods described above, we look for those points on curve $\boldsymbol{w}_s(\alpha)$ which lie on the boundary of the restricted region. The value of α for which $\boldsymbol{w}_s(\alpha)$ lies on the boundary of the restricted region must often be found iteratively. It would be simpler to replace the curve $\boldsymbol{w}_s(\alpha)$ by the line segment linking $\boldsymbol{A}_s \boldsymbol{c}_s$ to $\boldsymbol{w}_{\text{cal}}$. For the restricted region $R_{\boldsymbol{w}}$, the restricted weight vector would be $\boldsymbol{w}_{\text{res3}} = \boldsymbol{w}_{\text{cal}}$ if $\boldsymbol{w}_{\text{cal}}$ is in the restricted region, otherwise $\boldsymbol{w}_{\text{res3}}$ would be equal to the point at which the line segment crosses the boundary of restricted region, *i.e.*

$$\boldsymbol{w}_{\text{res3}} = \boldsymbol{A}_s \boldsymbol{c}_s + \xi(\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s \boldsymbol{c}_s),$$

where

$$\xi = \min_{\text{k}}\{\max[(\boldsymbol{w}^{(L)} - \boldsymbol{A}_s \boldsymbol{c}_s)/(\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s \boldsymbol{c}_s), (\boldsymbol{w}^{(H)} - \boldsymbol{A}_s \boldsymbol{c}_s)/(\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s \boldsymbol{c}_s)]\},$$

vector division being elementwise, the maximum of the two vectors being elementwise, and $\min_{\text{k}}$ providing the minimum element. We could also consider the weight vector of the restricted region, $\boldsymbol{w}_{\text{res4}}$, which comes closest to $\boldsymbol{w}_{\text{cal}}$. Again for restricted region $R_{\boldsymbol{w}}$, we would have

$$\boldsymbol{w}_{\text{res4}} = \min[\max(\boldsymbol{w}_{\text{cal}}, \boldsymbol{w}^{(L)}), \boldsymbol{w}^{(H)}].$$

The asymptotic properties of estimators using restricted weights $\boldsymbol{w}_{\text{res3}}$ or $\boldsymbol{w}_{\text{res4}}$ are identical to those of the calibration estimator, as long as $\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s \boldsymbol{c}_s$ converges in probability to $\boldsymbol{0}$, which is usually the case.

One interesting property of all the approaches discussed in this section is that, no matter what the restricted region, the existence of restricted weights is guaranteed. This is not always the case when using an approach based on distance measures. A simple example will now be introduced to allow comparisons between a few approaches.

We wish to estimate a total on the basis of a simple random sample of size 2 in a population of size 20. In other words, $\boldsymbol{c} = \boldsymbol{1}_{20\times 1}$ and $\boldsymbol{a} = 10(\boldsymbol{1}_{20\times 1})$. We use the auxiliary information vector $\boldsymbol{X} = (1, 2, 3, ..., 20)'$, assume that the selected sample is $s = \{2, 12\}$ and choose $\boldsymbol{U}$ as a diagonal matrix with $u_{kk} = x_k = k$. A rectangular restricted region is provided using points $\boldsymbol{w}^{(L)} = (0,\ 0)'$ and $\boldsymbol{w}^{(H)} = (20,\ 13)'$. In other words, the weight of the first sample unit must be greater than 0 and less than 20, whereas the weight of the second sample unit must be greater than 0 and less than 13.

Under these conditions, the calibrated weights $\boldsymbol{w}_{\text{cal}} = (15,\ 15)'$ lie outside the restricted region. Since $p = 1$, weights $\boldsymbol{w}_s(\alpha)$ lie on the line segment which links $\boldsymbol{A}_s \boldsymbol{c}_s = (10,\ 10)'$ to $\boldsymbol{w}_{\text{cal}}$. We therefore have $\boldsymbol{w}_{\text{res1}} = \boldsymbol{w}_{\text{res3}}$, which means that the two methods give the same result. In this case, we have $\boldsymbol{w}_{\text{res1}} = \boldsymbol{w}_{\text{res3}} = (13,\ 13)'$. The method which consists in choosing that point in the restricted region which lies closest to the calibrated weights yields $\boldsymbol{w}_{\text{res4}} = (15,\ 13)'$. On the other hand, if we look for $\boldsymbol{w}_{\text{res5}}$, the restricted weights obtained while requiring that the calibration equation be satisfied and while using a distance measurement which assumes an infinite value outside the restricted region, then there is no solution. In fact, for any weight in the restricted region $\boldsymbol{X}_s' \boldsymbol{w}_s \le 196$, whereas $\boldsymbol{X}'\boldsymbol{c} = 210$. If we had, say, $\boldsymbol{w}^{(H)} = (30,\ 13)'$, then using $D_s(\boldsymbol{w}_s)$ as a distance measurement within the restricted region we would have $\boldsymbol{w}_{\text{res5}} = (27,\ 13)'$. These weights are fairly distant from $\boldsymbol{w}_{\text{cal}} = (15,\ 15)'$ and from $\boldsymbol{A}_s \boldsymbol{c}_s = (10,\ 10)'$. Such is the price to be paid for insisting on having weights which meet the calibration equation.

6. Estimators for domains with a synthetic component

Restricted weights are used because of the properties of the calibration estimator for small sample sizes. For large sample sizes, we normally have $\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s \boldsymbol{c}_s$ converging in probability to $\mathbf{0}$, *i.e.* weights that are not problematic. A statistician faced with a problem of extreme weights must therefore in all likelihood also face another problem of small sample sizes, *i.e.* estimation for small domains. This section introduces an estimator whose asymptotic properties are those of the calibration estimator, but which uses restricted weights and takes advantage of a synthetic estimator.

Let $\tilde{\boldsymbol{Y}} = \boldsymbol{X}\tilde{\boldsymbol{\beta}}_s$ denote a synthetic estimate for $\boldsymbol{Y}$, we have

$$\begin{aligned} \tilde{\boldsymbol{Y}}'\boldsymbol{w}_s &= (\boldsymbol{X}_s\tilde{\boldsymbol{\beta}}_s)'\boldsymbol{w}_s \\ &= \tilde{\boldsymbol{\beta}}_s'\boldsymbol{X}_s'\boldsymbol{w}_s \\ &\simeq \tilde{\boldsymbol{\beta}}_s'\boldsymbol{X}'\boldsymbol{c} \\ &= (\boldsymbol{X}\tilde{\boldsymbol{\beta}}_s)'\boldsymbol{c} \\ &= \tilde{\boldsymbol{Y}}'\boldsymbol{c} \end{aligned} \qquad (13)$$

with equality at the third step if the weights satisfy the calibration equation $\boldsymbol{X}_s'\boldsymbol{w}_s = \boldsymbol{X}'\boldsymbol{c}$. The weights $\boldsymbol{w}_{\text{cal}}$ given by (1) minimize $||\boldsymbol{X}_s'\boldsymbol{w}_s - \boldsymbol{X}'\boldsymbol{c}||_{\boldsymbol{T}}^2$. We can therefore estimate $\boldsymbol{Y}'\boldsymbol{c}$ using

$$\hat{\boldsymbol{\tau}} = (\boldsymbol{Y}_s - \tilde{\boldsymbol{Y}}_s)'\boldsymbol{w}_{\text{res}} + \tilde{\boldsymbol{Y}}'\boldsymbol{c}. \qquad (14)$$

There will be equality between this estimator and estimator $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}}$ once the sample is sufficiently large for the calibration equation to be satisfied and for $\boldsymbol{w}_{\text{cal}}$ to lie in the restricted region, *i.e.* once $\boldsymbol{w}_{\text{res}} = \boldsymbol{w}_{\text{cal}}$. The asymptotic properties of these two estimators are therefore identical under certain conditions discussed in the previous section. The advantage of using estimator $\hat{\boldsymbol{\tau}}$ is that it provides a synthetic estimate when columns of $\boldsymbol{Y}_s$ and $\tilde{\boldsymbol{Y}}_s$ are zero.

7. Outliers

Outliers could be dealt with in much the same way as extreme weights. The strategy is the following: we adopt a restricted region for $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}}$, we show that when n is sufficiently large $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}}$ lies within the restricted region, and we adopt a more "reasonable" estimator to replace $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}}$ in those cases where $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}}$ lies outside the restricted region. For a stratified sample, we would normally have one restricted region per stratum.

In section 2, it was shown that under certain conditions for the asymptotic setup, $\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s \boldsymbol{c}_s = O_p(n^{-3/2}N^{\gamma})$. We thus have $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}} - \boldsymbol{Y}_s'\boldsymbol{A}_s\boldsymbol{c}_s = O_p(n^{-1/2}N^{\gamma})$, and if we assume that

$$\boldsymbol{Y}_s'\boldsymbol{A}_s\boldsymbol{c}_s - \boldsymbol{Y}'\boldsymbol{c} = O_p(n^{-1/2}N^{\gamma}), \qquad (15)$$

then $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}} - \boldsymbol{Y}'\boldsymbol{c} = O_p(n^{-1/2}N^{\gamma})$. An expert (or a group of experts) could determine on the basis of information gathered independently of survey data that it would not be reasonable to have $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}}$ outside a certain region. If $\boldsymbol{Y}'\boldsymbol{c}$ lies within the restricted region (*i.e.* if the expert does not find it unreasonable to have an estimate of the parameter which would be equal to the true value, $\boldsymbol{Y}'\boldsymbol{c}$, of the parameter), if $\gamma = 0$, and if the restricted region does not vary with n or N (or if $\gamma = 1$, and the restricted region varies in proportion to N), then for sufficiently large n, the probability that $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}}$ will lie within the restricted region is equal to one. In those cases where $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}}$ lies outside the restricted region, we could use as an estimate the point in the restricted region that lies closest to $\boldsymbol{Y}_s'\boldsymbol{w}_{\text{cal}}$ or we could assume that the weight of the few observations that are deemed outliers is equal to one, and distribute their original weights (less the number of outliers) among the observations that are not outliers. The asymptotic properties of this modified estimator used to deal with outliers are then identical to those of the unmodified estimator.

In the case of a non-stratified sample, this method is relatively easy to apply. If however the sample is stratified, and if constraints are imposed on estimates for each stratum, then we have two additional problems. First, if the asymptotic setup is such that the number of strata increases in proportion to the size of the sample, then the assumption given in (15) does not hold, since the mean sample size per stratum remains constant as $n \to \infty$. We need to determine whether it is reasonable to adopt an asymptotic setup in which the number of strata is constant (or increases less rapidly than n). Such an asymptotic setup is less plausible if the number of observations per stratum is small. The second problem is linked to the difficulty for the expert to impose constraints on estimates for each of the strata. The greater the number of strata, the greater the risk that $\boldsymbol{Y}'\boldsymbol{c}$ will not lie in the restricted region defined by the expert. In fact, in the case of a stratified sample, it is preferable for the expert to use information that is independent of the survey data, in order to ensure strata homogeneity, prior to finalizing stratification. In other words, it is preferable to use the information available before the survey, in order to prevent outliers, rather than to correct them. If the information has been used in such a way that, before the survey, there is no reason to believe that there is any unrepresentative observation in any stratum, then there is no justification for assuming the opposite once the data have been collected.

8. Conclusion

If for large sample sizes the calibrated weights remain within a restricted region, then the asymptotic properties of the estimator with restricted weights are obviously identical to those of the calibration estimator. For a given asymptotic setup, we can usually expect to have $\boldsymbol{w}_{\text{cal}} - \boldsymbol{A}_s \boldsymbol{c}_s$ converging in probability to $\boldsymbol{0}$, *i.e.* for sufficiently large sample sizes the calibrated weights $\boldsymbol{w}_{\text{cal}}$ will remain within the restricted region $R_{\boldsymbol{w}}$ if $\boldsymbol{A}_s \boldsymbol{c}_s$ lies within $R_{\boldsymbol{w}}$. However, we have seen that for the estimate of a total, we do not necessarily have convergence to 0 for $D_s(\boldsymbol{w}_{\text{cal}})$. We must therefore avoid having a restricted region defined by $\|\boldsymbol{w}_s - \boldsymbol{A}_s \boldsymbol{c}_s\|_{U_s}^2 \le l$ at least if we are estimating a total and not a mean.

We have provided necessary and sufficient conditions for the existence of weights restricted to intervals which satisfy the calibration equation. If such weights do not exist, the idea of satisfying the calibration equation exactly must be abandoned. The problem of calibration with restricted weights can be reformulated in such a way that a solution will always be possible. Some of the approaches described in this paper make it possible to obtain a solution without recourse to iterative methods. These are simple methods that are easy to interpret. The asymptotic properties of these estimators are usually identical to those of the calibration estimator without weight restrictions.

The problem of extreme weights is encountered for small sample sizes, thus the problem of estimating for small domains should be considered simultaneously. It is possible to take advantage of synthetic estimators while using an estimator with restricted weights having good asymptotic properties.

It is also possible to modify the calibration estimator, or any other asymptotically consistent estimator, so as to deal with outliers. The conditions under which this modified estimator will have asymptotic properties identical to those of the unmodified estimator are not easily verified, just as it is difficult to verify whether an outlier is in fact unrepresentative. However, such conditions make it possible to identify those factors which allow an estimator that is corrected for outliers to be statistically valid.

Acknowledgement

The author wishes to thank an associate editor and a referee for constructive comments which have helped improve the paper.

Appendix A

We wish to verify that $\Omega(\boldsymbol{\varphi}) = \boldsymbol{l}'(\boldsymbol{V}\boldsymbol{\varphi})_- - \boldsymbol{h}'(\boldsymbol{V}\boldsymbol{\varphi})_+$ has a value of zero or less. First, it is easily shown that this is true for a vector $\boldsymbol{\varphi}$, if and only if it is true for a vector $k\boldsymbol{\varphi}$ with arbitrary $k > 0$. Only the direction of $\boldsymbol{\varphi}$ matters. It is therefore sufficient to verify the condition for $\boldsymbol{\varphi}$ of norm equal to one. For this proof, we will use the 1_1-norm of $\boldsymbol{\varphi}$, $\|\boldsymbol{\varphi}\|_{l_1} = \sum_{i=1}^{p} |\varphi_i|$. Vectors $\boldsymbol{\varphi}$ with $\|\boldsymbol{\varphi}\|_{l_1} = 1$ are located in hyperplanes whose intersections lie on points orthogonal to the unit vectors, *i.e.* points at least one of whose coordinates is zero. Function Ω varies linearly except at points $\boldsymbol{\varphi}$ orthogonal to one or more rows of $\boldsymbol{V}$. Even when the domain of Ω is restricted to vectors $\boldsymbol{\varphi}$ with $\|\boldsymbol{\varphi}\|_{l_1} = 1$ that are orthogonal to $0 \le j < (p-1)$ linearly independent rows of $\boldsymbol{V}$, function Ω still varies linearly except at points orthogonal to other rows of $\boldsymbol{V}$ or orthogonal to unit vectors (which are likewise rows of $\boldsymbol{V}$). The maximum of Ω for $\|\boldsymbol{\varphi}\|_{l_1} = 1$ is therefore reached at a point $\boldsymbol{\varphi}$ orthogonal to $(p-1)$ linearly independent rows of $\boldsymbol{V}$. It is therefore sufficient to verify the condition for two vectors of opposite direction which are orthogonal to $(p-1)$ linearly independent rows of $\boldsymbol{V}$, and this for each subset of $(p-1)$ linearly independent rows of $\boldsymbol{V}$.

Appendix B

Let $\text{vec}(\boldsymbol{F})$ denote the vector obtained by piling successive columns of matrix $\boldsymbol{F} \in \mathbb{R}^{a \times b}$ with the first column on top, and let the Kronecker product of two matrices $\boldsymbol{F}$ and $\boldsymbol{G}$ be defined as

$$\boldsymbol{F} \otimes \boldsymbol{G} = \begin{pmatrix} f_{11}\boldsymbol{G} & \cdots & f_{1n}\boldsymbol{G} \\ \vdots & & \vdots \\ f_{m1}\boldsymbol{G} & \cdots & f_{mn}\boldsymbol{G} \end{pmatrix}. \qquad \text{(B1)}$$

The result is derived from the corollary in section 3 with

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{I}_{RC} \\ \boldsymbol{I}_R \otimes \boldsymbol{1}_{1\times C} \\ \boldsymbol{1}_{1\times R} \otimes \boldsymbol{I}_C \\ \boldsymbol{1}_{1\times RC} \end{pmatrix}, \quad \boldsymbol{w} = \text{vec}((\hat{N}_{ij})'),$$

$$\boldsymbol{l} = \begin{pmatrix} \text{vec}((N_{ij}^{(L)})') \\ N_{1.}^{(L)} \\ \vdots \\ N_{R.}^{(L)} \\ N_{.1}^{(L)} \\ \vdots \\ N_{.C}^{(L)} \\ N_{..}^{(L)} \end{pmatrix}, \ \boldsymbol{h} = \begin{pmatrix} \text{vec}((N_{ij}^{(H)})') \\ N_{1.}^{(H)} \\ \vdots \\ N_{R.}^{(H)} \\ N_{.1}^{(H)} \\ \vdots \\ N_{.C}^{(H)} \\ N_{..}^{(H)} \end{pmatrix}. \qquad \text{(B2)}$$

Only a finite set of conditions need be verified, first by noting that the columns of

$$\boldsymbol{V} = \begin{pmatrix} -\boldsymbol{I}_R \otimes \mathbf{1}_{C\times 1} & -\mathbf{1}_{R\times 1} \otimes \boldsymbol{I}_C & -\mathbf{1}_{RC\times 1} \\ \boldsymbol{I}_R & \mathbf{0}_{R\times C} & \mathbf{0}_{R\times 1} \\ \mathbf{0}_{C\times R} & \boldsymbol{I}_C & \mathbf{0}_{C\times 1} \\ \mathbf{0}_{1\times R} & \mathbf{0}_{1\times C} & 1 \end{pmatrix} \quad \text{(B3)}$$

form a basis for $N(\boldsymbol{M}')$. In other words, $\boldsymbol{M}'\boldsymbol{V} = \mathbf{0}$, the columns of $\boldsymbol{V}$ are linearly independent, and $N(\boldsymbol{M}')$ is of dimension $R + C + 1$. Note also that the last $R + C + 1$ rows of $\boldsymbol{V}$ are the unit vectors. Finally, we verify the conditions of the corollary for all vectors $\boldsymbol{\lambda} = \boldsymbol{V}\boldsymbol{\varphi}$ and $\boldsymbol{\lambda} = -\boldsymbol{V}\boldsymbol{\varphi}$, where $\boldsymbol{\varphi}$ is orthogonal to $R + C$ linearly independent rows of $\boldsymbol{V}$. This last step is described in greater detail in the following paragraph.

An arbitrary subset of $R + C$ linearly independent rows of $\boldsymbol{V}$ which includes the last row of $\boldsymbol{V}$ is denoted L, and the subset of all rows of $\boldsymbol{V}$ which are linear combinations of rows of L is denoted L^+. If L^+ includes row $RC + i$ $(i = 1, ..., R)$ if and only if $i \notin S \subseteq \{1, 2, ..., R\}$, and includes row $RC + R + j$ $(j = 1, ..., C)$ if and only if $j \notin T \subseteq \{1, 2, ..., C\}$, then we set $\boldsymbol{\varphi} = (\boldsymbol{\varphi}_S', -\boldsymbol{\varphi}_T', 0)'$, where the i^{th} element of $\boldsymbol{\varphi}_S \in \mathbb{R}^R$ is equal to 1 if $i \in S$ and to 0 otherwise, and the j^{th} element of $\boldsymbol{\varphi}_T \in \mathbb{R}^C$ is equal to 1 if $j \in T$ and to 0 otherwise. Then

$$\boldsymbol{V}\boldsymbol{\varphi} = ((-\boldsymbol{\varphi}_S \otimes \mathbf{1}_{C\times 1} + \mathbf{1}_{R\times 1} \otimes \boldsymbol{\varphi}_T)', \boldsymbol{\varphi}_S', -\boldsymbol{\varphi}_T', 0)',$$

therefore $\boldsymbol{\varphi}$ is orthogonal to all rows of L^+, and all the more so $\boldsymbol{\varphi}$ is orthogonal to all rows of L. Likewise, vector $\boldsymbol{\varphi}^* = (\boldsymbol{\varphi}_S', \boldsymbol{\varphi}_T', -1)'$ is orthogonal to all rows of a subset of $R + C$ linearly independent rows of $\boldsymbol{V}$ which includes row $RC + i$ $(i = 1, ..., R)$ if and only if $i \notin S$, and includes row $RC + R + j$ $(j = 1, ..., C)$ if and only if $j \notin T$, but does not include the last row of $\boldsymbol{V}$. The condition $-\boldsymbol{l}'\boldsymbol{\lambda}_- \leq \boldsymbol{h}'\boldsymbol{\lambda}_+$ with $\boldsymbol{\lambda} = \boldsymbol{V}\boldsymbol{\varphi}$ provides the fifth set of inequalities in (9). Likewise, by assuming $\boldsymbol{\lambda}$ equal to $-\boldsymbol{V}\boldsymbol{\varphi}$, $\boldsymbol{V}\boldsymbol{\varphi}^*$ and $-\boldsymbol{V}\boldsymbol{\varphi}^*$ we obtain the last three sets of inequalities in (9).

References

Bacharach, M. (1965). Estimating nonnegative matrices from marginal data. *International Economic Review*, 6, 294-310.

Bardsley, P., and Chambers, R.L. (1984). Multipurpose estimation from unbalanced samples. *Applied Statistics*, 33, 290-299.

Ben-Israel, A., and Greville, T.N.E. (1980). *Generalized Inverses*: *Theory and Applications*. Huntington, New York: Robert E. Krieger Publishing Company.

Brewer, K.R.W. (1979). A class of robust sampling designs for large scale surveys. *Journal of the American Statistical Association*, 74, 911-915.

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

Duchesne, P. (1999). Robust calibration estimators. *Survey Methodology*, 25, 43-56.

Fan, K. (1956). On systems of linear inequalities. *Annals of Mathematics Studies*, (Eds. H.W. Kuhn, and A.W. Tucker), 38, 99-156.

Graybill, F.A. (1983). *Matrices with Applications in Statistics*, (Second Edition). Belmont, California: Wadsworth Publishing.

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

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

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Théberge, A. (1999). Extensions of calibration estimators in survey sampling. *Journal of the American Statistical Association*, 94, 635-644.