Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 75(2) 2021 117–125

Rescaling Bootstrap Technique for Variance Estimation in Dual Frame Surveys

Rajeev Kumar[1], Anil Rai[1], Tauqueer Ahmad[1], Ankur Biswas[1] and Pramod Kumar Moury[2]
[1]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*
[1]*Amar Singh College, Bulandshahr*

Received 09 October 2020; Revised 08 October 2021; Accepted 18 October 2021

SUMMARY

In a Dual Frame (DF) surveys, set of two frames is used instead of a traditional single frame of sampling units from the target population. Dual frame surveys are applicable in those situations where one frame covers the entire population but very expensive to sample; so an alternate frame may be available that does not cover the entire population but is inexpensive to sample. As Hartley (1962) noted, variance estimation can be more complicated for dual frame surveys than for a single-frame survey. Unbiased variance estimator of parameter of interest is very tedious to obtain for estimator using dual frame surveys. In this article, we propose two rescaling bootstrap variance estimation techniques in dual frame surveys viz. Stratified Rescaling Bootstrap Without Replacement (SRBWO) and Post-stratified Rescaling Bootstrap Without Replacement (PRBWO) methods. Statistical properties of the proposed methods are compared through a simulation study. Simulation results suggest that the proposed SRBWO and PRBWO methods give an unbiased estimate of the variance of the dual frame estimator of population total and the SRBWO method performs better than the PRBWO method.

Keywords: Multiple frame surveys, Rescaling bootstrap, Post stratification, Simulation.

1. INTRODUCTION

In classical finite population sampling, a sample is selected using a probability sampling design from a single sampling frame containing all of the units in the target population. Hartley (1962) noted that in a Multiple Frame (MF) survey, a set of at least two frames is used instead of a traditional single frame of units from the target population. Each frame by itself may or may not be complete but the union is assumed to be complete. Dual Frame (DF) surveys are special case of MF surveys considering two frames covering entire population. The main purpose of using MF surveys is to reduce cost while maintaining estimation efficiency almost at par with a complete but single frame survey. In some cases, a frame that covers the entire population is very expensive, so an alternate frame may be available that does not cover the entire population but is cheaper to sample from. Sampling costs depend on many factors including the size of the sample or the mode of interview. For example, in case of agricultural surveys, in estimation of sugarcane production, making the list of all the farmers producing sugarcane in a particular region is expensive, whereas getting lists of the farmers supplying sugarcane to the sugarcane factory in a region is quite cheaper. Also, the frame developed with the help of the sugarcane factory is the overlapping with the frame of all the farmers producing sugarcane. Here, DF surveys can be used efficiently considering both the frames. Again in many agricultural surveys in different countries, an area frame consists of segments of land that completely covers the entire population, whereas a list frame consisting of the names and addresses of agricultural operators is not complete. Even though the area frame is complete but is very expensive to sample and on the other hand, list frames are usually less costly to sample. Such situations demand DF surveys for reducing the cost to obtain a given precision level.

Hartley (1962, 1974) developed the general theory for dual frames in the efficient estimation of population parameters and studied the special case of Simple Random Sampling (SRS) in both the frames and

Corresponding author: Ankur Biswas
E-mail address: ankur.biswas@icar.gov.in

obtained the optimal dual frame design by minimizing the variance of the estimator for a given cost. In a DF survey, two sampling frames *A* and *B*, together cover the population of interest *U*. Independent probability samples are taken from frames *A* and *B*, and information from the two samples is combined to estimate quantities of interest. In DF surveys, the population can be divided into three non-overlapping domains, viz. '*a*', '*b*' and '*ab*'. Capital letter subscripts are used to indicate population sizes, sample sizes, sample values, sample means, population means, population totals, and costs when they refer to a sampling frame. Small letter subscripts are used for the same quantities when they apply to domains. Usually, independent probability samples are drawn from frames *A* and *B* and samples in each frame are, further, post-stratified into two domains. It is notable that all the domain sizes are random in nature. When domain sizes are known, Hartley (1962) proposed the post stratified estimator of the population total (Y) under dual frame survey as given by

$$\hat{Y} = N_a \bar{y}_a + N_{ab}(p\,\bar{y}_{ab} + q\,\bar{y}_{ba}) + N_b \bar{y}_b \tag{1}$$

and the variance of the post stratified estimator of the population total was given by

$$V(\hat{Y}) = \frac{N_A}{n_A}\left[N_a\sigma_a^2 + N_{ab}p^2\sigma_{ab}^2\right] + \frac{N_B}{n_B}\left[N_b\sigma_b^2 + N_{ab}q^2\sigma_{ab}^2\right] \tag{2}$$

where, $\bar{y}_a = \sum_{i=1}^{n_a}\frac{y_i}{n_a} = \frac{y_a}{n_a}$, $\bar{y}_{ab} = \sum_{i=1}^{n_{ab}}\frac{y_i}{n_{ab}} = \frac{y_{ab}}{n_{ab}}$, $\bar{y}_b = \sum_{i=1}^{n_b}\frac{y_i}{n_b} = \frac{y_b}{n_b}$, $\bar{y}_{ba} = \sum_{i=1}^{n_{ba}}\frac{y_i}{n_{ba}} = \frac{y_{ba}}{n_{ba}}$, σ_a^2, σ_b^2 and σ_{ab}^2 are the within the post-strata variances, *p* and *q* represent proportion of frame *A* and proportion of frame *B* attached to the sample estimate of the domain (*ab*) such that $p + q = 1$. The optimum value for *p* was obtained by minimizing the variance function as presented in Cochran (1965) is given by

$$p = \frac{n_A N_B}{n_A N_B + n_B N_A} \text{ and } q = \frac{n_B N_A}{n_A N_B + n_B N_A}. \tag{3}$$

Following Hartley's approach, many authors have considered the estimation of population parameters under MF surveys. See Cochran (1965), Lund (1968), Fuller and Burmeister (1972), Saxena *et al.* (1984), Bankier (1986), Skinner (1991), Skinner *et al.* (1994), Skinner and Rao (1996), Rao and Skinner (1996), Lohr and Rao (2000, 2006), Singh and Wu (2003), Rao and Wu (2010).

A resampling method that is more promising, widely applicable and more dependable was introduced by Efron (1979) and named as Bootstrap Method. It correctly estimates the variance of the sample median and variances in the area of order statistics. The method is computationally intensive and provides the distribution of statistic and other important measures such as standard deviation of a statistic much more easily. The method proposed by Efron (1979) in the case of an independent and identically distribution (i.i.d.) sample where he suggested drawing SRS with replacement (SRSWR) sample of size *n* repeatedly from an original i.i.d. sample of size *n*. It is also known as the Naive Bootstrap method. Rao and Wu (1984) have shown that the bootstrap method of inference can easily be extended to stratified samples and proposed a method of bootstrapping known as "Rescaling bootstrap with replacement" in which a simple random sample of size m_h is drawn with replacement from an original sample of size n_h in stratum *h*, independently for each stratum. Ahmad (1997) proposed a "Rescaling Bootstrap Without Replacement (RSBWO)" technique of variance estimation for without replacement sampling designs that was more efficient than other bootstrap methods. Biswas *et al.* (2013, 2018) proposed rescaling jackknife and bootstrap technique for variance estimation for ranked set sampling in a finite population. Based on the findings, Dasgupta *et al.* (2018) developed a new sampling theory under dual frame surveys using Ranked Set Sampling in each frame.

Hartley (1962) noted that a dual frame design can result in considerable cost savings over a single frame design with comparable precision. Major issues and problems with estimation under multiple frame sampling include frame membership identification for all sampled units. This is required to post-stratify samples from different frames into appropriate population domains, estimation of domain totals using multiple samples, lack of information on the domain population sizes, identifying and removing duplicated units from multiple-frame samples, handling the extra variation induced by the random sample sizes and use of auxiliary information for estimation. Furthermore, variance estimation can be more complicated for multiple-frame surveys than for a single-frame survey.

An unbiased variance estimator is very tedious to obtain for an estimator under dual frame surveys. Therefore, in the present paper, an attempt has been made to propose two rescaling bootstrap variance estimation techniques in multiple frame surveys viz. Stratified Rescaling Bootstrap Without Replacement (SRBWO) and Post-stratified Rescaling Bootstrap Without Replacement (PRBWO) methods. In Section 2, we discuss the proposed two rescaling bootstrap variance estimation techniques in multiple frame surveys viz. SRBWO and PRBWO methods. In order to study the statistical performances of the proposed bootstrap methods, a simulation study was carried out. Details of the simulation study, simulation results, and discussions have been given in Section 3. Finally, the concluding remarks are given in Section 4.

2. PROPOSED RESCALING BOOTSTRAP VARIANCE ESTIMATION TECHNIQUES IN DUAL FRAME SURVEYS

In order to find the unbiased estimator of the variance of the estimator (Equation 2) of population total, two Rescaling Bootstrap methods are proposed in the following sub-sections.

2.1 Stratified Rescaling Bootstrap Without Replacement (SRBWO) method

Under dual frame surveys, a sample ( ) of size n_A is drawn from the Frame *A* of size N_A. This selected sample (S_A) is further divided into two post strata, i.e., S_a with size n_a and S_{ab} with size n_{ab} by using post stratification technique. Under this proposed method, since the sample membership for each domain or post strata is already known, each domain or post strata are considered as strata with a known frame. Bootstrap technique is used in the selection of resample $\left(S_a^*\right)$ of size n_a^* from the strata $\left(S_a\right)$ and the selection of resample $\left(S_{ab}^*\right)$ of size n_{ab}^* from the strata $\left(S_{ab}\right)$. The whole process is similarly repeated for Frame *B*. Here, it is notable that the strata S_{ab} with size n_{ab} and S_{ba} with size n_{ba} represent two different post-strata from S_A and S_B respectively Schematic diagram of sample and resample structures are presented in Fig. 1.

In order to unbiasedly estimate the variance of the estimator of population total under dual frame survey involving SRS in both the frames independently, the steps involved in the proposed Stratified Rescaling Bootstrap without replacement (SRBWO) technique are given below:

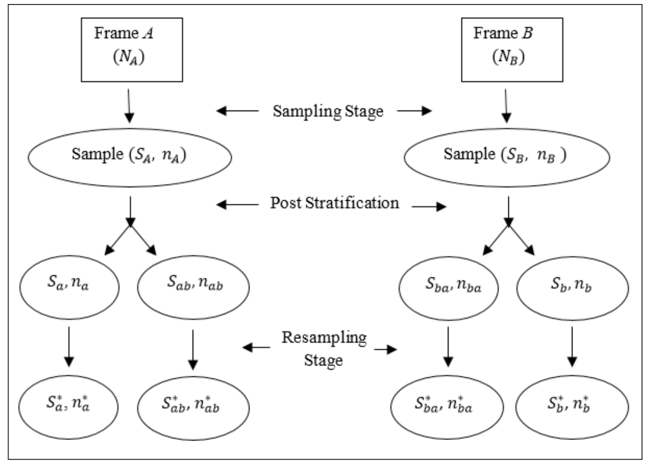

Fig. 1. Schematic diagram of the sample and resample structure under SRBWO method in Dual Frame surveys with two overlapping frames

Step 1. Draw a simple random sample $\left\{y_{ai}^*\right\}_{i=1}^{n_a^*}$ of size $n_a^* < n_a$ without replacement from the strata $\left(S_a\right)$ having observed value $y_{a1},\ y_{a2},\ \ldots,\ y_{an_a}$.

Step 2. Then calculate the following terms as

$$\tilde{y}_{ai} = \bar{y}_a + t_{1a}^{1/2} \frac{\sqrt{n_a^*}}{\sqrt{n_a - n_a^*}} \left(y_{ai}^* - \bar{y}_a\right) \quad \text{and} \quad \tilde{\bar{y}}_a^* = \frac{1}{n_a^*} \sum_{i=1}^{n_a^*} \tilde{y}_{ai} \tag{4}$$

where, $t_{1a} = \dfrac{N_A n_a}{N_a n_A}$.

Similarly, draw independent resamples from the other three domains i.e. *b*, *ab,* and *ba* domain. Then, calculate the similar functions, independently for the other three domains.

$$\tilde{\hat{Y}}^* = N_a \tilde{\bar{y}}_a^* + N_{ab}\left(p\tilde{\bar{y}}_{ab}^* + q\tilde{\bar{y}}_{ba}^*\right) + N_b \tilde{\bar{y}}_b^*.$$

Step 3. Then obtain the following term as

Step 4. Replace the resample in the original sample and independently replicate Step 1 to Step 3. Do this process for a large number of times, say M times and obtain $\tilde{\hat{Y}}_1^*,\ \tilde{\hat{Y}}_2^*,\ \ldots,\ \tilde{\hat{Y}}_M^*$.

Step 5. The bootstrap variance estimator of $\tilde{\hat{Y}}^*$ is given by

$$\hat{\tilde{V}}_{boot} = V_*\left(\tilde{\hat{Y}}^*\right) = E_*\left[\tilde{\hat{Y}}^* - E_*\left(\tilde{\hat{Y}}^*\right)\right]^2 \tag{5}$$

where, E_* and V_* denotes the expectation and variance with respect to the bootstrap sampling from a given sample.

The Monte Carlo estimator of variance as an approximation to $\hat{\hat{V}}_{boot}$ is given by

$$\hat{\hat{V}}_{boot}(a) = \frac{1}{M-1}\sum_{m=1}^{M}\left(\tilde{\hat{Y}}_m^* - \tilde{\hat{Y}}_a^*\right)^2 \tag{6}$$

where, Monte Carlo mean, $\tilde{\hat{Y}}_a^* = \frac{1}{M}\sum_{m=1}^{M}\tilde{\hat{Y}}_m^*$.

Now, it is important to prove that the proposed SRBWO technique provides an unbiased estimator of the variance of estimator of population total under dual frame survey. From Equation (5), we get

$$\hat{\hat{V}}_{boot} = V_*\left(\tilde{\hat{Y}}^*\right) = N_a^2 V_*\left(\tilde{\bar{y}}_a^*\right) + N_{ab}^2 p^2 V_*\left(\tilde{\bar{y}}_{ab}^*\right) + N_{ab}^2 q^2 V_*\left(\tilde{\bar{y}}_{ba}^*\right) + N_b^2 V_*\left(\tilde{\bar{y}}_b^*\right) \tag{7}$$

Now, using equation (4), we can write

$$V_*\left(\tilde{\bar{y}}_a^*\right) = V_*\left(\frac{1}{n_a^*}\sum_{i=1}^{n_a^*}\tilde{y}_{ai}\right) = t_{1a}\frac{n_a^*}{n_a - n_a^*}V_*\left(\bar{y}_a^*\right) = \frac{N_A}{n_A}\frac{s_a^2}{N_a}. \tag{8}$$

Similarly, we can obtain other bootstrap variance terms. Then, the $\hat{V}_{boot}$ becomes

$$\hat{\hat{V}}_{boot} = N_a^2\left[\frac{N_A}{n_A}\frac{s_a^2}{N_a}\right] + N_{ab}^2 p^2\left[\frac{N_A}{n_A}\frac{s_{ab}^2}{N_{ab}}\right] + N_{ab}^2 q^2\left[\frac{N_B}{n_B}\frac{s_{ab}^2}{N_{ab}}\right] + N_b^2\left[\frac{N_B}{n_B}\frac{s_b^2}{N_b}\right]. \tag{9}$$

Now, by taking expectation and considering for large N, i.e. $S_a^2 \cong \sigma_a^2$, we get

$$E\left[V_*\left(\tilde{\bar{y}}_a^*\right)\right] = \frac{N_A n_a}{N_a n_A}\frac{E\left(s_a^2\right)}{n_a} = \frac{N_A}{n_A}\frac{S_a^2}{N_a} \cong \frac{N_A}{n_A}\frac{\sigma_a^2}{N_a}.$$

Similar expectations can be obtained for the post-strata (*ab*), (*b*), and (*ba*) respectively. Then, by taking expectation on equation (9) and for large N considering $S_i^2 \cong \sigma_i^2$, we get

$$E\left[\hat{\hat{V}}_{boot}\right] \cong N_a^2\left[\frac{N_A\sigma_a^2}{N_a n_A}\right] + N_{ab}^2 p^2\left[\frac{N_A\sigma_{ab}^2}{N_{ab}n_A}\right] + N_{ab}^2 q^2\left[\frac{N_B\sigma_{ab}^2}{N_{ab}n_B}\right] + N_b^2\left[\frac{N_B\sigma_b^2}{N_b n_B}\right]$$

$$= \frac{N_A}{n_A}\left[N_a\sigma_a^2 + N_{ab}p^2\sigma_{ab}^2\right] + \frac{N_B}{n_B}\left[N_b\sigma_b^2 + N_{ab}q^2\sigma_{ab}^2\right] = V\left(\hat{Y}\right).$$

Hence, the proposed SRBWO technique results in an approximately unbiased estimator of the variance of the estimator of population total under a dual frame survey.

2.2 Post-stratified Rescaling Bootstrap Without Replacement (PRBWO) method

Under dual frame surveys, Frame A is of size N_A from which a sample (S_A) of size n_A is drawn. Under the proposed Post-stratified Rescaling Bootstrap Without Replacement (PRBWO) method, at first, the resampling technique is used in the selection of resample (S_A^*) of size n_A^* from the sample ( ). This selected resample (S_A^*) can further be divided into the following post-strata, i.e., S_a^* with size n_a^* and S_{ab}^* with size n_{ab}^*. It is notable that n_a^* and n_{ab}^* are random variables. The whole process is similarly repeated for Frame B. Schematic diagram of sample and resample structures are presented in Fig. 2.

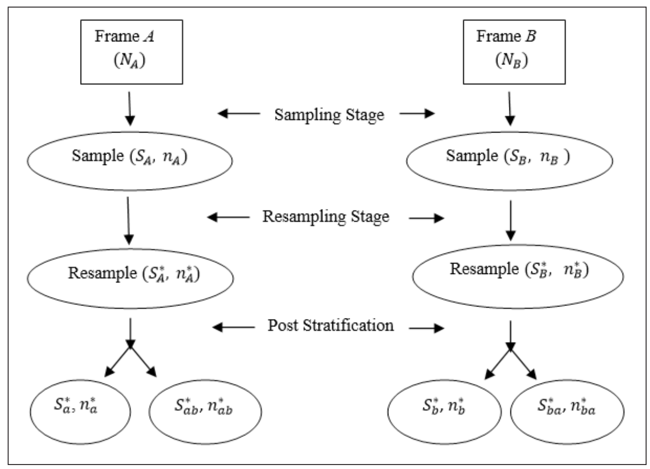

Fig. 2. Schematic diagram for sample and resample structure under PRBWO method in Dual Frame surveys with two overlapping frames

In order to unbiasedly estimate the variance of estimator of population total under dual frame surveys involving SRS in both the frames independently, the steps involved in the proposed PRBWO technique are given below:

Step 1. Draw a simple random sample without replacement $\{y_{Ai}^*\}_{i=1}^{n_A^*}$ of size $n_A^* < n_A$ from the observed sample (S_A) of the Frame A. The resample $\{y_{Ai}^*\}$ can be further divided into post-strata i.e. $\{y_{ai}^*\}_{i=1}^{n_a^*}$ and $\{y_{abi}^*\}_{i=1}^{n_{ab}^*}$ using post stratification technique.

Step 2. Then, calculate the following terms for the post-strata $\{y_{ai}^*\}_{i=1}^{n_a^*}$ as

$$\tilde{y}_{ai} = \bar{y}_a + t_{2a}^{1/2}\frac{\sqrt{n_A^*}}{\sqrt{n_A}}\left(y_{ai}^* - \bar{y}_a\right) \text{ and } \tilde{\bar{y}}_a^* = \frac{1}{n_a^*}\sum_{i=1}^{n_a^*}\tilde{y}_{ai} \tag{10}$$

where, $t_{2a} = \dfrac{N_A n_a}{N_a\left(n_A - n_A^*\right)}$.

Similarly, calculate the above functions for the other post-strata i.e. $\left\{y_{abi}^*\right\}_{i=1}^{n_{ab}^*}$.

Step 3. Repeat steps 1-2 for the other frame B and calculate similar rescaled functions using observed post-strata from the resample.

Step 4. Calculate,

$$\tilde{\hat{Y}}^* = N_a \tilde{\bar{y}}_a^* + N_{ab}\left(p\tilde{\bar{y}}_{ab}^* + q\tilde{\bar{y}}_{ba}^*\right) + N_b \tilde{\bar{y}}_b^*$$

Step 5. Replace the resample in the original sample and independently replicate Step 1 to Step 4. Do this process for a large number of times, say M times and obtain $\tilde{\hat{Y}}_1^*$, $\tilde{\hat{Y}}_2^*$, ..., $\tilde{\hat{Y}}_M^*$.

Step 6. The bootstrap variance estimator of $\tilde{\hat{Y}}^*$ is given by

$$\hat{\tilde{V}}_{boot} = V_*\left(\tilde{\hat{Y}}^*\right) = E_*\left[\tilde{\hat{Y}}^* - E_*\left(\tilde{\hat{Y}}^*\right)\right]^2 \tag{11}$$

where, E_* and V_* denotes the expectation and variance with respect to the bootstrap sampling from a given sample.

The Monte Carlo estimator of variance as an approximation to $\hat{\tilde{V}}_{boot}$ is given by

$$\hat{\tilde{V}}_{boot}(a) = \frac{1}{M-1}\sum_{m=1}^{M}\left(\tilde{\hat{Y}}_m^* - \tilde{\hat{Y}}_a^*\right)^2 \tag{12}$$

where, the Monte Carlo mean, $\tilde{\hat{Y}}_a^* = \dfrac{1}{M}\sum_{m=1}^{M}\tilde{\hat{Y}}_m^*$.

Now, it is important to prove that the proposed PRBWO technique provides an unbiased estimator of the variance of estimator of population total under a dual frame survey. From Equation (11), we get

$$\hat{\tilde{V}}_{boot} = V_*\left(\tilde{\hat{Y}}^*\right) = N_a^2 V_*\left(\tilde{\bar{y}}_a^*\right) + N_{ab}^2 p^2 V_*\left(\tilde{\bar{y}}_{ab}^*\right) + N_{ab}^2 q^2 V_*\left(\tilde{\bar{y}}_{ba}^*\right) + N_b^2 V_*\left(\tilde{\bar{y}}_b^*\right).$$

Following the approach of Cochran (1965) under the proposed PRBWO technique and using Equation (10), the first component of the above variance with respect to the bootstrap sampling can be written as

$$V_*\left(\tilde{\bar{y}}_a^*\right) = V_*\left[\bar{y}_a\left(1 - t_{2a}^{1/2}\frac{\sqrt{n_A^*}}{\sqrt{n_A}}\right)\right] + V_*\left[t_{2a}^{1/2}\frac{\sqrt{n_A^*}}{\sqrt{n_A}}\bar{y}_a^*\right]$$

$$= t_{2a}\frac{n_A^*}{n_A}V_*\left(\bar{y}_a^*\right) = t_{2a}\frac{n_A^*}{n_A}\frac{n_A^2}{n_a^2}V_*\left(W_a^*\bar{y}_a^*\right) \cong t_{2a}\frac{n_A^*}{n_A}\frac{n_A^2}{n_a^2}\left(\frac{1}{n_A^*} - \frac{1}{n_A}\right)W_a^* s_a^2$$

$$= \frac{N_A}{n_A}\frac{s_a^2}{N_a}$$

where, $\bar{y}_a^* = \dfrac{1}{n_a^*}\sum_{i=1}^{n_a^*} y_{ai}^*$, $\bar{y}_a = \dfrac{1}{n_a}\sum_{i=1}^{n_a} y_{ai}$, $V_*\left(\bar{y}_a\right) = 0$ and weight of post-strata (a) in the bootstrap resample as $W_a^* = n_a/n_A$.

Similarly, we can obtain other bootstrap variance terms. Then, using all these bootstrap variance terms for all post-strata, the $\hat{V}_{boot}$ becomes

$$\hat{\tilde{V}}_{boot} = N_a^2\left[\frac{N_A}{n_A}\frac{s_a^2}{N_a}\right] + N_{ab}^2 p^2\left[\frac{N_A}{n_A}\frac{s_{ab}^2}{N_{ab}}\right] + N_{ab}^2 q^2\left[\frac{N_B}{n_B}\frac{s_{ab}^2}{N_{ab}}\right] + N_b^2\left[\frac{N_B}{n_B}\frac{s_b^2}{N_b}\right]. \tag{13}$$

Now, by taking expectation of $V_*\left(\tilde{\bar{y}}_a^*\right)$ and for large sample N considering $S_a^2 \cong \sigma_a^2$, we get

$$E\left[V_*\left(\tilde{\bar{y}}_a^*\right)\right] = \frac{N_A n_a}{N_a n_A}\frac{E\left(s_a^2\right)}{n_a} = \frac{N_A}{n_A}\frac{S_a^2}{N_a} \cong \frac{N_A}{n_A}\frac{\sigma_a^2}{N_a}.$$

Similar expectations can be obtained for the post-strata (*ab*), (*b*), and (*ba*) respectively. Then, by taking expectation on equation (13) and for large N considering $S_i^2 \cong \sigma_i^2$, we get

$$E\left[\tilde{\hat{V}}_{boot}\right] \cong N_a^2\left[\frac{N_A\sigma_a^2}{N_a n_A}\right] + N_{ab}^2 p^2\left[\frac{N_A\sigma_{ab}^2}{N_{ab} n_A}\right] + N_{ab}^2 q^2\left[\frac{N_B\sigma_{ab}^2}{N_{ab} n_B}\right] + N_b^2\left[\frac{N_B\sigma_b^2}{N_b n_B}\right]$$

$$= \frac{N_A}{n_A}\left[N_a\sigma_a^2 + N_{ab}p^2\sigma_{ab}^2\right] + \frac{N_B}{n_B}\left[N_b\sigma_b^2 + N_{ab}q^2\sigma_{ab}^2\right]$$

$$= V\left(\hat{Y}\right).$$

Therefore, the proposed PRBWO technique gives an approximately unbiased estimator of the variance of the estimator of population total under dual frame surveys.

3. SIMULATION STUDY

The performance of the proposed SRBWO and PRBWO variance estimation procedures in the case of dual frame surveys was examined by a simulation study. Under the simulation study, first, a univariate normal population was generated using R software of size 4000. The mean and variance of the generated univariate normal population are 320 and 28.079 respectively. The size of the population domains (a), (ab), and (b) were chosen as 1000, 2000, and 1000 respectively, and accordingly, 4000 population units were randomly allocated to these domains. In this way, the size of each frame (A and B) becomes 3000. Further, 5000 different samples of several sample sizes were drawn from these simulated populations using dual frame sampling involving SRSWOR from both the frames. The estimates of post stratified estimator of population total under dual frame sampling were computed from these 5000 independent samples for each sample size separately. The theoretical variance of the dual frame estimator was computed based on Equation (2). Further, from each of these 5000 selected dual frame samples, 300 independent bootstrap resamples were drawn following both the proposed rescaling bootstrap variance estimation techniques in dual frame surveys i.e. Stratified Rescaling Bootstrap Without Replacement (SRBWO) and Post stratified Rescaling Bootstrap Without Replacement (PRBWO) methods. The Monte Carlo bootstrap estimates of the variance of the dual frame estimator of population total were obtained following the proposed SRBWO and PRBWO for different sample sizes using Equations (6) and (12) respectively. In order to compare the performance of these proposed rescaling bootstrap variance estimation techniques of the dual frame estimator of population total, percentage Relative Bias (% RB) and Relative Stability (RS) were obtained using the formulae given by

$$\%RB = \left[\frac{\frac{1}{s}\sum_{s}\left\{\hat{V}_s\left(\hat{Y}^*\right)\right\} - V\left(\hat{Y}\right)}{V\left(\hat{Y}\right)}\right] \times 100 \text{ and}$$

$$RS = \frac{\sqrt{\frac{1}{s}\sum_{s}\left\{\hat{V}_s\left(\hat{Y}^*\right) - V\left(\hat{Y}\right)\right\}^2}}{V\left(\hat{Y}\right)}.$$

where, $\hat{V}_s\left(\hat{Y}^*\right)$ is the Monte Carlo estimates of the variance of post stratified estimator obtained through a particular rescaling bootstrap variance estimation technique and s denotes the number of resamples selected for bootstrap variance estimation

3.1 Simulation results and discussion

The simulation results showing statistical properties of the proposed rescaling bootstrap variance estimation techniques in multiple frame surveys i.e. SRBWO and PRBWO methods are presented in this section. The Monte Carlo (MC) bootstrap estimate of the variance of post stratified estimator of population total under a dual frame survey, percentage Relative Bias (%RB),

Table 1. Statistical properties of the proposed Stratified Rescaling Bootstrap Without Replacement (SRBWO) technique along with its Standard version without any rescaling factor

Sample Size	Bootstrap Sample Size	Standard Bootstrap method (without Rescaling factor)				Proposed SRBWO method (with Rescaling factor)			
		Variance of estimator	Estimate of variance	% RB	RS	Variance of estimator	Estimate of variance	% RB	RS
200	40	33617762	135470659	302.973	3.056	33617762	33618978	0.003	0.092
200	60	33617762	79000490	134.996	1.370	33617762	33609110	-0.025	0.092
200	80	33617762	50826198	51.188	0.534	33617762	33636664	0.056	0.093
300	60	22411841	89959057	301.391	3.033	22411841	22361566	-0.224	0.079
300	90	22411841	52570150	134.564	1.360	22411841	22419602	0.035	0.079
300	120	22411841	33816150	50.885	0.524	22411841	22448011	0.161	0.079
600	120	11205921	44891853	300.608	3.017	11205921	11199268	-0.059	0.060
600	180	11205921	26203658	133.838	1.347	11205921	11210444	0.04	0.063
600	240	11205921	16860581	50.461	0.513	11205921	11217796	0.106	0.061
900	180	7470614	29962998	301.078	3.019	7470614	7480380	0.131	0.054
900	270	7470614	17489857	134.115	1.348	7470614	7486249	0.209	0.054
900	360	7470614	11234532	50.383	0.511	7470614	7481760	0.149	0.054
1200	240	5602960	22429299	300.312	3.010	5602960	5602749	-0.004	0.051
1200	360	5602960	13085116	133.539	1.341	5602960	5603758	0.014	0.051
1200	480	5602960	8420579	50.288	0.509	5602960	5609380	0.115	0.050

and Relative Stability (RS) of the proposed variance estimation techniques were calculated for different sample sizes and the corresponding bootstrap resample sizes and presented in Table 1 and Table 2 respectively. Standard versions of both the bootstrap techniques without using any rescaling factors are also given. The theoretical variance of the post stratified estimator of population total under dual frame survey for different sample sizes based Equation (2) is given in both the tables.

The following results can be observed in Table 1 and Table 2:

- The variance of the post stratified estimator of population total under dual frame survey decreases with the increase of sample size. It shows that this estimator is design consistent. .
- It can be observed that both the Standard Bootstrap version (without any rescaling factors) of the proposed SRBWO and PRBWO methods give a very large amount of %RB and RS for estimation of the variance of the post stratified estimator of population total under dual frame surveys. On the contrary, when proposed rescaling factors are used, both the SRBWO and PRBWO methods show a very less amount of %RB and RS in the estimates of the variance of the post stratified estimator. Thus, proposed rescaling factors were quite effective in reducing %RB and RS considerably as compared to the Standard Bootstrap version of SRBWO and PRBWO methods for variance estimation of the dual frame estimator.
- The %RB is very less for the proposed rescaling bootstrap variance estimation techniques for estimation of the variance of the dual frame estimator of population total. Therefore, the proposed variance estimation procedures are almost unbiased for the variance of the dual frame estimator of population total as established theoretically.
- RS of both methods are generally very less, close to zero, and decreases with the increase of sample size. Therefore, it can be concluded that the estimator of the variance of the dual frame estimator obtained following both the proposed rescaled bootstrap methods are stable in nature for different sample sizes.
- While comparing the %RB of the proposed SRBWO and PRBWO methods, it can be seen that %RB of the SRBWO method is consistently lesser than the PRBWO method and in most sample size cases %RB is quite negligible for the SRBWO method. The absolute value of % RB ranges from 0.004 to

Table 2. Statistical properties of the proposed Post-stratified Rescaling Bootstrap Without Replacement (PRBWO) method along with its Standard version without any rescaling factors

Sample Size	Bootstrap Sample Size	Standard Bootstrap method (without Rescaling factor)				Proposed PRBWO method (with Rescaling factor)			
		Variance of estimator	Estimate of variance	% RB	RS	Variance of estimator	Estimate of variance	% RB	RS
200	20	33617762	141263655	320.205	3.231	33617762	35057139	2.664	0.107
200	30	33617762	81172007	141.455	1.435	33617762	34513642	2.094	0.099
200	40	33617762	51868460	54.288	0.565	33617762	34321890	2.575	0.099
300	20	22411841	92477356	312.627	3.147	22411841	22989144	2.576	0.086
300	30	22411841	53554109	138.955	1.404	22411841	22841082	1.915	0.083
300	40	22411841	34178669	52.503	0.540	22411841	22686584	1.226	0.080
600	20	11205921	45506743	306.096	3.072	11205921	11352114	1.305	0.063
600	30	11205921	26412236	135.699	1.366	11205921	11300192	0.841	0.063
600	40	11205921	16953980	51.295	0.522	11205921	11279660	0.658	0.062
900	20	7470614	30182850	304.021	3.049	7470614	7535250	0.865	0.056
900	30	7470614	17570802	135.199	1.358	7470614	7521037	0.675	0.055
900	40	7470614	11267004	50.818	0.515	7470614	7503409	0.439	0.054
1200	20	5602960	22610327	303.543	3.043	5602960	5648049	0.805	0.052
1200	30	5602960	13162144	134.914	1.355	5602960	5636802	0.604	0.052
1200	40	5602960	8449207	50.799	0.514	5602960	5628542	0.457	0.052

0.22 and 0.44 to 2.66 in the case of the proposed SRBWO and PRBWO method respectively. Thus, with respect to %RB, the proposed SRBWO is more efficient than the PRBWO method in the estimation of the variance of the dual frame estimator.

- It can also be observed that in case of the SRBWO and PRBWO method %RB decreases with the increase of sample size.
- While comparing the RS of the proposed SRBWO and PRBWO methods, it can be seen that the values of RS of the SRBWO procedure are consistently lesser than PRBWO methods for a different combination of sample sizes. Thus, the estimator of the variance of the dual frame estimator obtained by the SRBWO method is more stable than the PRBWO method.

4. CONCLUSIONS

In this article, two different unbiased variance estimation procedures for the post stratified estimator of population total under dual frame surveys i.e. SRBWO and PRBWO method are proposed in order to unbiasedly estimate the sampling variance of the post stratified estimator of population total under dual frame surveys. Under these proposed procedures, resamples are taken domain-wise as well as frame-wise respectively and rescaling factors are obtained for each cases under a dual frame survey. It has been shown theoretically that the proposed estimators of variance in both the procedures become almost unbiased for the variance of the dual frame estimator of finite population total. Further, a statistical comparison of the proposed bootstrap variance estimation procedures was done through a simulation study. Due to the use of proposed rescaling factors, the %RB and RS of proposed SRBWO and PRBWO are reduced significantly from their standard versions without considering any rescaling factors. Simulation results suggest both the proposed procedures show very less amount of %RB and RS in the estimation of the variance of the dual frame estimator of population total. It was observed that the % RB in the PRBWO method is relatively more than the SRBWO method for all different combinations of sample sizes. Therefore, it can be concluded that the variance estimation procedure following the SRBWO method is more efficient and stable than the PRBWO method with respect to %RB and RS for different sample sizes. Thus, the SRBWO method is preferable over the PRBWO method. This resampling procedure can also be extended for MF surveys consisting of more than two frames.

ACKNOWLEDGEMENTS

Authors are very thankful to the reviewer for giving valuable suggestions which led to improvement of this article.

REFERENCES

Ahmad, T. (1997). A resampling technique in complex survey data. *J. Ind. Soc. Agril. Statist.*, **50(3)**, 364-379.

Bankier, M.D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. *J. Amer. Statist. Assoc.*, **81**, 1074-1079.

Berger, Y.G. (2007). A jackknife variance estimator for unistage stratified samples with unequal probabilities, *Biometrika*, **94(4)**, 953-964.

Biswas, A., Ahmad, T. and Rai, A. (2013). Variance estimation using Jackknife method in ranked set sampling under finite population framework. *J. Ind. Soc. Agril. Statist.*, **67(3)**, 345-353.

Biswas, A., Rai, A. and Ahmad, T. (2020). Rescaling bootstrap technique for variance estimation for ranked set samples in finite population. *Communications in Statistics - Simulation and Computation*, **49(10)**, 2704-2718. DOI: 10.1080/03610918.2018.1527349.

Chipperfield, J., and Preston, J. (2007). Efficient bootstrap for business surveys. *Survey Methodology*, **33**, 167-172.

Cochran, W.G. (1963). *Sampling Techniques*. Second edition. New York, N, Y., John Wiley and Sons, Inc.

Cochran, R.S. (1965). *Theory and application of Multiple frame surveys*. Iowa State University of Science and Technology Ames, Iowa.

Dasgupta, P., Ahmad, T., Biswas, A. and Rai, A. (2018). A dual frame approach for estimating finite population total using ranked set sampling. *Int. J. Agril. Statist. Sci.*, **14(1)**, 409-418, 2018.

Efron, B. (1979). Bootstrap methods: Another look at jackknife. *The Annals of Statistics*, **7**, 1-26.

Efron, B. (1982). Nonparametric estimates of standard error: The Jackknife, the bootstrap and other methods. *Biometrika*, **68**, 589-599.

Fuller, W.A. and Burmeister, L.F. (1972). Estimators for samples selected from two overlapping frames. Proceedings of the Social Statistics Section, *American Statistical Association*, 245-249.

Hartley, H.O. (1962). Multiple frame surveys. Proceedings of the social statistics sections. *American Statistical Association*, **19(6)**, 203-206.

Hartley, H.O. (1974). Multiple frame methodology and selected applications. *Sankhya, Series C*, **36(9)**, 99-118.

Lohr, S. (2011). Alternative survey sample designs: Sampling with multiple overlapping frames. *Survey Methodology*, **37(2)**, 197-213.

Lohr, S. and Rao, J.N.K. (2000). Inference from Dual Frame Surveys. *J. Amer. Statist. Assoc.*, **95(449)**, 271-280.

Lohr, S. and Rao, J.N.K. (2006). Estimation in multiple frame surveys. *J. Amer. Statist. Assoc.*, **101**, 1019-1030.

Lund, R. (1968). Estimators in multiple frame surveys. *Proceedings of the Social Statistics Sections, American Statistical Association*, 282-288.

Rao, J.N.K. and Skinner, C.J. (1996). Estimation in dual frame surveys with complex designs. *Proceedings of the Survey Methods Section, Statistical Society of Canada*, 63-68.

Rao, J.N.K. and Wu. C.F.J. (1984). Bootstrap inference for sample surveys. In *Proc. Sec. Sur. Res. Meth.*, 106-12, ASA, Washington, D.C.

Rao, J.N.K. and Wu, C.F.J. (1988). Resampling inference with complex survey data. *J. Amer. Statist. Assoc.*, **83(401)**, 231-241.

Rao, J.N.K. and Wu, C. (2010). Pseudo–Empirical Likelihood Inference for Multiple Frame Surveys. *J. Amer. Statist. Assoc.*, **105(492)**, 1494-1503.

Shao, J. (2003). Impact of bootstrap on sample surveys, *Statistical Science*, **18(2)**, 191-198.

Singh, A. and Wu, S. (2003). An extension of generalized regression estimator to dual frame surveys. *Proceedings of the Joint Statistical Meeting - Section on Survey Research Methods*, 3911–3918.

Sitter, R.R. (1992). A resampling procedure from complex survey data. *J. Amer. Statist. Assoc.*, **87(419)**, 755-765.

Skinner, C.J. (1991). On the efficiency of raking ratio estimation for multiple frame surveys. *J. Amer. Statist. Assoc.*, **86(415)**, 779-784.

Skinner, C.J., Holmes, D.J. and Holt, D. (1994). Multiple frame sampling for multivariate stratification. *International Statistical Review*, **62**, 333-347.