

## On the generalisation of Quatember's bootstrap

Tomasz Żądło<sup>1</sup>

### ABSTRACT

The problem of the estimation of the design-variance and the design-MSE of different estimators and predictors is considered. Bootstrap algorithms applicable to complex sampling designs are used. A generalisation of the bootstrap procedure studied by Quatember (2014) is proposed. In most of the cases considered in our simulation study it leads to more accurate estimates (or to very similar ones in remaining cases) of the design-MSE and the design-variance compared with the original algorithm and its other counterparts.

**Key words:** bootstrap for complex sampling designs, variance estimation, MSE estimation.

### 1. Introduction

Let the population of size  $N$  be denoted by  $\Omega$ . The population is divided into  $D$  disjoint subpopulations (domains)  $\Omega_d$ , each of size  $N_d$ , where  $d = 1, 2, \dots, D$ . Let the sample be denoted by  $s$  and its size by  $n$ . The set of sampled elements of  $d$ th domain is denoted by  $s_d$  and its size by  $n_d$ . Let the values of the variable of interest observed in the sample be denoted by  $y_k$  ( $k = 1, 2, \dots, n$ ). We additionally assume that vectors of auxiliary variables  $\mathbf{x}_l$  ( $l = 1, 2, \dots, N$ ) are known for all population elements. First and second order inclusion probabilities are denoted by  $\pi_k$  and  $\pi_{kl}$ , respectively. We consider the problem of estimation of the population (subpopulation) parameter  $\theta$  ( $\theta_d$ ) using estimator  $\hat{\theta}$  ( $\hat{\theta}_d$ ). The key issue is the estimation of the design-variance and the design-MSE of  $\hat{\theta}$  ( $\hat{\theta}_d$ ). In official statistics, the design-based accuracy is of primary interest and hence model-based methods, where the prediction accuracy is assessed, are not widely used. What is more, the comparison of the accuracy of methods based on different approaches (e.g. design-based and model-based under different

---

<sup>1</sup> University of Economics in Katowice, Katowice, Poland, E-mail: tomasz.zadlo@ue.katowice.pl.  
ORCID: <https://orcid.org/0000-0003-0638-0748>.

superpopulation models) is not appropriate if MSE is estimated under different approaches too. Hence, the aim of the paper is to present:

- a proposal of a generalisation of Quatember (2014) bootstrap valid for complex sampling designs, which can be used to estimate the design-precision and the design-accuracy of any estimator or predictor,
- a simulation study of properties of our proposals and other bootstrap estimators of the design-variance and the design-MSE not only in the case of estimation of population parameters but also in the case of estimation and prediction of subpopulations characteristics.

## 2. Bootstrap methods for complex sampling designs

The classic Efron's bootstrap (Efron, 1979) procedure, where simple random samples are drawn with replacement from the original sample, is correct under independence of random variables. In the case of complex sampling designs appropriate modifications must be used.

According to Ranalli and Mecatti (2012), majority of bootstrap methods for complex sampling designs can be classified into one out of two approaches. The first one is called an ad-hoc approach and is usually based on iid resampling and rescaling sample data. They classify, inter alia, the rescaling bootstrap (Rao and Wu, 1988), the mirror-match bootstrap (Sitter, 1992) and the generalised weighted bootstrap (Beaumont and Patak, 2012) as methods belonging to this approach. Proposals presented by Antal and Tillé (2011, 2014) are also taken into account in this approach. The Authors use mixtures of several sampling designs for resampling to meet two conditions – firstly, the expectation over the bootstrap distribution of the Horvitz-Thompson (1952) (HT) estimator must be equal to the value of the HT estimator computed based on the original sample; secondly, the variance over the bootstrap distribution of HT estimator must be equal (or approximately equal) to the HT variance estimator (Horvitz and Thompson, 1952) or Sen-Yates-Grundy variance estimator (Sen 1953, Yates and Grundy 1953). The second approach is the plug-in approach. It is based on the concept of pseudopopulation, although in some methods the pseudopopulation is not physically generated. The basic idea is as follows:

- 1) We built a pseudopopulation  $\Omega^* = \{1, 2, \dots, k^*, \dots, N^{pseudo}\}$ , where pseudoelements are replications of elements observed in the original sample. The element  $k$  observed in the original sample is replicated  $w_k$ -times.
- 2) A bootstrap sample  $s^*$  of size  $n$  (original sample size) is drawn from  $\Omega^*$  mimicking the original sampling design.

- 3) The value of estimator  $\hat{\theta}$  is computed based on  $s^*$  and it is denoted by  $\hat{\theta}^*$ .
- 4) Steps b) and c) are iterated  $B$  times providing  $\hat{\theta}_b^*$ , where  $b = 1, 2, \dots, B$ .

Bootstrap estimators of the design-variance and the design-bias are defined as follows (e.g. Rao and Wu 1988):

$$\hat{D}_{boot}^2(\hat{\theta}) = \frac{1}{B-1} \sum_{b=1}^B \left( \hat{\theta}_b^* - \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* \right)^2, \tag{1}$$

$$\hat{B}_{boot}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* - \hat{\theta}, \tag{2}$$

where  $\hat{\theta}$  is the value of the considered estimator based on the original sample.

The algorithm presented above allows for different definitions of weights  $w_k$ . One of the first proposals was presented by Holmberg (1998), who defined it as follows:  $w_k = \lfloor \pi_k^{-1} \rfloor + \epsilon_k$ , where  $\lfloor \pi_k^{-1} \rfloor$  is rounded down value of  $\pi_k^{-1}$ ,  $\epsilon_k$  is generated from Bernoulli distribution with probability  $\pi_k^{-1} - \lfloor \pi_k^{-1} \rfloor$ .

Other solutions include Barbiero and Mecatti (2010) 0.5 bootstrap, where inverses of inclusion probabilities are rounded to the nearest integer. Barbiero and Mecatti (2010) consider two x-balanced methods, where inverses of first order inclusion probabilities are rounded down and additional pseudoelements are included in the pseudopopulation to reach the minimum absolute difference between total values of an auxiliary variable in the real population and the pseudopopulation. Barbiero, Manzi and Mecatti (2015) define  $w_k$  as calibration weights rounded to the nearest integer.

There are two possible limitations of the above algorithms. Firstly, we require generation of the pseudopopulation of size (approximately) equal to the original population size, which may be problematic in the case of large real populations. Secondly, the number of replications  $w_k$  must be integer. The first problem is solved by Ranalli and Mecatti (2012) by directly re-sampling from the sample using appropriate sampling designs where  $n$  out of  $n$  elements are drawn at random, mimicking the original sample design, where  $N$  out of  $n$  elements are selected. The Quatember (2014) bootstrap omits both of the limitations but it is proposed only for simple random sampling without replacement and for probability proportional to size sampling.

Let us present the idea of the Quatember (2014) bootstrap. Although the pseudopopulation is not created, the process of sampling from the pseudopopulation is mimicked in the procedure of selecting a bootstrap sample of size  $n$  out of  $n$  elements observed in the original sample with appropriate probabilities by modification of the original sampling scheme. Firstly, let us present the algorithm of drawing  $b$ th ( $b = 1, 2, \dots, B$ ) bootstrap sample of size  $n$  for simple random sampling without replacement. Quatember (2014) assumes that the number of replications of sample

element  $k$  in the pseudopopulation, which is not physically created, equals its (possibly non-integer) design-weight - the inverse of the first order probability:  $Nn^{-1}$ . After draw  $j-1$  the number of remaining replications of element  $k$  in the pseudopopulation equals:  $Nn^{-1} - h_{k,j-1}$ , where  $h_{k,j-1}$  is the number of replications of element  $k$  selected in the bootstrap procedure in the first  $j-1$  draws. What is more, the probability of selecting a population element from the pseudopopulation of size  $N$  in the  $j$ th draw equals  $(N-j+1)^{-1}$ . Finally, element  $k$  is drawn from the original sample in the  $j$ th draw ( $j=1,2,\dots,n$ ) of the bootstrap algorithm with probability:

$$(Nn^{-1} - h_{k,j-1}) \times (N-j+1)^{-1}. \quad (3)$$

Secondly, we present the algorithm of drawing  $b$ th ( $b=1,2,\dots,B$ ) bootstrap sample of size  $n$  for probability proportional to size sampling. Quatember (2014) assumes that the number of replications of sample element  $k$  in the pseudopopulation, which is not physically created, equals its (possibly non-integer) design-weight given by:  $t_x(x_k n)^{-1}$ , where  $t_x = \sum_{i \in \Omega} x_i$ . After draw  $j-1$  the number of remaining replications of element  $k$  in the pseudopopulation equals:  $t_x(x_k n)^{-1} - h_{k,j-1}$ , where  $h_{k,j-1}$  is the number of replications of element  $k$  selected in the bootstrap procedure in the first  $j-1$  draws. What is more, Quatember (2014) assumes the following probability of selecting an population element from the pseudopopulation of size  $N$  in the  $j$ th draw in his algorithm:  $x_k \left( t_x - \sum_{i \in s_{bj-1}} x_i \right)^{-1}$ , where  $s_{bj-1}$  is the subset of  $b$ th bootstrap sample after draw  $j-1$ . The drawback of the Quatember (2014) bootstrap is that the assumed probability does not lead to the first order inclusion probabilities proportional to the values of the auxiliary variable (as they should be for probability proportional to size sampling). Finally, element  $k$  is drawn from the original sample in the  $j$ th draw ( $j=1,2,\dots,n$ ) of the bootstrap algorithm with probability:

$$\left( t_x(x_k n)^{-1} - h_{k,j-1} \right) \times x_k \left( t_x - \sum_{i \in s_{bj-1}} x_i \right)^{-1}. \quad (4)$$

### 3. The proposed bootstrap method

The idea of the proposed bootstrap results from motivating simulations studies where we usually observed properties of the design-variance estimators based on the original Quatember (2014) bootstrap better than that of competitors, but problems with

estimation of the design-MSE of some estimators and predictors using auxiliary information. To improve the method we propose to change the number of replications of sampled elements assumed by Quatember (2014) to be equal inverses of first order inclusion probabilities. Although these weights seem to be a natural choice, the choice is not the only and the best one – similarly to the choice between the Horvitz-Thompson estimator (using these weights to estimate the population total) and other estimators or predictors using different weighting systems, which usually lead to more accurate estimates than the Horvitz-Thompson estimator. Hence, below we propose to replace inverses of first order inclusion probabilities in the algorithm presented by Quatember (2014) by some calibration weights summing up to the population size, but other weighting systems are also possible.

To clarify considerations presented below, let us introduce the idea of the calibration estimator of the population total. It is given by (Deville, Särndal 1992):

$$\hat{\theta}^{CAL} = \sum_{k \in s} w_k y_k, \tag{5}$$

where weights  $w_k$  are solutions of:

$$\left\{ \begin{array}{l} f_s(w_k, \pi_k^{-1}, q_k) \rightarrow \min \\ \sum_{k \in s} w_k \mathbf{x}_k = \sum_{l \in \Omega} \mathbf{x}_l \end{array} \right., \tag{6}$$

where  $f_s(w_k, \pi_k^{-1}, q_k)$  is some distance measure between weights of the calibration estimator  $w_k$  and the inverses of the first order inclusion probabilities  $\pi_k^{-1}$ , where for more generality additional known weights  $q_k$  can be included. The minimization in (6) leads to the approximate design-unbiasedness of the calibration estimator. The equality in (6) is the condition of model-unbiasedness of the estimator (5) under the linear model. If in (6) we additionally assume that:

$$f_s(w_k, \pi_k^{-1}, q_k) = \sum_{k \in s} \frac{(w_k - \pi_k^{-1})^2}{\pi_k^{-1} q_k}, \tag{7}$$

then the resulting calibration estimator is called a generalised regression estimator (GREG) (Deville, Särndal 1992; Särndal, Swensson, Wretman 1992, p. 232; Rao, Molina 2015, p. 13). Deville and Särndal (1992) prove under some conditions that calibration estimators and the generalised regression estimator of the population total are asymptotically equivalent. But their values are very similar even for small sample sizes, as shown by Singh and Mohl (1996) and Stukel, Hidirolou and Särndal (1996).

Our proposal of the bootstrap algorithm for simple random sampling without replacement is as follows. In the  $b$  th bootstrap sample ( $b = 1, 2, \dots, B$ ) element  $k$  is

drawn from the original sample in the  $j$  th draw ( $j=1,2,\dots,n$ ) with probability (compare with (3)):

$$(w_k - h_{k,j-1}) \times (N - j + 1)^{-1}, \quad (8)$$

where  $w_k$ 's are some calibration weights such that  $\sum_{i \in \Omega} w_k = N$  (e.g. calibration weights considered by Deville and Särndal (1992)).

Our proposal of the bootstrap algorithm for probability proportional to size sampling is as follows. In the  $b$  th bootstrap sample ( $b=1,2,\dots,B$ ) element  $k$  is drawn from the original sample in the  $j$  th draw ( $j=1,2,\dots,n$ ) with probability (compare with (3)):

$$(w_k - h_{k,j-1}) \times x_k \left( t_x - \sum_{i \in s_{bj-1}} x_i \right)^{-1}, \quad (9)$$

where  $w_k$ 's are some calibration weights such that  $\sum_{i \in \Omega} w_k = N$  (e.g. calibration weights considered by Deville and Särndal (1992)).

Of course, the choice of  $w_k$ 's in the proposed algorithms is ambiguous (similarly to the choice of weights used in estimation). In the simulation studies, presented in the next section, we will consider four arbitrary chosen cases - calibration weights which fulfil four systems of calibration equations presented below. Firstly, we will consider weights  $w_{1k}$  ( $k=1,2,\dots,n$ ) such that (Deville and Särndal (1992)):

$$\sum_{k \in s} w_{1k} \mathbf{x}_k = \sum_{l \in \Omega} \mathbf{x}_l \wedge \sum_{k \in s} w_{1k} = N \wedge L_k \leq w_{1k} \leq U_k, \quad (10)$$

where in simulation studies, to avoid negative and extremely large calibration weights, we will assume that  $\forall_k L_k = 0$  and  $\forall_k U_k = 10\pi_k^{-1}$ . Secondly, we will consider weights  $w_{2k}$  ( $k=1,2,\dots,n$ ) defined similarly to (10) but for domains:

$$\forall_d \sum_{k \in s_d} w_{2k} \mathbf{x}_k = \sum_{l \in \Omega_d} \mathbf{x}_l \wedge \forall_d \sum_{k \in s_d} w_{2k} = N_d \wedge L_k \leq w_{2k} \leq U_k, \quad (11)$$

where  $L_k$  and  $U_k$  are defined as in (10). Thirdly, we will consider weights  $w_{3k}$  ( $k=1,2,\dots,n$ ), which leads to quantile calibration (similarly to Barbiero, Manzi and Mecatti 2015):

$$\sum_{k \in s} w_{3k} I(\mathbf{x}_k \leq \mathbf{x}_p) = Np \wedge \sum_{k \in s} w_{3k} = N \wedge L_k \leq w_{3k} \leq U_k, \quad (12)$$

where  $\mathbf{x}_p$  denotes the vector of population quantiles of auxiliary variables of order  $p = \{0.25, 0.5, 0.75\}$ ,  $L_k$  and  $U_k$  are defined as in (10). Fourthly, we will consider weights  $w_{4k}$  ( $k = 1, 2, \dots, n$ ) defined similarly to (12) but for domains:

$$\forall_d \sum_{k \in s_d} w_{4k} I(\mathbf{x}_k \leq \mathbf{x}_{dp}) = N_d p \wedge \forall_d \sum_{k \in s_d} w_{4k} = N_d \wedge L_k \leq w_{4k} \leq U_k, \quad (13)$$

where  $\mathbf{x}_{dp}$  denotes the vector of domain quantiles of auxiliary variables of order  $p = \{0.25, 0.5, 0.75\}$ ,  $L_k$  and  $U_k$  are defined as in (10).

In cases (10) and (12) calibration equations are solved based on the whole sample, which may be a good solution in the case of estimation of population parameters. We hope that taking into account information on auxiliary variables in building pseudopopulation will give better properties of the design-variance and the design-MSE bootstrap estimators than in case of the algorithm proposed by Quatember (2014). What is more, in cases (11) and (13) calibration equations are solved based on samples in domains, taking into account domain-specific information on auxiliary variables, which should additionally lead to better results in the case of estimation of domain parameters.

#### 4. Simulation study

We present results of a design-based simulation study conducted in R (R Development Core Team 2019). We use real data on  $N = 281$  Swedish municipalities (Särndal, Swensson and Wretman 1992). We assume a relatively large sample size  $n \approx 0.15N$  to show clearly differences between properties of different variance and MSE estimators. Revenues from 1985 municipal taxation (in millions of kronor) are the variable of interest, 1975 population (in thousands) – the auxiliary variable. We consider two subpopulations – the first one of size  $N_1 = 104$ , which consists of municipalities belonging to regions 1, 2 and 3; and the second of size  $N_2 = 177$ , which consists of municipalities belonging to regions 4-8. Large domain sizes will allow us to compare properties of estimators of design-variances and design-MSEs of direct and indirect estimators and predictors of domain totals. We consider probability proportional to size sampling using Brewer sampling scheme (Brewer 1975, Brewer and Hanif 1983). It is known to be a fast algorithm that does not cause problems in the case of asymmetry of the auxiliary variable as it can happen in the case of Rao-Sampford sampling scheme. However, in this sampling scheme there is a problem with computation of joint inclusion probabilities – a recursive formula is required and it implies a complete exploration of the splitting tree (Tillé 2006, p. 113).

In the simulation study we consider the problem of estimation of design-variances and design-MSEs of the following estimators and predictors:

- the Horvitz-Thompson (1952) estimator of the population total (which will be denoted by: HT) and of domains totals (HTd1, HTd2),
- the generalised regression estimator (e.g. Deville and Särndal 1992) of the population total (GREG) and of domains totals (GREGd1, GREGd2),
- the modified generalised regression estimator (e.g. Särndal 1981) of domains totals (MGREGd1, MGREGd2),
- the best linear unbiased predictor (e.g. Royall 1976) of domains totals (BLUPd1, BLUPd2) under the following model  $Y_k = \beta_1 x_k + \beta_1 + \xi_k$ , where  $\xi_k \sim iid(0, \sigma^2)$ .

We consider the following estimators of design-variances and design-MSEs of the above listed estimators and predictors:

- based on the Holmberg (1998) bootstrap (which will be denoted by H),
- based on the Antal and Tillé (2011) bootstrap (AT),
- based on the Quatember (2014) bootstrap (Q),
- the proposed generalised Quatember (2014) bootstrap with weights fulfilling calibration equations (10) (GQ1),
- the proposed generalised Quatember (2014) bootstrap with weights fulfilling calibration equations (11) (GQ2),
- the proposed generalised Quatember (2014) bootstrap with weights fulfilling calibration equations (12) (GQ3),
- the proposed generalised Quatember (2014) bootstrap with weights fulfilling calibration equations (13) (GQ4).

In the case of all bootstrap methods the number of bootstrap iterations equals  $B = 1000$ . Additionally, we consider classic design-variance estimators of the Horvitz-Thompson estimator and the GREG estimator (in both cases denoted by cl), where only first order inclusion probabilities are used. It results from the problems with computations of second order inclusion probabilities in Brewer sampling scheme described above. We use the following design-variance estimator of the Horvitz-Thompson estimator of the population total (Antal and Tillé 2011, p. 536):

$$\hat{D}^2(\hat{\theta}^{HT}) = \sum_{k=1}^n c_k \left( y_k \pi_k^{-1} - \sum_{k=1}^n c_k y_k \pi_k^{-1} \left( \sum_{k=1}^n c_k \right)^{-1} \right)^2, \quad (14)$$

where we use  $c_k = n(1 - \pi_k)(n - 1)^{-1}$  proposed by Hájek (1981), which gives efficient and only slightly biased design-variance estimator (Antal and Tillé 2014, p. 1348).



To estimate the design-variance of the GREG estimator we use the following one based on the Deville's method (Deville 1993):

$$\hat{D}^2(\hat{\theta}^{GREG}) = \left(1 - \sum_{k=1}^n a_k^2\right)^{-1} \sum_{k=1}^n (1 - \pi_k) (e_k \pi_k^{-1} - A)^2, \tag{15}$$

where  $a_k = (1 - \pi_k) \left(\sum_{k=1}^n (1 - \pi_k)\right)^{-1}$ ,  $A = \sum_{k=1}^n a_k e_k \pi_k^{-1}$ ,  $e_k = y_k - \mathbf{x}_k^T \mathbf{B}$ ,  $g_k$  - g-weights of GREG (see Deville and Särndal 1992),  $\mathbf{B} = \sum_{k=1}^n (g_k \pi_k^{-1} \mathbf{x}_k \mathbf{x}_k^T)^{-1} \sum_{k=1}^n (g_k \pi_k^{-1} \mathbf{x}_k y_k)$ . In the case of (14) and (15) replacing  $y_k$  with  $a_{dk} y_k$ , where  $a_{dk} = 1$  if  $k \in s_d$  and 0 otherwise, gives estimators of design-variances of estimators of domain totals.

In the simulation study we compute:

- the relative biases of the estimators of the design-variance of different estimators as

$$100\% \cdot V^{-1} \frac{1}{B} \sum_{r=1}^R (\hat{V}_r - V), \tag{16}$$

- the relative biases of the estimators of the design-MSE of different estimators as

$$100\% \cdot MSE^{-1} \frac{1}{B} \sum_{r=1}^R (M\hat{S}E_r - MSE), \tag{17}$$

- the relative RMSEs of the estimators of the design-variance of different estimators as

$$100\% \cdot V^{-1} \sqrt{\frac{1}{B} \sum_{r=1}^R (\hat{V}_r - V)^2}, \tag{18}$$

- the relative biases of the estimators of the design-MSE of different estimators as

$$100\% \cdot MSE^{-1} \sqrt{\frac{1}{B} \sum_{r=1}^R (M\hat{S}E_r - MSE)^2}, \tag{19}$$

where  $\hat{V}_r$  and  $M\hat{S}E_r$  are estimators of the design-variance and the design-MSE, respectively, obtained in the  $r$ th Monte Carlo iteration  $r = 1, 2, \dots, R$ , whereas  $V$  is the simulation design-variance given by  $V = \frac{1}{R} \sum_{r=1}^R \left(\hat{\theta}_d^r - \frac{1}{B} \sum_{r=1}^R \hat{\theta}_d^r\right)^2$ ,  $MSE$  is the simulation design-MSE given by  $MSE = \frac{1}{R} \sum_{r=1}^R (\hat{\theta}_d^r - \theta_d)^2$ ,  $\hat{\theta}_d^r$  is the value of the estimator of the subpopulation total (or its special case – the estimator of the population

total denoted by  $\hat{\theta}^r$ ) computed in the  $r$ th iteration,  $\theta_d$  is the value of the subpopulation total (or its special case – the population total denoted by  $\theta$ ), the number of samples drawn in the Monte Carlo simulation study equals  $R = 1000$ .

Firstly, we would like to present design-based properties of the considered estimators and predictors. The Horvitz-Thompson estimator is design-unbiased and hence we will consider only its design-variance estimators. GREG is asymptotically design-unbiased estimator (Deville and Särndal 1992), MGREG is approximately p-unbiased if the overall sample size increases even if the domain sample size is small (Molina and Rao 2015, p. 22) – for these estimators usually only design-variance is estimated. Although their relative design-biases obtained in the simulation study are small (see Table A1 in Appendix) we also analyze properties of estimators of their design-MSEs. We also consider best linear unbiased predictors for which prediction-MSEs (not design-MSEs) are usually estimated. Although in our simulation study, their design-biases and design-MSEs are not large (see Table A1 in Appendix), including them will allows us to check properties of the proposed design-MSE estimators not only for design-unbiased or approximately design-unbiased statistics.

Secondly, we present main results of the simulation study. RRMSEs of estimators of design-variances and design-MSEs are presented in Tables 1-3 below, their design-biases in Tables A2-A4 in Appendix. If we compare relative design-biases (see Table A2 and Table A3 in Appendix) and RRMSEs (Table 1 and Table 2) of our proposals of design-variance estimators with bootstrap competitors, we see that usually the best results are obtained for one of the proposed methods or the results for our method are very close to the best one (except results for the HT estimator). Among four proposals (GQ1-GQ4) the GQ1 method is the best choice in most of the cases. If we compare RRMSEs (see Table 3) of our proposals of design-MSE estimators with bootstrap competitors, we obtain similar conclusions – results for GQ1 are usually the best or close to the best.

**Table 1.** RRMSEs in % of bootstrap estimators of design-variances – part 1

Method	HT	HTd1	HTd2	GREG	GREGd1	GREGd2
cl	27.6	9.8	7.6	26.7	13.7	12.5
H	27.6	10.7	9.0	38.4	17.7	16.8
AT	28.1	10.9	8.6	46.8	22.4	21.9
Q	29.6	11.9	9.6	32.0	11.7	10.7
GQ1	31.3	12.5	10.2	27.5	10.5	9.3
GQ2	31.8	13.5	9.7	28.2	12.2	10.6
GQ3	32.8	12.6	10.1	28.4	10.8	9.7
GQ4	34.4	14.1	10.1	30.1	13.7	12.9

**Table 2.** RRMSEs in % of bootstrap estimators of design-variances – part 2

Method	MGREGd1	MGREGd2	BLUPd1	BLUPd2
H	40.4	33.9	43.6	36.9
AT	44.1	39.9	45.0	37.4
Q	41.5	30.7	38.7	32.0
GQ1	43.2	29.2	36.0	25.6
GQ2	44.9	29.8	36.9	25.7
GQ3	42.5	30.1	35.7	23.0
GQ4	46.4	31.4	38.6	25.9

**Table 3.** RRMSEs in % of bootstrap estimators of design-MSEs

Method	GREG	GREGd1	GREGd2	MGREGd1	MGREGd2	BLUPd1	BLUPd2
cl	26.4*	13.8*	12.6*	- **	- **	- ***	- ***
H	39.0	17.6	16.9	40.6	34.8	64.7	38.1
AT	47.2	22.8	22.6	44.2	40.5	64.5	35.3
Q	31.9	11.9	11.0	41.5	30.8	67.0	30.9
GQ1	27.4	10.6	9.3	44.7	32.8	67.3	34.6
GQ2	67.2	176.0	177.5	95.2	45.3	65.3	83.5
GQ3	77.0	42.7	42.5	75.2	112.6	65.4	60.6
GQ4	123.4	183.9	179.6	166.2	113.0	66.1	113.8

\* - design-variance estimator (15) is used to estimate design-MSE

\*\* - classic design-MSE estimator not available due to the lack of second order inclusion probabilities

\*\*\* - design-MSE estimator not available (prediction-MSE is usually estimated)

## 5. Conclusions

We present a generalisation of the bootstrap algorithm for complex sampling designs proposed by Quatember (2014), used to estimate the design-variance and the design-MSE. We study its properties in the case of estimation of population total using the HT and GREG estimators and in the case of estimation of subpopulation total using the HT, GREG, MGREG estimators and the BLUP. In the simulation study based on real data we show that our proposal gives more accurate design-MSE and design-variance estimators in most of cases (or of similar accuracy in other cases) for estimators and predictors which use auxiliary information compared with the original algorithm and other bootstrap methods considered in the paper.

## Acknowledgements

This paper was presented at the MSA 2019 conference, which financed its publication. Organization of the international conference “Multivariate Statistical Analysis 2019” (MSA 2019) was supported from resources for popularization of scientific activities of the Minister of Science and Higher Education in the framework of agreement No. 712/P-DUN/202019.

## References

- ANTAL, E., TILLÉ, Y., (2011). A Direct Bootstrap Method for Complex Sampling Designs From a Finite Population, *Journal of the American Statistical Association*, Vol. 106, No. 494, pp. 534–543.
- ANTAL, E., Tillé, Y., (2014). A New Resampling Method for Sampling Designs Without Replacement: The Doubled Half Bootstrap, *Computational Statistic*, Vol. 29, No. 5, pp. 1345–1363.
- BARBIERO, A., MANZI, G., MECATTI, F., (2015). Bootstrapping probability-proportional-to-size samples via calibrated empirical population, *Journal of Statistical Computation and Simulation*, Vol. 85, No. 3, pp. 608–620.
- BARBIERO, A., MECATTI, F., (2010). Bootstrap algorithms for variance estimation in  $\pi$ PS sampling, In *Complex Data Modeling and Computationally Intensive Statistical Methods* edited by P. Mantovan and P. Secchi, pp. 2019–2026. Springer-Verlag, Italia.
- BEAUMONT, J. F., PATAK, Z., (2012). On the Generalized Bootstrap for Sample Surveys with Special Attention to Poisson Sampling, *International Statistical Review*, Vol. 80, No. 1, pp. 127–148.
- BREWER, K. E. W., (1975). A simple procedure for sampling  $\pi$ pswor, *Australian & New Zealand Journal of Statistics*, Vol. 17, No. 3, pp. 166–172.
- BREWER, K. E. W., HANIF M., (1983). *Sampling with unequal probabilities*, Springer, New York.
- DEVILLE, J. C., (1993). *Estimation de la variance pour less enquêtes en deux phases*. Manuscript, INSEE, Paris.
- DEVILLE, J. C., SÄRNDAL, C. E., (1992). Calibration estimators in survey sampling, *Journal of the American Statistical Association*, Vol. 87, pp. 376–382.

- EFRON, B., (1979). Bootstrap methods: another look at the jackknife, *Annals of Statistics*, Vol. 7, pp. 1–26.
- HÁJEK, J., (1981). *Sampling From a Finite Population*, Marcel Dekker, New York.
- HOLMBERG, A., (1998). A bootstrap approach to probability proportional to size sampling, *Proceedings of Section on Survey Research Methods*, American Statistical Association, Washington, pp. 378–383.
- HORVITZ, D.G., THOMPSON, D. J., (1952). A Generalization of Sampling Without Replacement From a Finite Universe, *Journal of the American Statistical Association*, Vol. 47, No. 260, pp. 663–685.
- QUATEMBER, A., (2014). The Finite Population Bootstrap – from the Maximum Likelihood to the Horvitz-Thompson Approach, *Austrian Journal of Statistics*, Vol. 43, pp. 93–102.
- R DEVELOPMENT CORE TEAM, (2019). *A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna.
- RANALLI, M. G., MECATTI, F., (2012). Comparing Recent Approaches for Bootstrapping Sample Survey Data: A First Step Towards a Unified Approach, *Proceedings of Section on Survey Research Methods*, American Statistical Association, Washington, pp. 4088–4099.
- RAO, J. N. K, MOLINA, I., (2015). *Small area estimation. Second edition*, John Wiley and Sons, Hoboken, New Jersey.
- RAO, J. N. K., WU, C. F. J., (1988). Resampling Inference for Complex Survey Data, *Journal of American Statistical Association*, Vol. 83, pp. 231–241.
- ROYALL, R. M., (1976). The Linear Least Squares Prediction Approach to Two-Stage Sampling, *Journal of the American Statistical Association*, Vol. 71, pp. 657–473.
- SÄRNDAL, C. E, (1981). Frameworks for Inference in Survey Sampling with Applications to Small Area Estimation and Adjustment for Nonresponse, *Bulletin of the International Statistical Institute*, Vol. 49, pp. 494–513.
- SÄRNDAL, C. E., SWENSSON, B., WRETMAN, J., (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.
- SEN, A. R., (1953). On the estimate of variance in sampling with varying probabilities, *Journal of the Indian Society of Agricultural Statistics*, Vol. 5, No. 2, pp. 119–127.
- SINGH, A. C, MOHL, C. A., (1996). Understanding calibration estimators in survey sampling, *Survey Methodology*, Vol. 22, pp. 107–115.

- SITTER, R. R., (1992). A Resampling Procedure for Complex Survey Data, *Journal of the American Statistical Association*, Vol. 87, pp. 755–765.
- STUKEL, D. M., HIDIROGLOU, M. A., SÄRNDAL, C. E., (1996). Variance estimation for calibration estimators: A comparison of jackknifing versus Taylor linearization, *Survey Methodology*, Vol. 22, pp. 177–125.
- TILLÉ, Y., (2006). *Sampling algorithms*, Springer-Verlag, New York.
- YATES, F., GRUNDY, P. M., (1953). Selection Without Replacement from Within Strata with Probability Proportional to Size, *Journal of the Royal Statistical Society, Ser. B*, Vol. 15, pp. 235–261.

**APPENDIX**

**Table A1.** Relative design-biases and design-RRMSEs in % of considered estimators and predictors

estimator/predictor	relative bias (in %)	RRMSE (in %)
HT	-0.079	1.836
HTd1	-0.215	14.654
HTd2	0.037	11.498
GREG	-0.172	1.790
GREGd1	-0.799	15.592
GREGd2	0.359	12.232
MGREGd1	-0.208	2.804
MGREGd2	-0.148	2.190
BLUPd1	-3.100	3.809
BLUPd2	0.544	2.935

**Table A2.** Relative biases in % of bootstrap estimators of design-variances – part 1

Method	HT	HTd1	HTd2	GREG	GREGd1	GREGd2
Cl	9.5	0.6	-1.1	1.6	-6.8	-8.0
H	8.1	-1.4	-2.9	9.1	-0.8	-1.5
AT	9.6	0.6	-1.0	16.8	4.1	3.3
Q	13.5	5.6	3.5	6.2	0.6	0.2
GQ1	14.6	5.8	3.6	3.4	-2.5	-3.1
GQ2	14.8	8.8	6.3	3.8	3.3	3.2
GQ3	14.9	5.5	3.3	3.2	-2.9	-3.7
GQ4	15.2	9.4	7.0	3.8	1.7	1.7

**Table A3.** Relative biases in % of bootstrap estimators of design-variances – part 2

Method	MGREGd1	MGREGd2	BLUPd1	BLUPd2
H	3.2	1.8	7.8	9.3
AT	9.2	10.3	10.3	10.8
Q	7.6	-1.1	1.0	1.4
GQ1	8.6	-3.8	-1.7	-3.8
GQ2	11.4	-3.2	-1.7	-5.7
GQ3	7.0	-4.5	-1.8	-3.8
GQ4	10.2	-2.9	-2.5	-7.0

**Table A4.** Relative biases in % of bootstrap estimators of design-MSEs

Method	GREG	GREGd1	GREGd2	MGREGd1	MGREGd2	BLUPd1	BLUPd2
cl	0,6*	-7.1*	-8.1*	- **	- **	- ***	- ***
H	10.1	2.3	2.0	4.5	3.8	-63.0	9.6
AT	16.4	4.3	3.7	9.1	10.4	-62.8	7.1
Q	5.8	0.9	0.7	8.0	-1.0	-65.8	0.5
GQ1	2.9	-2.5	-2.9	10.3	-1.8	-66.1	2.6
GQ2	20.8	102.3	104.4	27.3	7.7	-60.9	18.1
GQ3	22.7	15.8	14.2	20.7	19.9	-63.1	11.2
GQ4	48.8	102.8	102.9	57.0	36.2	-51.5	43.5

\* - design-variance estimator (15) is used to estimate design-MSE

\*\* - classic design-MSE estimator not available due to the lack of second order inclusion probabilities

\*\*\* - design-MSE estimator not available (prediction-MSE is usually estimated)