ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 15, Number 2, 2011
Available online at www.math.ut.ee/acta/

Domain estimators calibrated on information from another survey

Carl-Erik Särndal and Imbi Traat

Abstract. We examine calibration estimation in a setting where two surveys are conducted on the same finite population. Some variables of study are common to the two surveys, but the second one requires greater detail in the statistics produced than the already published first one. More specifically, we require estimates for sub-populations, called domains, that are identified only in the second survey to add up consistently to known or estimated totals published for the common variables in the first survey. We outline and study several options for deriving calibration estimators for the domains identified in the second survey. We obtain explicit expressions showing how the calibration weights are related in the different approaches. The concluding section presents the results of a simulation study, comparing the precisions attained in the different options.

1. Introduction

It happens frequently nowadays that several surveys are carried out simultaneously, or almost simultaneously, on the same finite population. Typically, some variables are common in two or more surveys. It is natural to require that estimates for the common variables are consistent with each other in the different surveys.

In this paper we consider the situation where population totals are available, estimated or exactly known, from an already completed survey called the *reference survey* (RFS), while totals for certain domains need to be estimated in a subsequent survey, called the *present survey* (PRS). The term "RFS estimates" includes the case where they are true non-random values known from registers or other reliable data sources.

Received February 18, 2011.
2010 *Mathematics Subject Classification.* 62D05.
Key words and phrases. Auxiliary variables, calibration, common variables, consistent estimation, repeated weighting, multi-survey situation.
This work was supported by the Estonian Science Foundation Grant 8789.

The estimation of totals for domains requires information on domain membership for the units. Our starting point is that membership in the domains of interest is lacking for the units in the RFS, but observed for the units in the new survey, the PRS. Also observed in the PRS are the target variables, which include the common variables, and suitable auxiliary variables. Domain estimation then becomes possible, through the PRS.

Furthermore, consistency for variables common to the two surveys is required, so that the estimated domain totals in the PRS sum up to the corresponding already published totals estimated in the RFS for the whole population, or for larger domains of the population. Such consistency is often required by users of survey statistics; it is a cosmetic feature, but important for users. It may or may not be beneficial, in the sense of reducing variance in the domain estimates. As our paper shows, there is no simple answer to the question.

An example of this general situation is as follows. In a survey of individuals, estimates of total income by age class are available from the already completed RFS. Domains defined by age class are thus recorded in the RFS sample. The need then arises to have estimates of total income for domains not identified in the RFS sample, namely, those defined by the cross classification of age class and level of education. Both age and education will be observed for the PRS sample, and estimates of income by age class and education level can be produced. The estimates of income for these cross-classified domains have to be consistent, that is, they must add up over education classes to the already published estimates of income by age class.

In this paper we assume that both surveys, the RFS and the PRS, are carried out by probability sampling, with an arbitrary sampling design, from one and the same finite population. We consider the case where the two samples are independent of each other. Another obvious possibility, namely that the PRS sample is a subsample of the RFS sample, is not discussed.

The objective in the PRS is to produce consistent domain estimates through the classical calibration approach as presented for example in Deville and Särndal (1992). Two kinds of variables can intervene in the calibration of weights: (i) variables that are auxiliary in the usual sense of this term; for those *A-variables* we have known population totals, and (ii) variables common to the two surveys; estimates of totals for these *C-variables* are available from the RFS, and we now seek consistency with them in the PRS. We call *AC-calibration* the procedure consisting in calibration in one single step on both types of variables, with the PRS design weights as starting weights.

The A-variables, if reasonably powerful, are expected to be beneficial, in contributing to a lower variance in the domain estimates. But it is not obvious whether increased precision is realized by requiring consistency for the C-variables. There may be a price (in the form of some increase in variance) to pay for requiring consistent domain estimates.

We compare AC-calibration with *A-calibration*, that is, one where the weights are calibrated only on the A-variables. We derive an expression showing explicitly how the A-calibrated weights are modified to arrive at the C-variable consistency that characterizes the AC-calibrated weights.

Another procedure also examined in this paper consists in calibration in two steps. The first step is an A-calibration. Preliminary weights are thereby obtained. They are then used as starting weights for the calibration in the second step, which may involve both the A-variables and the C-variables, or just the C-variables alone. In both cases, the desired consistency of domain estimates for C-variables is realized. For obvious reasons, both varieties may be described as repeated weighting. This term has in fact been used in several articles on weighting of frequency tables from the Dutch national statistical agency CBS, see for example Kroese and Renssen (1999), Knottnerus and van Duin (2006). We derive the respective expression of weights for domain estimation in general, and compare it with the AC-calibrated weights.

The contribution of this paper can be summarized as follows. We outline several different ways to obtain, through calibration, estimates for domains that add up in a consistent manner to estimates published in an earlier survey. For this purpose, we distinguish an AC-calibration approach and a repeated weighting approach. The weights in these approaches have certain features in common, and both weight systems are uniformly applicable to all study variables. We show the differences of the two approaches by explicit expressions for the weights. But because of complex interrelationships among variables, we are not in a position to conclude that one approach is always better, from a variance point of view, than another. The preferred solution is data dependent.

The paper is arranged as follows. Section 2 presents AC-calibration. In Section 3 we note the special case of A-calibration and show how its calibrated weights are related to those of AC-calibration. Repeated weighting is developed in Section 4. Section 5 discusses related works, and Section 6 is devoted to illustrative special cases. The covariance matrix for AC-estimators is presented in Section 7, and the concluding Section 8 reports a simulation study.

2. The AC-calibration

We assume that two independent surveys are conducted on the same finite population $U = \{1, 2, \ldots, k, \ldots, N\}$. One of those, the reference survey (RFS) is already completed. An m-dimensional study variable vector $\mathbf{y}$, with value $\mathbf{y}_k$ for unit k is observed in the RFS and a design-based estimate $\hat{\mathbf{Y}}_0$

of the unknown total $\mathbf{Y} = \sum_U \mathbf{y}_k$ is produced. The estimator $\hat{\mathbf{Y}}_0$ is assumed to be design-consistent, nearly design-unbiased, and have high precision. A special case occurs when the RFS information $\hat{\mathbf{Y}}_0$ comes from registers. We use the term RFS for this case also, although there is no sampling, and thus, $\hat{\mathbf{Y}}_0$ is the true, non-random total $\mathbf{Y}$.

In another survey, called the present survey (PRS), the aim is to study specific domains. The finite population U is divided into non-overlapping and exhaustive domains U_d, $d \in \mathfrak{D} = \{1, 2, \dots, D\}$. These domains, called PRS domains, are identified in the PRS but not in the RFS. We observe the same m-dimensional study variable vector $\mathbf{y}$ in the PRS. The aim is to estimate the vector of domain totals $\mathbf{Y}_d = \sum_{U_d} \mathbf{y}_k$, for $d \in \mathfrak{D}$, consistently with the RFS information, i.e. we want the estimates of $\mathbf{Y}_d$ to sum up to $\hat{\mathbf{Y}}_0$. To realize that objective, we use the classical calibration approach extended to handle two types of calibration information.

In the PRS let a sample s of size n be drawn by a probability sampling design such that the inclusion probability of unit k is $\pi_k > 0$, so that the design weight of unit k is $a_k = 1/\pi_k$. The sample part falling into U_d is denoted by $s_d = U_d \cap s$. For units k in the PRS sample s we observe the vectors $\mathbf{x}_k$ and $\mathbf{y}_k$, where $\mathbf{x}_k$ is an auxiliary variable vector in the traditional sense, having known total $\mathbf{X} = \sum_U \mathbf{x}_k$. Calibration on this information will usually make estimates more precise. The calibration on the $\mathbf{y}_k$-information serves to achieve consistency with the RFS, and may or may not improve precision.

We define the $(p+m)$-dimensional vectors $(p + m \leq n)$

$$\begin{pmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{pmatrix}, \; k \in s, \; \begin{pmatrix} \mathbf{X} \\ \hat{\mathbf{Y}}_0 \end{pmatrix}, \; \begin{pmatrix} \hat{\mathbf{X}} \\ \hat{\mathbf{Y}} \end{pmatrix}, \tag{1}$$

where $\mathbf{X}$ and $\hat{\mathbf{Y}}_0$ are known, and $\hat{\mathbf{X}} = \sum_s a_k \mathbf{x}_k$ and $\hat{\mathbf{Y}} = \sum_s a_k \mathbf{y}_k$ are Horvitz-Thompson (HT) estimators computed in the PRS. Let $\mathbf{z}_{ACk}$ be an instrument vector; an arbitrary vector with dimension matching that of $(\mathbf{x}_k', \mathbf{y}_k')'$. A standard choice is $\mathbf{z}_{ACk} = q_k(\mathbf{x}_k', \mathbf{y}_k')'$ with specified constants q_k, often chosen as $q_k = 1$. The choice of the instrument vector has some, but usually only minor, impact on the variance of the calibration estimator. The instrument vector approach to calibration, see for example Särndal (2007), gives the weights

$$w_{ACk} = a_k(1 + \boldsymbol{\lambda}_{AC}' \mathbf{z}_{ACk}), \tag{2}$$

where

$$\boldsymbol{\lambda}_{AC}' = \begin{pmatrix} \mathbf{X} - \hat{\mathbf{X}} \\ \hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}} \end{pmatrix}' \mathbf{M}^{-1}, \; \mathbf{M} = \sum_s a_k \mathbf{z}_{ACk} \begin{pmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{pmatrix}'. \tag{3}$$

The w_{ACk} satisfy both the A-constraint and the C-constraint:

$$\sum_s w_{ACk}\mathbf{x}_k = \mathbf{X}, \quad \sum_s w_{ACk}\mathbf{y}_k = \hat{\mathbf{Y}}_0. \tag{4}$$

We assume that there is no linear dependence among the variables contained in the vectors $\mathbf{x}_k$ and $\mathbf{y}_k$.

Once computed, the weights w_{ACk} can be used for estimating all totals of interest in the PRS. In particular, $\mathbf{Y} = \sum_U \mathbf{y}_k$, is estimated by $\hat{\mathbf{Y}}_{AC} = \sum_s w_{ACk}\mathbf{y}_k$, and $\mathbf{Y}_d$ is estimated as

$$\hat{\mathbf{Y}}_{ACd} = \sum_{s_d} w_{ACk}\mathbf{y}_k, \; d \in \mathfrak{D}. \tag{5}$$

We have additive consistency with the RFS estimator $\hat{\mathbf{Y}}_0$, since (4) and (5) imply $\sum_{d\in\mathfrak{D}} \hat{\mathbf{Y}}_{ACd} = \hat{\mathbf{Y}}_0$. In other words, the estimates we present for the domain totals, $\hat{\mathbf{Y}}_{ACd}$, leave the estimate for the whole population total unchanged; it agrees with the earlier published $\hat{\mathbf{Y}}_0$.

Applications of the domain estimator $\hat{\mathbf{Y}}_{ACd}$ include the following.

(1) The vector $\mathbf{y}_k$ consists of different study variables such as salary, alcohol consumption, living expenditures and so on. Then $\hat{\mathbf{Y}}_{ACd}$ estimates the totals of these variables in domain d, for example, in an educational class. The sum of the estimates over all educational classes equals $\hat{\mathbf{Y}}_0$, in accordance with the RFS estimate.

(2) The vector $\mathbf{y}_k$ is itself a domain vector, so that $\mathbf{y}_k = y_k\boldsymbol{\gamma}_k$, where, for example, y_k is salary and $\boldsymbol{\gamma}_k$ identifies m age class domains. Then the vector $\hat{\mathbf{Y}}_{ACd}$ estimates total salary by age class in the education domain d. Letting $d = 1, 2, \ldots, D$, we get estimates of total salary for all mD sub-domains defined by the cross-classification of age and education. These sub-domain estimates have additive consistency: Their sum over all D education domains agrees with total salary by age class $\hat{\mathbf{Y}}_0$, as estimated in the RFS or known by registers.

(3) In an extension of case (2), we can let the RFS domain indicator $\boldsymbol{\gamma}_k$ in $\mathbf{y}_k = y_k\boldsymbol{\gamma}_k$ represent a cross-classification of two or more other categorical variables, and the PRS domains in the set $\mathfrak{D}$ may represent two or more other categorical variables. Then $\hat{\mathbf{Y}}_{ACd}$ for $d \in \mathfrak{D}$ produces PRS estimates for sub-domains defined by cross-classifying all the categorical variables involved. They add up consistently to the RFS estimate for corresponding domains.

Calibration is well established practice in most statistical agencies. So the AC-calibrated weights w_{ACk} can be routinely computed for any instrument vector $\mathbf{z}_{ACk}$ from the general expressions (2) – (3). A particular choice for the instrument vector allows us to derive a suggestive alternative form for the weights w_{ACk}, and we can make an easy comparison with the special case called A-calibration, where the C-information is omitted in (2) – (3).

3. A-calibration compared with AC-calibration

The special case where the C-information is omitted in (2) – (3), called A-calibration, is characterized by the A-calibrated weights,

$$w_{Ak} = a_k(1 + \boldsymbol{\lambda}'_A \mathbf{z}_{Ak}), \text{ where } \boldsymbol{\lambda}'_A = (\mathbf{X} - \hat{\mathbf{X}})' \left(\sum_s a_k \mathbf{z}_{Ak} \mathbf{x}'_k \right)^{-1}, \tag{6}$$

where the dimension of the instrument vector $\mathbf{z}_{Ak}$ matches that of $\mathbf{x}_k$. With these weights we form the calibration estimator of the population total $\mathbf{Y}$ as $\hat{\mathbf{Y}}_A = \sum_s w_{Ak}\mathbf{y}_k$, and that of the domain totals $\mathbf{Y}_d$ as

$$\hat{\mathbf{Y}}_{Ad} = \sum_{s_d} w_{Ak}\mathbf{y}_k, \ \forall d \in \mathfrak{D}. \tag{7}$$

By construction these domain estimators add up to the estimate $\hat{\mathbf{Y}}_A$ for the whole population total,

$$\sum_{d\in\mathfrak{D}} \hat{\mathbf{Y}}_{Ad} = \hat{\mathbf{Y}}_A, \tag{8}$$

but $\hat{\mathbf{Y}}_A$ does not agree with $\hat{\mathbf{Y}}_0$, so consistency with the RFS is no longer present.

Different choices of $\mathbf{z}_{Ak}$ give different weights. When $\mathbf{y}_k$ is one-dimensional, the asymptotically optimal choice of $\mathbf{z}_{Ak}$ for a given sampling design and given $\mathbf{x}$-vector is $\mathbf{z}_{Ak} = a_k^{-1} \sum_{l\in s}(a_k a_l - a_{kl})\mathbf{x}_l$ (Estevao and Särndal 2004; Kott 2004), where a_{kl} is the inverse of the second order inclusion probability. The need to involve the a_{kl} makes this choice less appealing.

We use $\mathbf{z}_{Ak} = q_k\mathbf{x}_k$, with specified positive constants q_k. For a univariate $\mathbf{y}_k$ this choice gives the classical GREG estimator (e.g. Särndal et al. 1992, pp. 225–228). We get the multivariate GREG estimator,

$$\hat{\mathbf{Y}}_A = \hat{\mathbf{Y}} + (\mathbf{X} - \hat{\mathbf{X}})'\hat{\mathbf{B}}, \tag{9}$$

where $\hat{\mathbf{Y}} = \sum_s a_k\mathbf{y}_k$, and $\hat{\mathbf{B}}$ is a coefficient matrix for a multivariate regression fit (Rencher 1988). Each component of the vector $\mathbf{y}_k$, when regressed on $\mathbf{x}_k$, gives rise to a regression coefficient vector given by the corresponding column of

$$\hat{\mathbf{B}} = \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\hat{\mathbf{T}}_{\mathbf{xy}} : \ p \times m, \tag{10}$$

where

$$\hat{\mathbf{T}}_{\mathbf{xx}} = \sum_s a_k q_k \mathbf{x}_k \mathbf{x}'_k : \ p \times p, \quad \hat{\mathbf{T}}_{\mathbf{xy}} = \sum_s a_k q_k \mathbf{x}_k \mathbf{y}'_k : \ p \times m. \tag{11}$$

The matrix $\hat{\mathbf{B}}$ generates for unit k a vector of predicted values and a vector of residuals given respectively by

$$\hat{\mathbf{y}}_k = \hat{\mathbf{B}}'\mathbf{x}_k,\ \ \mathbf{e}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k. \tag{12}$$

We want to compare the weights in AC-calibration with those in A-calibration. Let us fix the instrument vector for the AC-calibration as $\mathbf{z}'_{ACk} = q_k(\mathbf{x}'_k, \mathbf{y}'_k)$. Then the AC calibrated weights (2) – (3) can be written on the form (see Appendix),

$$w_{ACk} = w_{Ak} + a_k q_k \mathbf{e}'_k \hat{\mathbf{Q}}^{-1}(\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}}_A), \tag{13}$$

where the w_{Ak} are the A calibrated weights (6), $\hat{\mathbf{Y}}_A$ is given in (9), $\mathbf{e}_k$ in (12), and

$$\hat{\mathbf{Q}} = \sum_s a_k q_k \mathbf{y}_k \mathbf{e}'_k :\ m \times m. \tag{14}$$

Note that the dimensionality is reduced when w_{ACk} is computed from (13) rather than from (2) – (3); in (13) the inversions of the $p \times p$ matrix $\hat{\mathbf{T}}_{\mathbf{xx}}$ and the $m \times m$ matrix $\hat{\mathbf{Q}}$ are carried out separately.

The dependence of the matrix $\hat{\mathbf{Q}}$ on the residuals $\mathbf{e}_k$ is further illustrated by an alternative expression. In (14), $\mathbf{y}_k$ can be replaced by $\mathbf{e}_k = \mathbf{y}_k - \hat{\mathbf{B}}'\mathbf{x}_k$ because

$$\sum_s a_k q_k \mathbf{x}_k \mathbf{e}'_k = \sum_s a_k q_k \mathbf{x}_k(\mathbf{y}'_k - \mathbf{x}'_k\hat{\mathbf{B}}) = \hat{\mathbf{T}}_{\mathbf{xy}} - \hat{\mathbf{T}}_{\mathbf{xx}}\hat{\mathbf{B}} = \mathbf{0}.$$

Consequently,

$$\hat{\mathbf{Q}} = \sum_s a_k q_k \mathbf{e}_k \mathbf{e}'_k. \tag{15}$$

The weight system (13) can be used for all variables whose totals or domain totals need to be estimated in the PRS. We focus here on estimates for the common variables, for which the known RFS vector $\hat{\mathbf{Y}}_0$ contains informative control totals. The AC-calibrated estimator of $\mathbf{Y}_d$ is obtained as

$$\hat{\mathbf{Y}}_{ACd} = \sum_{s_d} \mathbf{y}_k w_{ACk} = \hat{\mathbf{Y}}_{Ad} + \hat{\mathbf{Q}}_d \hat{\mathbf{Q}}^{-1}(\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}}_A), \tag{16}$$

where $\hat{\mathbf{Y}}_{Ad}$ is the A-calibrated estimator (7), and

$$\hat{\mathbf{Q}}_d = \sum_{s_d} a_k q_k \mathbf{y}_k \mathbf{e}'_k. \tag{17}$$

Noting that $\sum_{d\in\mathfrak{D}} \hat{\mathbf{Y}}_{Ad} = \hat{\mathbf{Y}}_A$ and $\sum_{d\in\mathfrak{D}} \hat{\mathbf{Q}}_d = \hat{\mathbf{Q}}$, the additive consistency with the RFS follows: $\sum_{d\in\mathfrak{D}} \hat{\mathbf{Y}}_{ACd} = \hat{\mathbf{Y}}_0$. In the unlikely case that the RFS and the PRS agree in their estimates of the whole population total, so that $\hat{\mathbf{Y}}_A = \hat{\mathbf{Y}}_0$, then the A calibrated domain estimator is unchanged, i.e. $\hat{\mathbf{Y}}_{ACd} = \hat{\mathbf{Y}}_{Ad}$.

We can interpret (16) to say that the additive consistency with $\hat{\mathbf{Y}}_0$ is a consequence of distributing the difference $\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}}_A$ over the D domains, with the aid of the adjustment matrix $\hat{\mathbf{Q}}_d\hat{\mathbf{Q}}^{-1}$. The next section discusses the technique known as repeated weighting. Placed in a context of domain estimation, this technique provides another example of distributing the difference $\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}}_A$ over the set of domains, but with a different adjustment matrix.

4. Repeated weighting

Repeated Weighting (RW) was developed at the Central Bureau of Statistics (CBS) in the Netherlands. The original idea is to produce frequency tables in a consistent manner, so that estimated marginal frequencies in the PRS agree with their counterparts in another survey, one that precedes the present one in a temporal or other ordering. Some references are Kroese and Renssen (1999), Houbiers (2004), and Knottnerus and van Duin (2006). We formulate the RW method somewhat differently, so as to facilitate comparison with AC-calibration.

The RW method can be described as calibration in two steps. The A-calibrated weights w_{Ak} given by (6) are based on a given auxiliary vector $\mathbf{x}_k$ and a given instrument vector $\mathbf{z}_{Ak}$. In a second step, the w_{Ak} are used as starting weights for a new calibration, leading to weights w_{RWk}, $k \in s$, that satisfy the constraint

$$\sum_s w_{RWk}\mathbf{y}_k = \hat{\mathbf{Y}}_0, \tag{18}$$

where $\mathbf{y}_k$ is the vector of common variables, and as before, $\hat{\mathbf{Y}}_0$ is the RFS estimate. Let $\mathbf{z}_{RWk}$ be an instrument vector with dimension matching that of $\mathbf{y}_k$. Calibrated weights are then given by $w_{RWk} = w_{Ak}(1 + \boldsymbol{\lambda}'_{RW}\mathbf{z}_{RWk})$, where $\boldsymbol{\lambda}_{RW}$ is determined to satisfy the constraint (18). This leads to the RW weights

$$w_{RWk} = w_{Ak}\{1 + \mathbf{z}'_{RWk}\hat{\mathbf{Q}}_R^{-1}(\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}}_A)\}, \tag{19}$$

for $k \in s$, where $\hat{\mathbf{Y}}_A = \sum_s w_{Ak}\mathbf{y}_k$, and

$$\hat{\mathbf{Q}}_R = \sum_s w_{Ak}\mathbf{z}_{RWk}\mathbf{y}'_k. \tag{20}$$

The resulting estimator of the domain total $\mathbf{Y}_d = \sum_{U_d}\mathbf{y}_k$ is $\hat{\mathbf{Y}}_{RWd} = \sum_{s_d} w_{RWk}\mathbf{y}_k$ which we can write, for easy comparison with (16), on the alternative form

$$\hat{\mathbf{Y}}_{RWd} = \hat{\mathbf{Y}}_{Ad} + \hat{\mathbf{Q}}_{dR}\hat{\mathbf{Q}}_R^{-1}(\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}}_A), \tag{21}$$

where $\hat{\mathbf{Y}}_{Ad} = \sum_{s_d} w_{Ak}\mathbf{y}_k$, and

$$\hat{\mathbf{Q}}_{dR} = \sum_{s_d} w_{Ak}\mathbf{z}_{RWk}\mathbf{y}'_k. \tag{22}$$

Summing over domains we get $\hat{\mathbf{Y}}_{RW} = \sum_{d\in D} \hat{\mathbf{Y}}_{RWd} = \hat{\mathbf{Y}}_0$.

The applications of RW at CBS aim primarily at producing frequency tables in the present survey to achieve consistency with corresponding estimates in preceding surveys. We can describe their objective as choosing $\mathbf{z}_{RWk} = \mathbf{y}_k = \boldsymbol{\gamma}_k$, where $\boldsymbol{\gamma}_k$ codifies the table cells with which consistency is required. But more generally, we can allow the vector $\mathbf{y}_k$ in (18) to be defined, for example, as in the cases (1), (2) and (3) discussed in Section 2. Additionally, we have the freedom to choose $\mathbf{z}_{RWk}$.

The weight system w_{RWk} given by (19) with $\mathbf{z}_{RWk} = \mathbf{y}_k$ is not A-consistent, that is, not consistent with the known vector total $\mathbf{X} = \sum_U \mathbf{x}_k$. However, this consistency can be achieved in two ways: (1) by including $\mathbf{x}_k$ in the vector $\mathbf{y}_k$ in (18) − (19), or (2) by choosing $\mathbf{z}_{RWk} = \mathbf{e}_k a_k q_k / w_{Ak}$, where $\mathbf{e}_k$ are regression residuals given by (12), and the w_{Ak} are given by (6) with $\mathbf{z}_{Ak} = q_k\mathbf{x}_k$. In case (2), the RW weight (19) and the AC calibrated weight (13) coincide, and so do the domain estimators.

We cannot conclude, neither theoretically nor by our empirical experience, that $\hat{\mathbf{Y}}_{ACd}$ is better than $\hat{\mathbf{Y}}_{RWd}$ from the point of view of smaller variance, despite of the fact that the former but not the latter is A-calibrated. And although $\hat{\mathbf{Y}}_{RW} = \hat{\mathbf{Y}}_{AC} = \hat{\mathbf{Y}}_0$, it does not follow that $\hat{\mathbf{Y}}_{RWd}$ and $\hat{\mathbf{Y}}_{ACd}$ will be close for all domains $d \in \mathfrak{D}$.

5. Discussion

The construction of weights that use both C-information and A-information has been considered by other authors. Zieschang (1990) considers a survey with two different simultaneously processed samples from the same population and having some variables in common. For each sample, he derives weights that are adjusted by the general regression estimator (GREG) technique so that estimated totals agree for comparable domains. He finds that efficiency is gained by letting the weights take common variables into account. Renssen and Nieuwenbroek (1997) consider the case of two simultaneous surveys. They build an adjusted GREG estimator from information on both C-variables and A-variables. The respective weights produce consistent estimates of totals for variables common to the two surveys. In the absence of known C-totals, they strive, by combining data over surveys, for highly precise estimates of the C-variable totals. This increases the efficiency of their estimator. Domain estimation is not examined. Further development for more precise population level estimates through combining information from multiple surveys is made in Merkuris (2004), and for domain level estimates in Merkuris (2010).

These references differ from our work in that they consider simultaneously processed surveys or samples, whereas we assume an already completed RFS survey that cannot be processed simultaneously with the PRS.

In a recent work by Dever and Valliant (2010) calibration information is also combined from different surveys. Their post-stratification estimator with estimated control totals from another survey can be seen, in certain setting, as a special case of our AC calibration. They do not consider domain estimation, and consistency between surveys is not their aim. But their message about increased variance due to the estimated control totals, is also visible in our simulation study for domains. Also, their linearized variance estimator has some common features with our approximate covariance matrix of the AC-calibrated domain estimators. Similarly, the incomplete post-stratification estimator by Deville et al. (1993) has connection points with our approach, in particular, if their post-strata were domains, and the domain counts under marginal restrictions were of interest.

Our calibration-based domain estimators (16) and (21) achieve consistency by distributing a difference $\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}}_A$ over domains with two different adjustment matrices. There are yet other estimation techniques that achieve consistency in a similar manner, but with other adjustment matrices. One example is the General Restriction (GR) estimator, proposed by Knottnerus (2003, Chapter 12) and studied in the domain estimation framework by Sõstra and Traat (2009). The GR estimator, optimal under certain assumptions, uses an adjustment matrix that involves the true, but unknown, covariance matrix of the A-calibrated domain estimators. This covariance matrix needs to be estimated; the resulting modified GR estimator is then no longer optimal but only asymptotically so. In addition, the design based estimation of the covariance matrix requires the second order inclusion probabilities, which makes this procedure less appealing for practice. By contrast, our AC-calibrated estimator (16) and our RW estimator (21) do not include any unknown quantities, furthermore, they require only first order inclusion probabilities.

6. Special cases

In this section we choose simple special cases to illustrate general formulas in Sections 2 and 3. We want to compare the AC-calibrated estimator with the RW estimator, to illustrate the differences. For Sections 6.2 and 6.3 we fix $\mathbf{z}'_{ACk} = q_k(\mathbf{x}'_k, \mathbf{y}'_k)$, $\mathbf{z}_{Ak} = q_k\mathbf{x}_k$ and $\mathbf{z}_{RWk} = q_k\mathbf{y}_k$.

6.1. The case with no A-information. Consider first the special case where no A-information is available in the PRS. Then the vector of domain totals $\mathbf{Y}_d = \sum_{U_d} \mathbf{y}_k$ can be unbiasedly estimated by $\hat{\mathbf{Y}}_d = \sum_{s_d} a_k \mathbf{y}_k$. If $\mathbf{y}_k$ consists of variables common with the RFS, we may want to estimate $\mathbf{Y}_d$ consistently with the RFS estimate $\hat{\mathbf{Y}}_0$. In the absence of A-information, AC-calibration reduces to C-calibration with weights obtained from (2) and (3) as

$$w_{Ck} = a_k(1 + \boldsymbol{\lambda}'_C \mathbf{z}_{Ck}), \text{ where } \boldsymbol{\lambda}'_C = (\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}})' \left(\sum_s a_k \mathbf{z}_{Ck} \mathbf{y}'_k \right)^{-1},$$

where $\mathbf{z}_{Ck}$ has the dimension of $\mathbf{y}_k$, and $\hat{\mathbf{Y}}$ is the HT estimator for the whole population, $\hat{\mathbf{Y}} = \sum_{d \in \mathfrak{D}} \hat{\mathbf{Y}}_d = \sum_s a_k \mathbf{y}_k$.

In the absence of A-information, the starting weights in the second step of the RW approach are the design weights a_k. We get from (19) and (20), with $w_{Ak} = a_k$, and with $\hat{\mathbf{Y}}_A$ replaced by $\hat{\mathbf{Y}}$,

$$w_{RWk} = a_k(1 + \mathbf{z}'_{RWk} \hat{\mathbf{Q}}_R^{-1} (\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}})),$$

where $\hat{\mathbf{Q}}_R = \sum_s a_k \mathbf{z}_{RWk} \mathbf{y}'_k$. Thus if $\mathbf{z}_{RWk} = \mathbf{z}_{Ck}$, C-calibration and RW-calibration produce the same weights.

6.2. The case of one-dimensional A-information. Consider the special case where $\mathbf{x}_k = x_k$ is a one-dimensional positive auxiliary variable in the PRS, with known total $X = \sum_U x_k$. Let $\boldsymbol{\gamma}_k : m \times 1$ be the indicator of m domains that were of interest in the RFS. The values of the common variable vector defined as $\mathbf{y}_k = \boldsymbol{\gamma}_k y_k$ are observed in the PRS. The vector of domain totals $\sum_U \boldsymbol{\gamma}_k y_k$ was estimated in the RFS by $\hat{\mathbf{Y}}_0$, and an HT estimator of it in the PRS would be $\hat{\mathbf{Y}} = \sum_s a_k \boldsymbol{\gamma}_k y_k$.

In the PRS, suppose there are $D \times m$ domains of interest, defined by the cross-classification of the RFS domains in $\mathfrak{G} = \{1, 2, \ldots, m\}$ with D "new" domains, $d \in \mathfrak{D}$. Let us choose $q_k = 1/x_k$. Then the A-calibrated weight is $w_{Ak} = a_k X/\hat{X}$, and with these weights the domain total $\mathbf{Y}_d = \sum_{U_d} \mathbf{y}_k = \sum_{U_d} \boldsymbol{\gamma}_k y_k$ is estimated by

$$\hat{\mathbf{Y}}_{Ad} = \hat{\mathbf{Y}}_d X/\hat{X} : \; m \times 1,$$

where $\hat{\mathbf{Y}}_d = \sum_{s_d} a_k \boldsymbol{\gamma}_k y_k$. The sum of $\hat{\mathbf{Y}}_{Ad}$ over $d \in \mathfrak{D}$ is the A-calibrated estimator

$$\hat{\mathbf{Y}}_A = \hat{\mathbf{Y}} X/\hat{X}. \tag{23}$$

Now when we impose also the calibration on $\hat{\mathbf{Y}}_0$, the A-calibrated weight w_{Ak} will be adjusted in the manner of (13). An analysis shows that

$$\hat{\mathbf{Q}} = \sum_s a_k \frac{y_k^2}{x_k} \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k' - \frac{\hat{\mathbf{Y}}\hat{\mathbf{Y}}'}{\hat{X}}. \tag{24}$$

This matrix has a simple form, a diagonal matrix minus a product of vectors, and can be explicitly inverted (Rao, 1973, p. 33). Finally, the weights (13) take the form

$$w_{ACk} = a_k \frac{X}{\hat{X}} + a_k \left(\frac{y_k}{x_k} \boldsymbol{\gamma}_k' - \frac{1}{\hat{X}} \hat{\mathbf{Y}}' \right) \hat{\mathbf{Q}}^{-1} (\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}}_A), \tag{25}$$

where $\hat{\mathbf{Y}}_A$ is is given in (23), and $\mathbf{Q}$ in (24).

By comparison, the RW weight for the choice $\mathbf{z}_{RWk} = q_k \mathbf{y}_k$ with $q_k = 1/x_k$ follows from (19),

$$w_{RWk} = a_k \frac{X}{\hat{X}} \{1 + \frac{y_k}{x_k} \boldsymbol{\gamma}_k' \hat{\mathbf{Q}}_R^{-1} (\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}}_A)\}, \tag{26}$$

with

$$\hat{\mathbf{Q}}_R = \frac{X}{\hat{X}} \sum_s a_k \frac{y_k^2}{x_k} \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k'.$$

Formulas (25) and (26) show two different weight systems, w_{ACk} and w_{RWk}. Both are consistent with the estimator $\hat{\mathbf{Y}}_0$ for the RFS domains, but only the w_{ACk} will reproduce the known total X.

6.3. The case with domain sizes as A-information. We consider the special case where both $\mathbf{y}_k = y_k$ and $\hat{\mathbf{Y}}_0 = \hat{Y}_0$, are one-dimensional. In the PRS we are interested in estimating y-totals in the domains $d = 1, 2, \ldots, D$ so that their sum is $\hat{Y}_0$. Let us fix $\mathbf{x}_k$ to be the domain indicator $\mathbf{x}_k = \boldsymbol{\delta}_k = (\delta_{1k}, \delta_{2k}, \ldots, \delta_{Dk})'$; $\delta_{dk} = 1$ if unit $k \in U_d$ and $\delta_{dk} = 0$ otherwise. The vector of known auxiliary totals is the vector of domain sizes, $\mathbf{X} = \sum_U \boldsymbol{\delta}_k = (N_1, N_2, \ldots, N_D)'$, and the corresponding HT estimator is $\hat{\mathbf{X}} = \sum_s a_k \boldsymbol{\delta}_k = (\hat{N}_1, \hat{N}_2, \ldots, \hat{N}_D)'$ with $\hat{N}_d = \sum_{s_d} a_k$. The A-calibrated weight (6) and the domain estimator (7) simplify to

$$w_{Ak} = a_k N_d / \hat{N}_d, \text{ if } k \in s_d, \text{ and } \hat{Y}_{Ad} = N_d \hat{Y}_d / \hat{N}_d,$$

where $\hat{Y}_d = \sum_{s_d} a_k y_k$. Summing over the domains gives for the whole population total the well-known post-stratified estimator,

$$\hat{Y}_A = \sum_{d \in \mathfrak{D}} N_d \hat{Y}_d / \hat{N}_d. \tag{27}$$

Let $\mathbf{z}_{ACk} = (\boldsymbol{\delta}'_k, y_k)'$. Then the residuals, $e_k = y_k - \hat{\mathbf{B}}'\boldsymbol{\delta}_k$, take a simple form,

$$e_k = y_k - \hat{Y}_d/\hat{N}_d, \text{ if } k \in s_d. \tag{28}$$

With $\hat{Y}_A$ and e_k given by (27) and (28), the AC-calibrated weight system, following from (13) and (15), is

$$w_{ACk} = a_k \frac{N_d}{\hat{N}_d} + \frac{a_k e_k}{\sum_{d\in\mathfrak{D}} \sum_{s_d} a_k e_k^2} (\hat{Y}_0 - \hat{Y}_A), \; k \in s_d, \tag{29}$$

To compare with repeated weighting, we fix $\mathbf{z}_{RWk} = y_k$. Then from (19) the RW weight system is

$$w_{RWk} = a_k \frac{N_d}{\hat{N}_d} \left\{ 1 + \frac{y_k}{\sum_{d\in\mathfrak{D}} (N_d/\hat{N}_d) \sum_{s_d} a_k y_k^2} (\hat{Y}_0 - \hat{Y}_A) \right\}, \; k \in s_d.$$

Again, the two weight systems differ. Both are calibrated on $\hat{Y}_0$, but only the system(29) will reproduce the domain sizes N_d.

7. Covariance matrix of the AC-calibrated estimator

The covariance matrix of the RW estimator for the frequency table case is developed in Knottnerus and van Duin (2006). In this section we derive the covariance matrix of the AC-calibrated estimators $\hat{\mathbf{Y}}_{ACd}, d \in \mathfrak{D}$, given by (16). More particularly, we find the covariance matrix of the linearized estimator. The matrices $\hat{\mathbf{Q}}$ and $\hat{\mathbf{Q}}_d$, given in (14) and (17), consist of the design-weighted sums. They are unbiased consistent estimators of the corresponding population sums. Consequently, $\hat{\mathbf{Q}}_d \hat{\mathbf{Q}}^{-1}$ is a consistent estimator of the population quantity $\mathbf{Q}_d \mathbf{Q}^{-1}$, where $\mathbf{Q} = \sum_U q_k \mathbf{y}_k \mathbf{e}'_k$ and $\mathbf{Q}_d = \sum_{U_d} q_k \mathbf{y}_k \mathbf{e}'_k$. The linearized form of the AC calibrated domain estimator (16) is

$$\hat{\mathbf{Y}}_{ACd} \approx \hat{\mathbf{Y}}_{Ad} + \mathbf{Q}_d \mathbf{Q}^{-1} (\hat{\mathbf{Y}}_0 - \hat{\mathbf{Y}}_A).$$

We present the covariance matrix of the linearized estimator in a single matrix expression taking into considerations all domains $d \in \mathfrak{D}$. Therefore, we stack the m-dimensional vectors $\hat{\mathbf{Y}}_{ACd}$ on top of each other to obtain the mD-dimensional vector denoted by $\hat{\boldsymbol{\theta}}_{AC}$. Analogously, we stack the A-calibrated domain estimators $\hat{\mathbf{Y}}_{Ad}$ on top of each other and obtain the mD-dimensional vector $\hat{\boldsymbol{\theta}}_A$. The additivity property of the A-calibrated domain estimators can be expressed as

$$\sum_{d\in\mathfrak{D}} \hat{\mathbf{Y}}_{Ad} = \hat{\mathbf{Y}}_A = \mathbf{R}\hat{\boldsymbol{\theta}}_A.$$

Here $\mathbf{R} = \mathbf{1}'_D \otimes diag(\mathbf{1}_m)$, where $\mathbf{1}_D$ and $\mathbf{1}_m$ are the vectors of ones with the indicated dimensionality, and $diag(\cdot)$ indicates a diagonal matrix. Now in matrix form,

$$\hat{\boldsymbol{\theta}}_{AC} \approx \hat{\boldsymbol{\theta}}_A + \mathbf{C}\mathbf{Q}^{-1}(\hat{\mathbf{Y}}_0 - \mathbf{R}\hat{\boldsymbol{\theta}}_A),$$

where $\mathbf{C} : mD \times m$ is a matrix with the $m \times m$ matrices $\mathbf{Q}_d$, $d \in \mathfrak{D}$, stacked on top of each other. Reorganizing terms, we get

$$\hat{\boldsymbol{\theta}}_{AC} \approx \mathbf{A}_1\hat{\boldsymbol{\theta}}_A + \mathbf{A}_2\hat{\mathbf{Y}}_0,$$

where $\mathbf{A}_1 = \mathbf{I} - \mathbf{C}\mathbf{Q}^{-1}\mathbf{R}$ and $\mathbf{A}_2 = \mathbf{C}\mathbf{Q}^{-1}$. Since $\hat{\mathbf{Y}}_0$ and $\hat{\boldsymbol{\theta}}_A$ are uncorrelated, due to different surveys, the desired approximate covariance matrix is

$$\text{Acov}(\hat{\boldsymbol{\theta}}_{AC}) = \mathbf{A}_1\text{cov}(\hat{\boldsymbol{\theta}}_A)\mathbf{A}_1' + \mathbf{A}_2\text{cov}(\hat{\mathbf{Y}}_0)\mathbf{A}_2'. \tag{30}$$

Since $\text{cov}(\hat{\boldsymbol{\theta}}_A)$ is the covariance matrix of the traditional A-calibrated domain estimators, its form, as well its estimator, are known (Estevao and Särndal 2006; Knottnerus van Duin 2006). The same can be said for $\text{cov}(\hat{\mathbf{Y}}_0)$ if a calibration estimator is used in the RFS. Furthermore, consistent estimates of $\mathbf{A}_1$ and $\mathbf{A}_2$ can be obtained, replacing $\mathbf{Q}_d$ by $\hat{\mathbf{Q}}_d$ and $\mathbf{Q}^{-1}$ by $\hat{\mathbf{Q}}^{-1}$. In the special case of a non-random $\hat{\mathbf{Y}}_0$, only the first term in (30) remains.

8. Simulation

The simulation reported in this section illustrates application (2) in Section 2. That is, the common variable vector has the form $\mathbf{y}_k = y_k\boldsymbol{\gamma}_k$, where the vector $\boldsymbol{\gamma}_k$ indicates one out of three possible regions. We use a population of $N = 2000$ persons with real data from the Estonian Labour Force Survey. The study variable value y_k is *salary* for person k (monthly, in thousands of kroons). Two cases of common variable information were studied: (1) the case of a constant (non-random) $\hat{\mathbf{Y}}_0$ composed of a priori known values of total salary by region, $\hat{\mathbf{Y}}_0 = (4999.0, 4614.8, 1396.3)'$, and (2) the case of random $\hat{\mathbf{Y}}_0$, composed of three unbiased regional estimates derived from an RFS, as explained later.

The PRS in this experiment involves two educational categories, $d = 1, 2$, not observed in the RFS. The objective in the PRS is to estimate total salary for each of the $3 \times 2 = 6$ sub-domains formed by the cross classification of region and education categories. The consistency that we require for each region is that the two sub-domain (education category) estimates add up to the common variable information for regions as specified in $\hat{\mathbf{Y}}_0$.

5000 PRS samples, each with size $n = 200$, were drawn by simple random sampling without replacement (SI). For each sample and each sub-domain we computed the HT estimator and an A-calibrated estimator based on $\mathbf{z}_{Ak} = \mathbf{x}_k = \boldsymbol{\gamma}_k$ whose known population total is the vector of sizes of regions, $(1019, 733, 248)'$.

In the simulation we also studied the AC estimator and the RW estimator. The AC weights were computed with (2) – (3) having $\mathbf{z}_{ACk} = (\mathbf{x}_k', \mathbf{y}_k')'$, and the RW weights with (19) having $\mathbf{z}_{RWk} = \mathbf{y}_k$. Both non-random and random $\hat{\mathbf{Y}}_0$ were used. With these weights we computed the AC estimator and the RW estimator for each sample and each sub-domain. For the case of the random $\hat{\mathbf{Y}}_0$, we associated with each of the 5000 PRS samples an independently drawn RFS sample (SI, sample size $n = 400$) from U, and the realized vector of estimates $\hat{\mathbf{Y}}_0$ was used as a random constraint vector in computing the AC weights and the RW weights. More specifically, $\hat{\mathbf{Y}}_0$ was calculated for each RFS sample as an A-calibrated estimator of the total salary by region, using here also $\mathbf{z}_{Ak} = \mathbf{x}_k = \gamma_k$ and the region sizes $\sum_U \gamma_k = (1019, 733, 248)'$ as auxiliary information. The standard deviations of the components of $\hat{\mathbf{Y}}_0$ were $(205.4, 236.5, 105.4)'$.

Table 1. Standard Deviations of Estimators for Region × Education Sub-domain Totals and Regional Domain Totals

				Non-rand. $\hat{\mathbf{Y}}_0$		Random $\hat{\mathbf{Y}}_0$	
Reg.	Educ.	HT	A	AC	RW	AC	RW
1	1	309	278	268	269	278	282
	2	410	359	268	269	315	299
2	1	287	256	262	265	266	272
	2	501	416	262	265	345	321
3	1	189	154	159	155	164	161
	2	234	199	159	155	184	169
1	-	455	308	0	0	205	205
2	-	541	349	0	0	236	236
3	-	291	158	0	0	105	105

The simulation validated the theoretical expectation that all estimators are unbiased or nearly so. Therefore the negligible biases are not presented. The simulation-based standard deviations of the estimators in the six sub-domains are given in Table 1. It shows that both the AC and the RW estimators are consistent with $\hat{\mathbf{Y}}_0$; by construction, they have zero standard deviations at the regional level for the non-random constraint case, and the standard deviations for the random case are equal, determined solely by the variability of $\hat{\mathbf{Y}}_0$ in the RFS.

Compared to the HT estimator, the A-calibrated estimator is as expected more efficient in every domain. Neither is consistent with $\hat{\mathbf{Y}}_0$. However, AC and RW have such consistency. They are in all cells more efficient than

HT, but not always more efficient than the A-calibrated estimator. The latter is a bit unexpected for the case of constant $\hat{\mathbf{Y}}_0$, although the loss of efficiency is then small. This might be caused by very different variances of the A-calibrated estimators in the educational sub-domains. After putting on summation restriction, the estimators are forced to take on equal variances inside regions which has caused an huge decrease for the higher variance and, in some cases, a small increase for the smaller variance.

For the six sub-domains, the AC estimator and the RW estimator have very similar standard deviations in the constant $\hat{\mathbf{Y}}_0$ case, while for the random case, the standard deviations are increased, for both AC and RW, and the differences between them become more pronounced. Neither estimator is superior to the other over all sub-domains.

Additional simulations are reported in Traat (2010) for variations of the situation considered in this section, with alternative ways of specifying the finite populations from which repeated samples are drawn. The results from these simulations confirm the general pattern of the results reported here.

Appendix: Derivation of the relationship between AC- and A-calibrated weights

We consider the instrument vector, $\mathbf{z}'_{ACk} = q_k(\mathbf{x}'_k, \mathbf{y}'_k)$. With this vector the $(p+m)\times(p+m)$ matrix $\mathbf{M}$ in (3) has the following blocks:

$$\mathbf{M} = \sum_s a_k \mathbf{z}_{ACk}(\mathbf{x}'_k, \mathbf{y}'_k) = \begin{pmatrix} \hat{\mathbf{T}}_{\mathbf{xx}} & \hat{\mathbf{T}}_{\mathbf{xy}} \\ \hat{\mathbf{T}}'_{\mathbf{xy}} & \hat{\mathbf{T}}_{\mathbf{yy}} \end{pmatrix},$$

where $\hat{\mathbf{T}}_{\mathbf{xx}}$ and $\hat{\mathbf{T}}_{\mathbf{xy}}$ are given in (11), and

$$\hat{\mathbf{T}}_{\mathbf{yy}} = \sum_s a_k q_k \mathbf{y}_k \mathbf{y}'_k : \; m \times m.$$

The inverse of the block matrix $\mathbf{M}$ is (Kollo and von Rosen 2005, p. 74)

$$\mathbf{M}^{-1} = \begin{pmatrix} \hat{\mathbf{T}}_{\mathbf{xx}}^{-1} + \hat{\mathbf{B}}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{B}}' & -\hat{\mathbf{B}}\hat{\mathbf{Q}}^{-1} \\ -\hat{\mathbf{Q}}^{-1}\hat{\mathbf{B}}' & \hat{\mathbf{Q}}^{-1} \end{pmatrix}, \qquad (A.1)$$

where $\hat{\mathbf{B}}$ is given in (10), and

$$\hat{\mathbf{Q}} = \hat{\mathbf{T}}_{\mathbf{yy}} - \hat{\mathbf{T}}'_{\mathbf{xy}}\hat{\mathbf{B}} : \; m \times m. \qquad (A.2)$$

In arriving at $(A.1)$ we have also used the symmetricity of $\mathbf{M}$ and the formula for inverting a sum of matrices. The general expression (2) for w_{ACk} contains the term

$$\boldsymbol{\lambda}'_{AC}\mathbf{z}_{ACk} = (\mathbf{X}' - \hat{\mathbf{X}}', \; \hat{\mathbf{Y}}'_0 - \hat{\mathbf{Y}}')\mathbf{M}^{-1}\mathbf{z}_{ACk}, \qquad (A.3)$$

where

$$\mathbf{M}^{-1}\mathbf{z}_{ACk} = q_k \begin{pmatrix} (\hat{\mathbf{T}}_{\mathbf{xx}}^{-1} + \hat{\mathbf{B}}\hat{\mathbf{Q}}^{-1}\hat{\mathbf{B}}')\mathbf{x}_k - \hat{\mathbf{B}}\hat{\mathbf{Q}}^{-1}\mathbf{y}_k \\ -\hat{\mathbf{Q}}^{-1}\hat{\mathbf{B}}'\mathbf{x}_k + \hat{\mathbf{Q}}^{-1}\mathbf{y}_k \end{pmatrix}.$$

The expression simplifies with the aid of the predicted values and the residuals in (12):

$$\mathbf{M}^{-1}\mathbf{z}_{ACk} = q_k \begin{pmatrix} \hat{\mathbf{T}}_{\mathbf{xx}}^{-1}\mathbf{x}_k - \hat{\mathbf{B}}\hat{\mathbf{Q}}^{-1}\mathbf{e}_k \\ \hat{\mathbf{Q}}^{-1}\mathbf{e}_k \end{pmatrix}.$$

Now $(A.3)$ can be written as

$$\boldsymbol{\lambda}_{AC}'\mathbf{z}_{ACk} = \boldsymbol{\lambda}_A' q_k \mathbf{x}_k - q_k(\hat{\mathbf{Y}}_A - \hat{\mathbf{Y}}_0)'\hat{\mathbf{Q}}^{-1}\mathbf{e}_k, \tag{A.4}$$

where the vector $\boldsymbol{\lambda}_A$ is as in the A-calibrated weight (6) with $\mathbf{z}_{Ak} = q_k\mathbf{x}_k$, and $\hat{\mathbf{Y}}_A$ is given in (9). Finally, it follows that the AC-calibrated weight $w_{ACk} = a_k(1 + \boldsymbol{\lambda}_{AC}'\mathbf{z}_{ACk})$ can be expressed as in (13). For convenience, we have transposed the last matrix product in $(A.4)$. The expression (14) of the matrix $\hat{\mathbf{Q}}$ follows by inserting $\hat{\mathbf{T}}_{\mathbf{yy}}$ and $\hat{\mathbf{T}}_{\mathbf{xy}}'$ into $(A.2)$.

Acknowledgement. We thank the anonymous referee for the motivating discussion around the paper.

References

Dever, J. A., Valliant, R. (2010), *A comparison of variance estimators for poststratification to estimated control totals*, Surv. Methodol. **36**, 45–56.

Deville, J. C., Särndal, C.-E. (1992), *Calibration estimators in survey sampling*, J. Amer. Statist. Assoc. **87**, 376–382.

Deville, J. C., Särndal, C.-E., Sautory, O. (1993), *Generalized raking procedures in survey sampling*, J. Amer. Statist. Assoc. **88**, 1013–1020.

Estevao, V. M., Särndal, C.-E. (2004), *Borrowing strength is not the best technique within a wide class of design-consistent domain estimators*, J. Official Statist. **20**, 645–660.

Estevao, V. M., Särndal, C.-E. (2006), *Survey estimates by calibration on complex auxiliary information*, Internat. Statist. Rev. **74**, No.2, 127–147.

Houbiers, M. (2004), *Towards a social statistical database and unified estimates at Statistics Netherlands*, J. Official Statist. **20**, 55–75.

Knottnerus, P. (2003), *Sample Survey Theory. Some Pythagorean Perspectives*, Springer, New York.

Knottnerus, P., van Duin, C. (2006), *Variances in repeated weighting with an application to the Dutch Labour Force Survey*, J. Official Statist. **22**, 565–584.

Kollo, T., von Rosen, D. (2005), *Advanced Multivariate Statistics with Matrices*, Springer, Dordrecht.

Kott, P. S. (2004), *Comment on Demnati and Rao: linearization variance estimators for survey data*, Surv. Methodol. **30**, 27–28.

Kroese, A. H., Renssen, R. H. (1999), *Weighting and imputation at Statistics Netherland*; In: Proceedings of the IASS Conference on Small Area Estimation, Riga, pp. 109–120.

Merkouris, T. (2004), *Combining independent regression estimators from multiple surveys*, J. Amer. Statist. Assoc. **99**, 1131–1139.

Merkouris, T. (2010), *Combining information from multiple surveys by using regression for efficient small domain estimation*, J. Roy. Statist. Soc. Ser. B **72**, 27–48.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, Wiley, New York.

Rencher, A. C., (1998) *Multivariate Statistical Inference and Applications*, Wiley, New York.

Renssen, R. H., Nieuwenbroek, N. J. (1997), *Aligning estimates for common variables in two or more sample surveys*, J. Amer. Statist. Assoc. **92**, 368–374.

Sõstra, K., Traat, I. (2009), *Optimal domain estimation under summation restriction*, J. Statist. Plann. Inference **139**, 3928–3941.

Särndal, C.-E. (2007), *The calibration approach in survey theory and practice*, Surv. Methodol. **33**, 99–119.

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

Traat, I. (2010), *Comparison of calibration-based consistent domain estimators*; In: Carlson, Nyquist and Villani (eds), Official Statistics - Methodology and Applications in Honour of Daniel Thorburn, pp. 137–147.

Zieschang, K. D. (1990), *Sample weighting methods and estimation of totals in the consumer expenditure survey*, J. Amer. Statist. Assoc. **85**, 986–1001.

Statistics Sweden, SE-70189 Örebro, Sweden
E-mail address: carl.sarndal@scb.se

Institute of Mathematical Statistics, University of Tartu, 50409 Tartu, Estonia
E-mail address: imbi.traat@ut.ee