Nonlinear Analysis: Modelling and Control, 2010, Vol. 15, No. 3, 325–340

Estimation of the finite population covariance using calibration

A. Plikusas, D. Pumputis

Institute of Mathematics and Informatics, Vilnius University
Akademijos str. 4, LT-08663 Vilnius, Lithuania

Vilnius Pedagogical University
Studentų str. 39, LT-08106 Vilnius, Lithuania

plikusas@ktl.mii.lt; dpumputis@vpu.lt

Received: 2009-11-11 **Revised:** 2010-04-30 **Published online:** 2010-09-17

Abstract. Estimators of the finite population covariance with several systems of weights are considered. New calibrated estimators of the finite population covariance (variance) are derived, using two and three weighting systems that are defined by various calibration equations and loss functions. The expressions of approximate variance for some of these estimators are presented. The estimators derived are compared by simulation. Finally, it is shown how the calibrated estimators of the covariance may be applied in regression estimation of the finite population total.

Keywords: finite-population covariance, auxiliary variables, calibration, calibrated weights, regression estimator.

1 Introduction

Survey statisticians are always concerned with the improvement of methods for estimation of the finite population total, mean, proportion and other parameters. Auxiliary information may be used for that purpose. The estimators that use auxiliary variables are often much more accurate than the standard ones. The calibrated estimators belong to this class of estimators. The idea of the calibration technique for estimating the population totals is presented in [1].

The estimation of more complicated parameters using the calibration methods is not widely studied in the literature. The calibrated estimator of the ratio of two totals is considered by Plikusas [2], Krapavickaitė and Plikusas [3]. Calibration estimation for quantiles is studied by Harms and Duchesne [4], Rueda et al. [5]. Sitter and Wu [6] proposed a model-calibrated method to estimate the quadratic finite population functions. Singh et al. [7] applied the calibration technique in the estimation of variance of the Horvitz–Thompson estimator.

Some calibrated estimators of the finite population covariance are introduced in the paper [8]. They use one weighting system, which is defined using various calibration equations and loss functions. In the following section, we recall these estimators and provide some new estimators with several systems of weights.

An overview of the calibration theory and application of the calibrated estimators in survey practice is given by Särndal in [9].

2 Calibrated estimators of the finite population covariance

2.1 Estimators with one system of weights

Consider a finite population $\mathcal{U} = \{u_1, u_2, \dots, u_N\}$ of N elements. Without loss of generality, we can assume $\mathcal{U} = \{1, 2, \dots, N\}$. Let y and z be two study variables defined on the population $\mathcal{U}$, taking real nonnegative values $y_1, \dots, y_N$ and $z_1, \dots, z_N$, respectively. The values of the variables y and z are not known.

Let the covariance

$$\mathrm{Cov}(y, z) = \frac{1}{N-1} \sum_{k=1}^{N} \left(y_k - \frac{1}{N} \sum_{k=1}^{N} y_k \right) \left(z_k - \frac{1}{N} \sum_{k=1}^{N} z_k \right)$$

be the parameter of interest.

Denote by s, $s \subset \mathcal{U}$, a probability sample set drawn from the population $\mathcal{U}$, by π_k – the inclusion probability of element k into the sample s, and by $d_k = \frac{1}{\pi_k}$ – sample design weight of element k, $k = 1, 2, \dots, N$.

In the case of none auxiliary information, we can estimate the population covariance using the well-known only design based estimator

$$\widehat{\mathrm{Cov}}(y, z) = \frac{1}{N-1} \sum_{k \in s} d_k \left(y_k - \frac{1}{N} \sum_{k \in s} d_k y_k \right) \left(z_k - \frac{1}{N} \sum_{k \in s} d_k z_k \right). \tag{1}$$

It is considered in Särndal, Swensson and Wretman's book [10, p. 187].

The weights d_k of estimator (1) may be modified using auxiliary variables and calibration approach to obtain estimators with smaller variance. Denote the auxiliary variables taking values $a_1, \dots, a_N$ and $b_1, \dots, b_N$ by a and b. It should be noted that, depending on the calibration equations used, in addition to the values of auxiliary variables for sampled elements, only the covariance of auxiliary variables, or covariance and totals of these auxiliary variables are needed for the construction of calibrated estimators. In the paper [8], we apply the calibration technique to modify the design weights d_k, provided that the auxiliary variables are given. We consider here the calibrated estimator of the covariance of the following shape

$$\widehat{\mathrm{Cov}}_{1w}(y, z) = \frac{1}{N-1} \sum_{k \in s} w_k \left(y_k - \frac{1}{N} \sum_{k \in s} w_k y_k \right) \left(z_k - \frac{1}{N} \sum_{k \in s} w_k z_k \right). \tag{2}$$

The new (calibrated) weights w_k are defined under the following conditions:

a) The weights w_k satisfy some calibration equation;

b) The distance between the weights d_k and w_k is minimal according to some loss function $L(w,d)$.

Conditions a) and b) can be specified in different ways. The following calibration equations are used in the paper:

$$\text{I.}\qquad \frac{1}{N-1}\sum_{k\in s} w_k(a_k-\widehat{\mu}_{aw})(b_k-\widehat{\mu}_{bw}) = \mathrm{Cov}(a,b), \tag{3}$$

$$\widehat{\mu}_{aw} = \frac{1}{N}\sum_{k\in s} w_k a_k, \quad \widehat{\mu}_{bw} = \frac{1}{N}\sum_{k\in s} w_k b_k.$$

$$\text{II.}\qquad \frac{1}{N-1}\sum_{k\in s} w_k(a_k-\mu_a)(b_k-\mu_b) = \mathrm{Cov}(a,b), \tag{4}$$

$$\mu_a = \frac{1}{N}\sum_{k=1}^{N} a_k, \quad \mu_b = \frac{1}{N}\sum_{k=1}^{N} b_k,$$

$$\text{III.}\qquad \sum_{k\in s} w_k a_k = \sum_{k=1}^{N} a_k, \quad \sum_{k\in s} w_k b_k = \sum_{k=1}^{N} b_k. \tag{5}$$

The loss function

$$L_1(w,d) = \sum_{k\in s} \frac{(w_k-d_k)^2}{d_k q_k} \tag{6}$$

and some other ones are applied in the final specification of calibrated weights w_k. Here $q_k,\ k\in\mathcal{U}$, are free additional positive constants or additional weights. The calibrated estimators can be modified by choosing q_k.

The calibrated estimators with one weighting system are denoted by $\widehat{\mathrm{Cov}}_{1w}^{(non)}(y,z)$, $\widehat{\mathrm{Cov}}_{1w}^{(lin)}(y,z)$, $\widehat{\mathrm{Cov}}_{1w}^{(tot)}(y,z)$, depending on the calibration equation used: (3), (4), (5). For example, $\widehat{\mathrm{Cov}}_{1w}^{(non)}(y,z)$ denotes the estimator whose weights w_k satisfy the calibration equation (3) and minimize the loss function (6).

Next, we extend the definitions, given in this subsection, to the case of multiple weighting systems.

2.2 Estimators with several systems of weights

Let us consider some other, more general estimators of the finite population covariance, which are constructed using several weighting systems. The new calibrated estimators of the covariance are of the following shape:

$$\widehat{\mathrm{Cov}}_{mw}(y,z) = \frac{1}{N-1}\sum_{k\in s} w_k^{(1)}\left(y_k-\frac{1}{N}\sum_{l\in s} w_l^{(2)} y_l\right)\left(z_k-\frac{1}{N}\sum_{l\in s} w_l^{(3)} z_l\right). \tag{7}$$

Several calibration equations may be used for definition of the calibrated weights $w_k^{(1)}$, $w_k^{(2)}$, $w_k^{(3)}$. Let us consider some of them.

Case 1. The nonlinear calibration equation

$$\widehat{\mathrm{Cov}}_{mw}(a,b) = \mathrm{Cov}(a,b). \tag{8}$$

Case 2. The systems of weights $w_k^{(1)}$, $w_k^{(2)}$, $w_k^{(3)}$ are defined by calibration equations:

$$\frac{1}{N-1}\sum_{k\in s} w_k^{(1)}(a_k-\mu_a)(b_k-\mu_b) = \mathrm{Cov}(a,b), \tag{9}$$

$$\sum_{k\in s} w_k^{(2)} a_k = \sum_{k=1}^{N} a_k, \quad \sum_{k\in s} w_k^{(3)} b_k = \sum_{k=1}^{N} b_k. \tag{10}$$

Case 3. The first system of weights $w_k^{(1)}$ is defined by the nonlinear calibration equation (3). Calibration equations (10) define the other two systems of the weights $w_k^{(2)}$ and $w_k^{(3)}$.

Case 4. We can consider the estimator of covariance which uses two systems of weights:

$$\widehat{\mathrm{Cov}}_{mw}(y,z) = \frac{1}{N-1}\sum_{k\in s} w_k^{(1)}\Big(y_k - \frac{1}{N}\sum_{l\in s} w_l^{(2)} y_l\Big)\Big(z_k - \frac{1}{N}\sum_{l\in s} w_l^{(2)} z_l\Big). \tag{11}$$

The first system of weights $w_k^{(1)}$ is defined by equation (9), whereas the second system $w_k^{(2)}$ satisfies the following equations

$$\sum_{k\in s} w_k^{(2)} a_k = \sum_{k=1}^{N} a_k, \quad \sum_{k\in s} w_k^{(2)} b_k = \sum_{k=1}^{N} b_k. \tag{12}$$

Case 5. We can use another combination of two systems of calibrated weights: the first one $w_k^{(1)}$ satisfies nonlinear calibration equation (3), where the system $w_k^{(2)}$ is defined by (12).

Case 6. The system of weights $w_k^{(1)}$ satisfies equation (9), whereas the system $w_k^{(2)}$ is obtained using nonlinear calibration equation (3).

The following loss function may be used for final definition of calibrated weights:

$$L(w,d) = \sum_{i\in \mathbf{r}}\sum_{k\in s} \frac{(w_k^{(i)}-d_k)^2}{d_k\, q_k}, \tag{13}$$

where $\mathbf{r} = \{1,2,3\}$, if the estimators with three weighting systems are considered, and $\mathbf{r} = \{1,2\}$, in the case of two weighting systems.

The first case is most complicated analytically, the expressions for the approximate iterative solutions of calibration equation (8) are cumbersome.

The following proposition defines the weights $w_k^{(1)}$, $w_k^{(2)}$, $w_k^{(3)}$ of estimator (7) for all the six cases mentioned in this subsection.

Let us introduce some additional notation:

$$\widehat{\mu}_{aw}^{(i)} = \frac{1}{N}\sum_{k\in s} w_k^{(i)} a_k, \quad \widehat{\mu}_{bw}^{(i)} = \frac{1}{N}\sum_{k\in s} w_k^{(i)} b_k, \quad \widehat{N}_w^{(i)} = \sum_{k\in s} w_k^{(i)}, \quad i = 1,2,3.$$

Proposition 1. *The weights $w_k^{(i)}$, $k \in s$, $i = 1,2,3$, which satisfy calibration equation* (8) *and minimize loss function* (13)*, satisfy the equation $w_k^{(i)} = d_k u_k^{(i)}$. Here $u_k^{(i)} = 1 + \lambda q_k c_k^{(i)}$,*

$$\begin{aligned}
c_k^{(1)} &= \big(a_k - \widehat{\mu}_{aw}^{(2)}\big)\big(b_k - \widehat{\mu}_{bw}^{(3)}\big),\\
c_k^{(2)} &= -a_k\bigg(\widehat{\mu}_{bw}^{(1)} - \frac{\widehat{N}_w^{(1)}}{N}\widehat{\mu}_{bw}^{(3)}\bigg),\\
c_k^{(3)} &= -b_k\bigg(\widehat{\mu}_{aw}^{(1)} - \frac{\widehat{N}_w^{(1)}}{N}\widehat{\mu}_{aw}^{(2)}\bigg),\\
\lambda &= \widehat{A}\bigg(\sum_{k\in s} d_k q_k\big(a_k b_k c_k^{(1)} + c_k^{(2)} - c_k^{(3)}\big)\bigg)^{-1},\\
\widehat{A} &= (N-1)\mathrm{Cov}(a,b) + N\widehat{\mu}_{bw}^{(3)}\bigg(\widehat{\mu}_{aw}^{(1)} - \frac{\widehat{N}_w^{(1)}}{N}\widehat{\mu}_{aw}^{(2)}\bigg)\\
&\quad + N\widehat{\mu}_{aw}^{(2)}\widehat{\mu}_{bw}^{(1)} + \widehat{N}_w^{(2)} - \widehat{N}_w^{(3)} - \sum_{k\in s} d_k a_k b_k.
\end{aligned}$$

In Cases 2, 4, *and* 6*, the first system of weights $w_k^{(1)}$ is defined by the equations:*

$$w_k^{(1)} = d_k\Bigg(1 + q_k\bigg(\sum_{l=1}^{N} c_l - \sum_{l\in s} d_l c_l\bigg)\bigg(\sum_{l\in s} d_l q_l c_l^2\bigg)^{-1} c_k\Bigg), \tag{14}$$

where $c_k = (a_k - \mu_a)(b_k - \mu_b)$.

The equations

$$w_k^{(i)} = d_k\Bigg(1 + \widehat{A}\bigg(\sum_{l\in s} d_l q_l f_l a_l b_l\bigg)^{-1} q_k f_k\Bigg)$$

define the first system of weights $w_k^{(1)}$ in Cases 3 *and* 5*, and the system $w_k^{(2)}$ in Case* 6*. Here*

$$\widehat{A} = (N-1)\mathrm{Cov}(a,b) + N\bigg(2 - \frac{\widehat{N}_w^{(i)}}{N}\bigg)\widehat{\mu}_{aw}^{(i)}\widehat{\mu}_{bw}^{(i)} - \sum_{k\in s} d_k a_k b_k,$$

$$f_k = (a_k - \widehat{\mu}_{aw}^{(i)})(b_k - \widehat{\mu}_{bw}^{(i)}) - \left(1 - \frac{\widehat{N}_w^{(i)}}{N}\right)\left(\frac{\widehat{\mu}_{aw}^{(i)}}{a_k} + \frac{\widehat{\mu}_{bw}^{(i)}}{b_k}\right)a_k b_k, \quad i = 1, 2, 3.$$

In Cases 2 *and* 3*, the system of weights* $w_k^{(2)}$ *is defined by*

$$w_k^{(2)} = d_k\left(1 + q_k\left(\sum_{l=1}^{N} a_l - \sum_{l\in s} d_l a_l\right)\left(\sum_{l\in s} d_l q_l a_l^2\right)^{-1} a_k\right), \tag{15}$$

and the system $w_k^{(3)}$ *is defined by the same equation* (15) *by replacing* a_k *with* b_k*.*

In Cases 4 *and* 5*, the second system of weights* $w_k^{(2)}$ *satisfies these equations:*

$$w_k^{(2)} = d_k\left(1 + q_k\left(\sum_{l=1}^{N} \mathbf{x}_l' - \sum_{l\in s} d_l \mathbf{x}_l'\right)\left(\sum_{l\in s} d_l q_l \mathbf{x}_l \mathbf{x}_l'\right)^{-1} \mathbf{x}_k\right),$$

where $\mathbf{x}_k = (a_k, b_k)'$*.*

Proof. Let us take the loss function (13) and calibration equation (8), and define the Lagrange function

$$\Lambda = \sum_{i\in\{1,2,3\}} \sum_{k\in s} \frac{(w_k^{(i)} - d_k)^2}{d_k\, q_k} - \lambda\left(\frac{1}{N-1}\sum_{k\in s} w_k^{(1)}\left(a_k - \widehat{\mu}_{aw}^{(2)}\right)\left(b_k - \widehat{\mu}_{bw}^{(3)}\right) - \mathrm{Cov}(a,b)\right).$$

By solving the equations

$$\frac{\partial \Lambda}{\partial w_k^{(i)}} = 0, \quad i = 1, 2, 3, \ \ k \in s,$$

we get

$$w_k^{(i)} = d_k\left(1 + \frac{1}{2(N-1)}\lambda q_k c_k^{(i)}\right). \tag{16}$$

Hence

$$w_k^{(1)} a_k b_k + w_k^{(2)} - w_k^{(3)} = d_k\left(a_k b_k + \frac{1}{2(N-1)}\lambda q_k\left(a_k b_k c_k^{(1)} + c_k^{(2)} - c_k^{(3)}\right)\right). \tag{17}$$

Then, summing derived equations (17) over the sample elements and taking into account the calibration equation (8), we get the expression for λ. Inserting this expression into (16), we get iterative equations for $w_k^{(i)}$, $i = 1, 2, 3$.

The proof for other cases of calibration equations and the loss function is similar. □

The calibrated estimators of covariance, corresponding to the cases of calibration equations mentioned above, are denoted by $\widehat{\mathrm{Cov}}^{(i)}_{mw}(y,z)$, $i = 1,2,3,4,5,6$. For example, $\widehat{\mathrm{Cov}}^{(1)}_{mw}(y,z)$ denotes the estimator which uses three weighting systems that satisfy calibration equation (8) and minimize loss function (13).

3 Estimation of variance

The presented calibrated estimators of the covariance are complicated enough, there is no explicit expression for the calibrated weights in some cases.

Provided calibration equations (4), (5) are used for the definition of calibrated weights, we get the explicit solution of the calibration problem and the Taylor linearization technique may be applied to derive an approximate variance of estimators. The following proposition gives an approximate variance for the estimator $\widehat{\mathrm{Cov}}^{(4)}_{mw}(y,z)$.

Proposition 2. *The approximate variance of the estimator*

$$\widehat{\mathrm{Cov}}^{(4)}_{mw}(y,z) = \frac{1}{N-1}\sum_{k\in s} w_k^{(1)}\Big(y_k - \frac{1}{N}\sum_{i\in s} w_i^{(2)} y_i\Big)\Big(z_k - \frac{1}{N}\sum_{i\in s} w_i^{(2)} z_i\Big), \quad (18)$$

the weights $w_k^{(1)}$, $w_k^{(2)}$ of which satisfy the corresponding equations (9), (12) *and minimize the loss function $L(w,d)$ defined by equation* (13)*, is given by*

$$\mathrm{AVar}\big(\widehat{\mathrm{Cov}}^{(4)}_{mw}(y,z)\big) = \frac{1}{(N-1)^2}\sum_{k=1}^{N}\sum_{l=1}^{N}\frac{\pi_{kl}-\pi_k\pi_l}{\pi_k\pi_l}e_k e_l,$$

where π_{kl}, $k,l = 1,\dots,N$, is the inclusion probability of the elements k and l into the sample,

$$e_k = (y_k-\mu_y)(z_k-\mu_z) + Bc_k,$$

$$c_k = (a_k-\mu_a)(b_k-\mu_b), \quad \mu_y = \frac{1}{N}\sum_{k=1}^{N} y_k, \quad \mu_z = \frac{1}{N}\sum_{k=1}^{N} z_k,$$

$$B = t_{qcc}^{-1}(-t_{qcyz} + \mu_z t_{qyc} + \mu_y t_{qcz} - \mu_y\mu_z t_{qc}),$$

$$t_{qcc} = \sum_{k=1}^{N} q_k c_k^2, \quad t_{qcyz} = \sum_{k=1}^{N} q_k c_k y_k z_k, \quad t_{qcy} = \sum_{k=1}^{N} q_k c_k y_k,$$

$$t_{qcz} = \sum_{k=1}^{N} q_k c_k z_k, \quad t_{qc} = \sum_{k=1}^{N} q_k c_k.$$

Proof. Estimator (18) can be expressed in the following form

$$\widehat{\mathrm{Cov}}^{(4)}_{mw}(y,z) = \frac{1}{N-1}\Bigg(\sum_{k\in s} w_k^{(1)} y_k z_k - \frac{1}{N}\sum_{k\in s} w_k^{(1)} y_k \sum_{k\in s} w_k^{(2)} z_k - \frac{1}{N}\sum_{k\in s} w_k^{(1)} z_k \sum_{k\in s} w_k^{(2)} y_k + \frac{1}{N^2}\sum_{k\in s} w_k^{(1)} \sum_{k\in s} w_k^{(2)} y_k \sum_{k\in s} w_k^{(2)} z_k\Bigg). \tag{19}$$

It follows from Proposition 1 that the weights $w_k^{(1)}$, $w_k^{(2)}$ that satisfy the corresponding equations (9), (12) and minimize the loss function $L(w,d)$, are given by

$$w_k^{(1)} = d_k\left(1 + \left(\sum_{l=1}^{N} c_l - \sum_{l\in s} d_l c_l\right)\left(\sum_{l\in s} d_l q_l c_l^2\right)^{-1} q_k c_k\right), \tag{20}$$

$$w_k^{(2)} = d_k\left(1 + \left(\sum_{l=1}^{N} \mathbf{x}_l' - \sum_{l\in s} d_l \mathbf{x}_l'\right)\left(\sum_{l\in s} d_l q_l \mathbf{x}_l \mathbf{x}_l'\right)^{-1} q_k \mathbf{x}_k\right), \tag{21}$$

where $\mathbf{x}_l = (a_l, b_l)'$.

Inserting expressions (20), (21) of weights into (19) we find:

$$\begin{aligned}
&\widehat{\mathrm{Cov}}^{(4)}_{mw}(y,z)\\
&= \frac{1}{N-1}\Bigg(\hat{t}_{yz} + (t_c - \hat{t}_c)\hat{t}_{qcc}^{-1}\hat{t}_{qcyz}\\
&\quad - \frac{1}{N}\big(\hat{t}_y + (t_c - \hat{t}_c)\hat{t}_{qcc}^{-1}\hat{t}_{qcy}\big)\big(\hat{t}_z + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}_{q\mathbf{x}}^{-1}\hat{\mathbf{t}}_{qz\mathbf{x}}\big)\\
&\quad - \frac{1}{N}\big(\hat{t}_z + (t_c - \hat{t}_c)\hat{t}_{qcc}^{-1}\hat{t}_{qcz}\big)\big(\hat{t}_y + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}_{q\mathbf{x}}^{-1}\hat{\mathbf{t}}_{qy\mathbf{x}}\big)\\
&\quad + \frac{1}{N^2}\big(\widehat{N} + (t_c - \hat{t}_c)\hat{t}_{qcc}^{-1}\hat{t}_{qc}\big)\big(\hat{t}_y + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}_{q\mathbf{x}}^{-1}\hat{\mathbf{t}}_{qy\mathbf{x}}\big)\\
&\quad \times \big(\hat{t}_z + (\mathbf{t}_{\mathbf{x}}' - \hat{\mathbf{t}}_{\mathbf{x}}')\widehat{\mathbf{A}}_{q\mathbf{x}}^{-1}\hat{\mathbf{t}}_{qz\mathbf{x}}\big)\Bigg)\\
&= f\big(\hat{t}_{yz}, \hat{t}_c, \hat{t}_{qcc}, \hat{t}_{qcyz}, \hat{t}_y, \hat{t}_{qyc}, \hat{t}_z, \hat{\mathbf{t}}_{\mathbf{x}}, \widehat{\mathbf{A}}_{q\mathbf{x}}, \hat{\mathbf{t}}_{qz\mathbf{x}}, \hat{t}_{qcz}, \hat{\mathbf{t}}_{qy\mathbf{x}}, \widehat{N}, \hat{t}_{qc}\big).
\end{aligned} \tag{22}$$

Here

$$\hat{t}_{yz} = \sum_{k\in s} d_k y_k z_k, \quad t_c = \sum_{k=1}^{N} c_k, \quad \hat{t}_c = \sum_{k\in s} d_k c_k, \quad \hat{t}_{qcc} = \sum_{k\in s} d_k q_k c_k^2,$$

$$\hat{t}_{qcyz} = \sum_{k\in s} d_k q_k c_k y_k z_k, \ \ \hat{t}_y = \sum_{k\in s} d_k y_k, \ \ \hat{t}_{qyc} = \sum_{k\in s} d_k q_k c_k y_k, \ \ \hat{t}_z = \sum_{k\in s} d_k z_k,$$

$$\mathbf{t}_{\mathbf{x}} = \sum_{k=1}^{N} \mathbf{x}_k, \quad \hat{\mathbf{t}}_{\mathbf{x}} = \sum_{k\in s} d_k \mathbf{x}_k, \quad \widehat{\mathbf{A}}_{q\mathbf{x}} = \sum_{k\in s} d_k q_k \mathbf{x}_k \mathbf{x}_k^{'}, \quad \hat{\mathbf{t}}_{qz\mathbf{x}} = \sum_{k\in s} d_k q_k z_k \mathbf{x}_k,$$
$$\hat{t}_{qcz} = \sum_{k\in s} d_k q_k c_k z_k, \quad \hat{\mathbf{t}}_{qy\mathbf{x}} = \sum_{k\in s} d_k q_k y_k \mathbf{x}_k, \quad \widehat{N} = \sum_{k\in s} d_k, \quad \hat{t}_{qc} = \sum_{k\in s} d_k q_k c_k.$$

The estimators $\hat{t}_{yz}$, $\hat{t}_c$, $\hat{t}_{qcc}$, $\hat{t}_{qcyz}$, $\hat{t}_y$, $\hat{t}_{qcy}$, $\hat{t}_z$, $\hat{\mathbf{t}}_{\mathbf{x}}$, $\widehat{\mathbf{A}}_{q\mathbf{x}}$, $\hat{\mathbf{t}}_{qz\mathbf{x}}$, $\hat{t}_{qcz}$, $\hat{\mathbf{t}}_{qy\mathbf{x}}$, $\widehat{N}$, $\hat{t}_{qc}$ are Horvitz–Thompson (or also called π) estimators (see e. g. [10, p. 43]), and therefore are unbiased estimators of the totals

$$t_{yz} = \sum_{k=1}^{N} y_k z_k, \quad t_c = \sum_{k=1}^{N} c_k, \quad t_{qcc} = \sum_{k=1}^{N} q_k c_k^2, \quad t_{qcyz} = \sum_{k=1}^{N} q_k c_k y_k z_k,$$
$$t_y = \sum_{k=1}^{N} y_k, \quad t_{qcy} = \sum_{k=1}^{N} q_k c_k y_k, \quad t_z = \sum_{k=1}^{N} z_k, \quad \mathbf{t}_{\mathbf{x}} = \sum_{k=1}^{N} \mathbf{x}_k,$$
$$\mathbf{A}_{q\mathbf{x}} = \sum_{k=1}^{N} q_k \mathbf{x}_k \mathbf{x}_k^{'}, \quad \mathbf{t}_{qz\mathbf{x}} = \sum_{k=1}^{N} q_k z_k \mathbf{x}_k, \quad t_{qcz} = \sum_{k=1}^{N} q_k c_k z_k,$$
$$\mathbf{t}_{qy\mathbf{x}} = \sum_{k=1}^{N} q_k \mathbf{x}_k y_k, \quad N = \sum_{k=1}^{N} 1, \quad t_{qc} = \sum_{k=1}^{N} q_k c_k,$$

respectively.

It follows from expression (22) that $\widehat{\mathrm{Cov}}_{mw}^{(4)}(y,z)$ is a function of the unbiased estimators mentioned above. Using the Taylor linearization method, we approximate the function $\widehat{\mathrm{Cov}}_{mw}^{(4)}(y,z)$ by a linear one. The linear part of the Taylor series expansion of $\widehat{\mathrm{Cov}}_{mw}^{(4)}(y,z)$ at the mean point

$$\begin{aligned}&(\hat{t}_{yz}, \hat{t}_c, \hat{t}_{qcc}, \hat{t}_{qcyz}, \hat{t}_y, \hat{t}_{qcy}, \hat{t}_z, \hat{\mathbf{t}}_{\mathbf{x}}, \widehat{\mathbf{A}}_{q\mathbf{x}}, \hat{\mathbf{t}}_{qz\mathbf{x}}, \hat{t}_{qcz}, \hat{\mathbf{t}}_{qy\mathbf{x}}, \widehat{N}, \hat{t}_{qc})\\ &\quad = (t_{yz}, t_c, t_{qcc}, t_{qcyz}, t_y, t_{qcy}, t_z, \mathbf{t}_{\mathbf{x}}, \mathbf{A}_{q\mathbf{x}}, \mathbf{t}_{qz\mathbf{x}}, t_{qcz}, \mathbf{t}_{qy\mathbf{x}}, N, t_{qc})\end{aligned}$$

is

$$\begin{aligned}&\widehat{\mathrm{Cov}}_{mwL}^{(4)}(y,z)\\ &\quad = \frac{1}{N-1}\left(-Bt_c + \hat{t}_{yz} + B\hat{t}_c - \frac{1}{N} t_z \hat{t}_y - \frac{1}{N} t_y \hat{t}_z + \frac{1}{N^2} t_y t_z \widehat{N}\right)\\ &\quad = \frac{1}{N-1}\Bigg(-Bt_c + \sum_{k\in s} d_k y_k z_k + B\sum_{k\in s} d_k c_k - \frac{1}{N} t_z \sum_{k\in s} d_k y_k\\ &\qquad\qquad - \frac{1}{N} t_y \sum_{k\in s} d_k z_k + \frac{1}{N^2} t_y t_z \sum_{k\in s} d_k\Bigg)\\ &\quad = \frac{1}{N-1}\left(-Bt_c + \sum_{k\in s} d_k e_k\right).\end{aligned}$$

The approximate variance of the estimator $\widehat{\mathrm{Cov}}_{mw}^{(4)}(y,z)$ is equal to

$$\mathrm{AVar}\big(\widehat{\mathrm{Cov}}_{mw}^{(4)}(y,z)\big) = \mathrm{Var}\big(\widehat{\mathrm{Cov}}_{mwL}^{(4)}(y,z)\big)$$
$$= \frac{1}{(N-1)^2}\mathrm{Var}\bigg(-Bt_c + \sum_{k\in s} d_k e_k\bigg) = \frac{1}{(N-1)^2}\mathrm{Var}\bigg(\sum_{k\in s} d_k e_k\bigg).$$

The final expression of the approximate variance of the calibrated estimator $\widehat{\mathrm{Cov}}_{mw}^{(4)}(y,z)$ is obtained using the expression of the variance (see, for example [10]) for the Horvitz–Thompson estimator of the total of the variable $yz + Bc - \mu_z y - \mu_y z + \mu_y \mu_z$. □

Expressions (14), (15) of the weights $w_k^{(1)}$, $w_k^{(2)}$, $w_k^{(3)}$ for the estimator

$$\widehat{\mathrm{Cov}}_{mw}^{(2)}(y,z) = \frac{1}{N-1}\sum_{k\in s} w_k^{(1)}\bigg(y_k - \frac{1}{N}\sum_{l\in s} w_l^{(2)} y_l\bigg)\bigg(z_k - \frac{1}{N}\sum_{l\in s} w_l^{(3)} z_l\bigg)$$

are also explicit. Thus, the Taylor linearization method may be employed to derive an approximate variance for this estimator. The solution is presented by the following proposition.

Proposition 3. *The Taylor linearization approach gives the same approximate variance for the estimators $\widehat{\mathrm{Cov}}_{mw}^{(2)}(y,z)$ and $\widehat{\mathrm{Cov}}_{mw}^{(4)}(y,z)$.*

The proof is similar to that of Proposition 2.

Remark 1. We propose the estimator

$$\widehat{\mathrm{Var}}(\widehat{\mathrm{Cov}}_{mw}(y,z)) = \frac{1}{(N-1)^2}\sum_{k\in s}\sum_{l\in s}\bigg(1 - \frac{\pi_k\pi_l}{\pi_{kl}}\bigg)\frac{\hat{e}_k}{\pi_k}\frac{\hat{e}_l}{\pi_l},$$

for estimating the variances of the estimators $\widehat{\mathrm{Cov}}_{mw}^{(2)}(y,z)$ and $\widehat{\mathrm{Cov}}_{mw}^{(4)}(y,z)$, because the approximate variances of these estimators are equal.

The values $\hat{e}_k$ are defined by replacing unknown parameters t_{qcyz}, t_{qcc}, t_{qcy}, t_{qcz}, t_{qc}, μ_y and μ_z in the expression of e_k, given in Proposition 2, with their estimates: $\hat{t}_{qcyz}$, $\hat{t}_{qcc}$, $\hat{t}_{qcy}$, $\hat{t}_{qcz}$, $\hat{t}_{qc}$, $\hat{\mu}_y = N^{-1}\hat{t}_y$ and $\hat{\mu}_z = N^{-1}\hat{t}_z$.

Remark 2. Replication methods, such as the jackknife, bootstrap and balanced half-samples, may be used for the estimation of variances of the estimators $\widehat{\mathrm{Cov}}_{mw}^{(i)}(y,z)$, $i = 1,3,5,6$. All these methods are described, for example, in [10]. Some bootstrap methods for survey sampling are considered in [12].

4 Simulation study

4.1 Influence of different weighting systems on the accuracy of estimation

The simulation study is performed to observe the efficiency of calibrated estimators of the covariance. The calibrated estimators that use one weighting system and are derived using the same calibration equation, are very similar despite the loss function used. This is the reason, why only three estimators $\widehat{\mathrm{Cov}}_{1w}^{(non)}(y,z)$, $\widehat{\mathrm{Cov}}_{1w}^{(tot)}(y,z)$, $\widehat{\mathrm{Cov}}_{1w}^{(lin)}(y,z)$ that use one system of weights are included into the simulation.

The subset of a real population of size 300 from the Lithuanian Enterprise Survey is used for the simulation. Two variables (a and b) are the numbers of employees for a different time period, and the other two variables (y and z) are the profit of the enterprise at the same periods. The population is stratified into two strata by the size of the survey variable y. The stratified simple random sample is used as a sample design. The sample size $n = 100$ is allocated to strata, using Neyman's optimal allocation. $M = 1000$ samples were drawn and for each of them the calibrated estimators $\widehat{\mathrm{Cov}}_{1w}^{(non)}(y,z)$, $\widehat{\mathrm{Cov}}_{1w}^{(tot)}(y,z)$, $\widehat{\mathrm{Cov}}_{1w}^{(lin)}(y,z)$ that use one weighting system, the estimators $\widehat{\mathrm{Cov}}_{mw}^{(i)}(y,z)$, $i = 1,2,3,4,5,6$, that use two or three weighting systems, and the design based estimator $\widehat{\mathrm{Cov}}(y,z)$ were computed. As it has been shown, the calibrated estimators contain free additional constants. In the sequel it is assumed $q_k = 1$ for all $k \in \mathcal{U}$. The empirical relative bias (RB), variance (Var), relative root mean square error ($RRMSE$), and the coefficient of variation (cv) for each estimator and for some different sets of auxiliaries, having different correlation ρ with the study variables, are presented in Table 1. For any estimator $\widehat{\theta}$ of the finite population parameter θ, all these characteristics of accuracy are defined by the following equations:

$$RB(\widehat{\theta}) = \frac{1}{M}\sum_{i=1}^{M}\frac{\widehat{\theta}_i - \theta}{\theta}, \quad \mathrm{Var}(\widehat{\theta}) = \frac{1}{M}\sum_{i=1}^{M}\left(\widehat{\theta}_i - \frac{1}{M}\sum_{j=1}^{M}\widehat{\theta}_i\right)^2,$$

$$RRMSE(\widehat{\theta}) = \frac{1}{\theta}\sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(\widehat{\theta}_i - \theta\right)^2}, \quad cv(\widehat{\theta}) = \frac{\sqrt{\mathrm{Var}(\widehat{\theta})}}{\frac{1}{M}\sum_{i=1}^{M}\widehat{\theta}_i},$$

where $\widehat{\theta}_i$ is the estimate of θ computed from the ith simulated sample.

Table 1. The main estimated characteristics of accuracy for the estimators of the finite population covariance (sample size: $n = 100$).

Estimator	RB	$\mathrm{Var} \times 10^{-13}$	$RRMSE$	cv
$\rho(y,a) = 0.81$ $\rho(z,b) = 0.90$ $\rho(y,b) = 0.63$ $\rho(z,a) = 0.60$				
$\widehat{\mathrm{Cov}}_{1w}^{(non)}(y,z)$	-0.0495	2.7493	0.0935	0.0835
$\widehat{\mathrm{Cov}}_{1w}^{(tot)}(y,z)$	-0.0796	5.3133	0.1360	0.1198
$\widehat{\mathrm{Cov}}_{1w}^{(lin)}(y,z)$	-0.0065	2.2129	0.0715	0.0716
$\widehat{\mathrm{Cov}}_{mw}^{(1)}(y,z)$	-0.0019	2.1657	0.0704	0.0705
$\widehat{\mathrm{Cov}}_{mw}^{(2)}(y,z)$	-0.0049	2.1194	0.0698	0.0700
$\widehat{\mathrm{Cov}}_{mw}^{(3)}(y,z)$	-0.0510	2.8040	0.0950	0.0844
$\widehat{\mathrm{Cov}}_{mw}^{(4)}(y,z)$	-0.0046	2.1211	0.0698	0.0700
$\widehat{\mathrm{Cov}}_{mw}^{(5)}(y,z)$	-0.0505	2.7920	0.0946	0.0842

$\widehat{\mathrm{Cov}}^{(6)}_{mw}(y,z)$	-0.0050	2.1078	0.0696	0.0698
$\widehat{\mathrm{Cov}}(y,z)$	-0.0735	10.3861	0.1708	0.1665
$\rho(y,a)=0.21$	$\rho(z,b)=0.90$	$\rho(y,b)=0.63$	$\rho(z,a)=0.15$	
$\widehat{\mathrm{Cov}}^{(non)}_{1w}(y,z)$	-0.0635	6.7417	0.1395	0.1327
$\widehat{\mathrm{Cov}}^{(tot)}_{1w}(y,z)$	-0.0743	5.2115	0.1321	0.1180
$\widehat{\mathrm{Cov}}^{(lin)}_{1w}(y,z)$	-0.0858	9.4940	0.1706	0.1613
$\widehat{\mathrm{Cov}}^{(1)}_{mw}(y,z)$	-0.0792	9.8254	0.1696	0.1629
$\widehat{\mathrm{Cov}}^{(2)}_{mw}(y,z)$	-0.0814	9.3788	0.1676	0.1595
$\widehat{\mathrm{Cov}}^{(3)}_{mw}(y,z)$	-0.0643	6.7424	0.1399	0.1328
$\widehat{\mathrm{Cov}}^{(4)}_{mw}(y,z)$	-0.0784	9.2041	0.1650	0.1575
$\widehat{\mathrm{Cov}}^{(5)}_{mw}(y,z)$	-0.0619	6.6470	0.1380	0.1315
$\widehat{\mathrm{Cov}}^{(6)}_{mw}(y,z)$	-0.0805	9.4446	0.1677	0.1599
$\widehat{\mathrm{Cov}}(y,z)$	-0.0738	9.7766	0.1668	0.1615
$\rho(y,a)=0.23$	$\rho(z,b)=0.31$	$\rho(y,b)=0.19$	$\rho(z,a)=0.16$	
$\widehat{\mathrm{Cov}}^{(non)}_{1w}(y,z)$	-0.0627	12.1333	0.1781	0.1778
$\widehat{\mathrm{Cov}}^{(tot)}_{1w}(y,z)$	-0.0703	10.2911	0.1688	0.1651
$\widehat{\mathrm{Cov}}^{(lin)}_{1w}(y,z)$	-0.0767	10.2916	0.1716	0.1663
$\widehat{\mathrm{Cov}}^{(1)}_{mw}(y,z)$	-0.0764	10.2927	0.1715	0.1662
$\widehat{\mathrm{Cov}}^{(2)}_{mw}(y,z)$	-0.0763	10.2829	0.1714	0.1661
$\widehat{\mathrm{Cov}}^{(3)}_{mw}(y,z)$	-0.0666	11.4251	0.1749	0.1733
$\widehat{\mathrm{Cov}}^{(4)}_{mw}(y,z)$	-0.0757	10.3007	0.1712	0.1662
$\widehat{\mathrm{Cov}}^{(5)}_{mw}(y,z)$	-0.0660	11.4427	0.1748	0.1733
$\widehat{\mathrm{Cov}}^{(6)}_{mw}(y,z)$	-0.0722	10.3695	0.1702	0.1661
$\widehat{\mathrm{Cov}}(y,z)$	-0.0730	10.2602	0.1698	0.1654

In the case of a highly correlated auxiliary variables (if $\rho(y,a)=0.81$ and $\rho(z,b)=0.90$), the combination of linear and nonlinear calibration gives the best results, i.e., the most accurate estimator is $\widehat{\mathrm{Cov}}^{(6)}_{mw}$. The first system of weights $w_k^{(1)}$ of this estimator is defined by the linear equation (9), while the second system $w_k^{(2)}$ satisfies the nonlinear equation (3).

If the first system of weights is defined by the nonlinear equation and the two additional systems satisfy the traditional equations (5), we get the estimators $\widehat{\mathrm{Cov}}^{(3)}_{mw}$ and $\widehat{\mathrm{Cov}}^{(5)}_{mw}$, a relative root mean square error of which is larger than that of some calibrated estimators which use one weighting system. The reason is that the estimators $\widehat{\mathrm{Cov}}^{(3)}_{mw}$ and $\widehat{\mathrm{Cov}}^{(5)}_{mw}$ have higher relative bias. The accuracy of estimators $\widehat{\mathrm{Cov}}^{(1)}_{mw}$, $\widehat{\mathrm{Cov}}^{(2)}_{mw}$, $\widehat{\mathrm{Cov}}^{(4)}_{mw}$ is similar to that of $\widehat{\mathrm{Cov}}^{(6)}_{mw}$. The estimator $\widehat{\mathrm{Cov}}^{(1)}_{mw}$ has the lowest relative bias.

In the case of one well correlated auxiliary variable (if $\rho(y,a)=0.21$ and $\rho(z,b)=0.90$) the estimators $\widehat{\mathrm{Cov}}^{(3)}_{mw}$ and $\widehat{\mathrm{Cov}}^{(5)}_{mw}$ are most accurate among those that use several systems of weights. The accuracy characteristics of these estimators are close to that of the estimator $\widehat{\mathrm{Cov}}^{(non)}_{1w}$. This may be explained by the fact that the same nonlinear equation is used for the definition of the first weighting system $w_k^{(1)}$ of the estimators $\widehat{\text{Cov}}_{mw}^{(3)}$ and $\widehat{\text{Cov}}_{mw}^{(5)}$. The accuracy of other estimators that use several systems of weights is similar to that of the estimator $\widehat{\text{Cov}}_{1w}^{(lin)}$ which uses one weighting system defined by a linear calibration equation. In the case of estimators $\widehat{\text{Cov}}_{mw}^{(2)}$, $\widehat{\text{Cov}}_{mw}^{(4)}$ and $\widehat{\text{Cov}}_{mw}^{(6)}$ this may be explained by the linear calibration equation (9) that is used to define the first weighting system $w_k^{(1)}$.

In the case of low correlated auxiliary variables, all the calibrated estimators and the standard estimator (1) are of a similar quality. The standard estimator has a simple analytical form and all its characteristics of accuracy are close to that of the calibrated estimators. We can suggest to use it for estimating the finite population covariance, when no correlated auxiliary variables are available.

4.2 The performance of the variance estimator proposed

The empirical study of the quality of the variance estimator proposed in Remark 1 is presented in Table 2. The same data and the same sample design is used for simulation. Note that this variance estimator is applicable only to the estimators $\widehat{\text{Cov}}_{mw}^{(2)}(y,z)$ and $\widehat{\text{Cov}}_{mw}^{(4)}(y,z)$. The mean value of the variance estimators of 1000 samples is given in the fourth column of Table 2. It seems that the proposed variance estimators slightly underestimate the empirical variance (EmpVar). The approximate variance (AVar) is given in the second column.

Table 2. The main estimated characteristics of accuracy of the variance estimators of the finite population covariance (true value of covariance: $\text{Cov}(y,z) = 66083066$, sample size: $n = 100$).

Estimator	$\text{AVar} \times 10^{-13}$	$\text{EmpVar} \times 10^{-13}$	$\widehat{\text{Var}} \times 10^{-13}$
$\rho(y,a) = 0.81$	$\rho(z,b) = 0.90$	$\rho(y,b) = 0.63$	$\rho(z,a) = 0.60$
$\widehat{\text{Cov}}_{mw}^{(2)}(y,z)$	1.9112	2.1194	1.9394
$\widehat{\text{Cov}}_{mw}^{(4)}(y,z)$	1.9112	2.1211	1.9394
$\rho(y,a) = 0.21$	$\rho(z,b) = 0.90$	$\rho(y,b) = 0.63$	$\rho(z,a) = 0.15$
$\widehat{\text{Cov}}_{mw}^{(2)}(y,z)$	10.1780	9.3788	7.4862
$\widehat{\text{Cov}}_{mw}^{(4)}(y,z)$	10.1780	9.2041	7.4862
$\rho(y,a) = 0.23$	$\rho(z,b) = 0.31$	$\rho(y,b) = 0.19$	$\rho(z,a) = 0.16$
$\widehat{\text{Cov}}_{mw}^{(2)}(y,z)$	10.4471	10.2829	7.5708
$\widehat{\text{Cov}}_{mw}^{(4)}(y,z)$	10.4471	10.3007	7.5708

5 New regression estimators of the population total

An important question is "how the calibrated estimators of covariance may be applied in survey sampling?" In this section, we present how they can be applied to estimate the finite population total.

Consider a finite population $\mathcal{U} = \{u_1, u_2, \ldots, u_N\}$ of N elements. Assume, in this section, y to be a study variable and variables z, a, b to be known auxiliary variables. Let a population total

$$t_y = \sum_{k=1}^{N} y_k$$

be a parameter of interest. In the presence of a multivariate auxiliary variable, the generalized regression estimator (GREG) (see, for example, [10, p. 219–244],) is mainly used for the estimation of the finite population total. In our case, we denote the auxiliary vector, attributed to the element k, by $\mathbf{x}_k$, $k = 1, \ldots, N$, and put $\mathbf{x}_k = (1, z_k, a_k, b_k)'$. The GREG estimator is expressed as follows:

$$\hat{t}_{yGREG} = \sum_{k \in s} d_k y_k + \left(\sum_{k=1}^{N} \mathbf{x}_k - \sum_{k \in s} d_k \mathbf{x}_k \right)' \widehat{B},$$

where

$$\widehat{B} = \left(\sum_{k \in s} d_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} \sum_{k \in s} d_k \mathbf{x}_k y_k.$$

In the case of one auxiliary variable, say z, the regression estimator of the total t_y is

$$\hat{t}_{yr} = \sum_{k \in s} d_k y_k + \left(\sum_{k=1}^{N} z_k - \sum_{k \in s} d_k z_k \right) \frac{\widehat{\mathrm{Cov}}(y, z)}{\widehat{S}_z^2},$$

where $\widehat{\mathrm{Cov}}(y, z)$ is standard estimator (1) of the covariance; $\widehat{S}_z^2 = \widehat{\mathrm{Cov}}(z, z)$ is an estimator of the variance of the variable z of the same type. Note that actually we know the true variance $S_z^2 = \mathrm{Cov}(z, z)$. Despite this fact, the estimator $\widehat{\mathrm{Cov}}(y, z)/\widehat{S}_z^2$ of the regression coefficient $\mathrm{Cov}(y, z)/S_z^2$ is used in statistical theory and practice. In most cases, it is more stable and has lower variance.

Now we shall modify the estimator $\hat{t}_{yr}$, using calibrated estimators of the covariance considered in the paper, and introduce three new estimators of the total t_y:

$$\hat{t}_{yrw}^{(1)} = \sum_{k \in s} d_k y_k + \left(\sum_{k=1}^{N} z_k - \sum_{k \in s} d_k z_k \right) \frac{\widehat{\mathrm{Cov}}_{1w}^{(lin)}(y, z)}{\widehat{\mathrm{Cov}}_{1w}^{(lin)}(z, z)}, \tag{23}$$

$$\hat{t}_{yrw}^{(2)} = \sum_{k \in s} d_k y_k + \left(\sum_{k=1}^{N} z_k - \sum_{k \in s} d_k z_k \right) \frac{\widehat{\mathrm{Cov}}_{mw}^{(6)}(y, z)}{\widehat{\mathrm{Cov}}_{mw}^{(6)}(z, z)}, \tag{24}$$

$$\hat{t}_{yrw}^{(3)} = \sum_{k \in s} d_k y_k + \left(\sum_{k=1}^{N} z_k - \sum_{k \in s} d_k z_k \right) \frac{\widehat{\mathrm{Cov}}_{mw}^{(2)}(y, z)}{\widehat{\mathrm{Cov}}_{mw}^{(2)}(z, z)}. \tag{25}$$

These estimators are obtained using the estimators $\widehat{\text{Cov}}_{1w}^{(lin)}$, $\widehat{\text{Cov}}_{mw}^{(6)}$, $\widehat{\text{Cov}}_{mw}^{(2)}$ of the covariance that employ one, two, and three weighting systems, respectively.

A short simulation study is performed to compare these estimators of total. We employ the same data of Section 4 from the Lithuanian Enterprise Survey.

The variable z is used to define the initial regression estimator, the variables a and b serve as the auxiliaries for the variables y and z, respectively, when estimating the covariance $\text{Cov}(y, z)$ and variance S_z^2 in (23), (24) and (25). The population is stratified into two strata by the size of the survey variable y. The stratified simple random sample is used as a sample design. The sample size $n = 30$ is allocated to strata, using Neyman's optimal allocation. 1000 samples were drawn and the average of the estimates is taken.

In Tables 3 and 4, the relative empirical bias, variance, relative root mean square error and coefficient of variation for the regression estimators are presented. The results of Table 4 are obtained from a modified data set which was produced from the initial data set by replacing the values of the variable y with the values of the variable a. The regression estimators that are obtained using the calibrated estimators of covariance are at least of the same accuracy (Table 3) or more accurate (Table 4) as compared to the GREG. A simple regression estimator that uses one auxiliary variable can also be more effective in comparison with GREG, which uses three auxiliaries.

Consequently, more accurate estimators of the covariance may be useful for estimating the finite population total or mean.

Table 3. The main estimated characteristics of accuracy for the regression estimators of the population total (sample size: $n = 30$).

Estimator	RB	Var $\times 10^{-13}$	$RRMSE$	cv
$\rho(y,z) = 0.70$		$\rho(y,a) = 0.81$	$\rho(z,b) = 0.90$	
$\hat{t}_{yGREG}$	0.0446	0.4031	0.0655	0.0460
$\hat{t}_{yr}$	-0.0126	1.6011	0.0965	0.0969
$\hat{t}_{yrw}^{(1)}$	0.0061	0.6894	0.0631	0.0624
$\hat{t}_{yrw}^{(2)}$	0.0028	0.7490	0.0655	0.0653
$\hat{t}_{yrw}^{(3)}$	0.0027	0.7536	0.0657	0.0655

Table 4. The main estimated characteristics of accuracy for the regression estimators of the population total (sample size: $n = 30$).

Estimator	RB	Var	$RRMSE$	cv
$\rho(a,z) = 0.54$		$\rho(y,a) = 0.81$	$\rho(z,b) = 0.90$	
$\hat{t}_{yGREG}$	-0.0234	70151	0.1096	0.1096
$\hat{t}_{yr}$	-0.0156	54841	0.0959	0.0962
$\hat{t}_{yrw}^{(1)}$	0.0038	37226	0.0781	0.0777
$\hat{t}_{yrw}^{(2)}$	-0.0010	36293	0.0770	0.0771
$\hat{t}_{yrw}^{(3)}$	-0.0002	35796	0.0765	0.0765

References

1. J.C. Deville, C.E. Särndal, Calibration estimators in survey sampling, *J. Am. Stat. Assoc.*, **87**, pp. 376–382, 1992.
2. A. Plikusas, Calibrated weights for the estimators of the ratio, *Lith. Math. J.*, **43**(special iss.), pp. 543–547, 2003.
3. D. Krapavickaitė, A. Plikusas, Estimation of a Ratio in the Finite Population, *Informatica*, **16**, pp. 347–364, 2005.
4. T. Harms, P. Duchesne, On calibration estimation for quantiles, *Surv. Methodol.*, **52**, pp. 37–52, 2006.
5. M. Rueda, S. Martínez, H. Martínez, A. Arcos, Calibration methods for estimating quantiles, *Metrika*, **66**, pp. 355–371, 2007.
6. R.R. Sitter, C. Wu, Efficient Estimation of Quadratic Finite Population Functions in the Presence of Auxiliary Information, *J. Am. Stat. Assoc.*, **97**, pp. 535–543, 2002.
7. S. Singh, S. Horn, S. Chowdhury, F. Yu, Calibration of the estimators of variance, *Aust. N.Z. J. Stat.*, **41**, pp. 199–212, 1999.
8. A. Plikusas, D. Pumputis, Calibrated estimators of the population covariance, *Acta Appl. Math.*, **97**, pp. 177–187, 2007.
9. C.E. Särndal, The calibration approach in survey theory and practice, *Surv. Methodol.*, **33**(2), pp. 99–119, 2007.
10. C.E. Särndal, B. Swensson, J. Wretman, *Model Assisted Survey Sampling*, Springer-Verlag, New York, 1992.
11. M. Rueda, S. Martínez, H. Martínez, A. Arcos, Estimation of the distribution function with calibration methods, *J. Stat. Plan. Infer.*, **137**, pp. 435–448, 2007.
12. R.R. Sitter, Comparing three bootstrap methods for survey data, *Can. J. Stat.*, **20**(2), pp. 135–154, 1992.