Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 74(2) 2020 137–147

Rescaled Spatial Bootstrap Variance Estimation of Spatial Estimator of Finite Population Parameters under Ranked Set Sampling

Ankur Biswas, Anil Rai, Tauqueer Ahmad and Prachi Misra Sahoo

ICAR-Indian Agricultural Statistics Research Institute, New Delhi

Received 31 December 2019; Revised 08 June 2020; Accepted 14 July 2020

SUMMARY

Ranked Set Sampling (RSS) is preferred over Simple Random Sampling (SRS) when measuring an observation is expensive or time consuming, but can be easily ranked at a negligible cost. Biswas *et al.* (2015) proposed a Spatial Estimator (SE) of population mean under RSS through prediction approach incorporating spatial dependency among sampling units of a spatial finite population. In this present article, an attempt has been made to propose bootstrap techniques viz. Rescaled Spatial Stratified Bootstrap (RSSB) and Rescaled Spatial Clustered Bootstrap (RSCB) methods for unbiased variance estimation of the SE under RSS from finite populations. Simulation study reveals that both the proposed methods give approximately unbiased estimation of variance of the SE under RSS for different combination of sample and bootstrap sample sizes, but while considering relative stability, RSSB method was found to be more stable.

Keywords: Ranked set sampling, Prediction approach, Inverse distance weighting, Ranks, Cycles.

1. INTRODUCTION

Ranked Set Sampling (RSS), proposed by McIntyre (1952), is a well-known sampling scheme for estimation of population parameters which provides more precise estimator than Simple Random Sampling Without Replacement (SRSWOR) when actual measurements of target variables are either difficult or expensive in terms of time, money or labour, but ranking of sampling units on the basis of visual inspection or any other cheaper method, can be done easily. RSS has received increasing attention from statisticians due to its potential for observational economy. RSS has been found to be useful in many research areas like ecological and environmental studies (Dell and Clutter, 1972; Martin *et al.*, 1980; Al-Saleh and Zheng, 2002; Ozturk *et al.*, 2005), agricultural studies (Cobby *et al.*, 1985; Halls and Dell, 1985; Chen *et al.*, 2004; Husby *et al.*, 2005; Bocci *et al.*, 2010), reliability theory (Kvam and Samaniego, 1994; Ghitany, 2005; Dong *et al.*, 2012) and medical studies (Samawi and Al-Sagheer, 2001; Nahhas *et al.*, 2002). In case of crop yield (production) estimation surveys of a region, RSS technique can be useful for ranking of fields with respect to crop yield through visual inspection or any other convenient ways, whereas, actual measurement of crop yield in each selected field can be obtained either by randomly selecting a small plots for harvesting by Crop Cutting Experiments technique or by whole field harvest method. Crop yield estimation in these actual measurement cases are quite costly and time consuming process. Patil *et al.* (1994) have discussed various aspects of RSS in detail. Takahasi and Futatsuya (1988, 1998), Patil *et al.* (1995), Krishna (2002), Sud and Mishra (2006, 2007), Kankure and Rai (2008), Rai and Krishna (2013) made attempts to extend the theory of RSS without replacement in the context of finite population. Deshpande *et al.* (2006) classified without replacement sampling design of RSS under finite population under Level-0, Level-1 and Level-2. Several attempts were made to compute inclusion probabilities (Al-Saleh and Samawi, 2007; Özdemir and Gökpinar, 2007 and 2008; Gökpinar and Özdemir, 2010, 2012, 2014; Jozani and Johnson, 2011, 2012; Frey, 2011; Ozturk and Jozani, 2013; Ozturk, 2014,

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

2016a, 2016b). Wolfe (2012) presents a comprehensive review on RSS.

In most of the agricultural and environmental surveys, often, the parameters of interest are spatial in nature. Neighbouring sampling units share similar kind of attribute values as compared to distant units. This phenomenon suggests spatial dependency in the study data. Classical statistical methods fail to capture such dependency present in the underlying data in case of spatial population (Zhang and Griffith, 2000). Several authors (Hedayat *et al.*, 1988; Arbia, 1993; Sahoo *et al.*, 2006; Kankure and Rai, 2008) have considered these type of situations and proposed techniques to select samples through spatial sampling by assigning less probability of selection to contiguous units of already selected units in the sample. Biswas *et al.* (2015) considered RSS randomization framework to select samples from a spatial finite population and a Spatial Estimator (SE) has been proposed for estimation of population mean using Inverse Distance Weighting (IDW) method (Donald, 1968) through prediction approach (Royall, 1970).

1.1 Spatial Estimator (SE) in case of RSS

Let, there be a spatial finite population of size N consisting of units U_i, $i\in\Omega=\{1, 2, \ldots, N\}$ with mean $\bar{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i$ and finite variance $\sigma_Y^2=\frac{1}{N}\sum_{i=1}^{N}(Y_i-\bar{Y})^2$. Without loss of generality, let, respective population values of the character of interest Y are $Y_1\le Y_2\le\ldots\le Y_N$. Let the parameter of interest be linear in nature i.e. population mean $\bar{Y}=\frac{1}{N}\sum_{q\in\Omega}Y_q$. Let D_{qj} denotes distance between the population units U_q and U_j, q, $j\in\Omega$. In order to select a without replacement sample through RSS design, m^2 units are randomly drawn by SRSWOR from the finite population. Then, this sample is randomly divided into m sets of size m. The units within each set are ranked on the basis of character of interest or some auxiliary information. Smallest ranked unit is accurately quantified from first set, 2nd smallest ranked unit is quantified from 2nd set and so forth, until the largest ranked unit is quantified from mth set. The entire cycle is replicated r times until in total n = mr observations for a ranked set sample are quantified. Let, s denotes the set of all sampled units. The unbiased estimator of the population mean i.e. sample mean (Patil *et al.*, 1995) is given by

$$\bar{y}_{RSS}=\frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m}y_{(i:m)k}\,, \tag{1.1}$$

where $y_{(i:m)k}$ denotes the value of the $(i{:}m)k^{th}$ unit i.e. i^{th} ranked unit in the i^{th} set of size m in the k^{th} cycle of the sample.

Let $d_{(i:m)k,\,j}$ denote the distance between the $(i{:}m)k^{th}$ sampled unit and any other non-sampled unit U_j, $j\in\bar{s}$, where $\bar{s}$ is the set of all non-sample units. Following the prediction approach of Royall (1970), Biswas *et al.* (2015) proposed following Spatial Estimator (SE) of population mean, $\bar{Y}$, under RSS design

$$\hat{\bar{Y}}_{SE,RSS}=\left[mr\,\bar{y}_{RSS}+(N-mr)\,\bar{y}_{p,RSS}\right]\Big/N\,. \tag{1.2}$$

where,

$$\bar{y}_{p,RSS}=\frac{1}{N-mr}\sum_{j\in\bar{s}}\hat{y}_{j,p}\quad\text{and}$$

$$\hat{y}_{j,p}=\left(\sum_{k=1}^{r}\sum_{i=1}^{m}(y_{(i:m)k}/d_{(i:m)k,j})\right)\Big/\left(\sum_{k=1}^{r}\sum_{i=1}^{m}(1/d_{(i:m)k,j})\right);$$

$$(i:m)k\in s\text{ and }j\in\bar{s}$$

is the predicted value of j^{th} non-sampled unit, Y_j, $j\in\bar{s}$ by IDW technique. It is notable that value of $d_{(i:m)k,\,j}$ is random as it depends on selection of $(i{:}m)k^{th}$ sampled unit.

In case of large population size i.e. for large N, Biswas *et al.* (2015) provided the approximate sampling variance of the Spatial estimator, $\hat{\bar{Y}}_{SE,RSS}$, under RSS as

$$V\left(\hat{\bar{Y}}_{SE,RSS}\right)\cong\left(\frac{mr+(N-mr)\bar{\bar{D}}}{N}\right)^2\frac{1}{mr}\left(\frac{N-1-mr}{N-1}\sigma_Y^2-\bar{\gamma}_Y\right), \tag{1.3}$$

where,

$$\bar{\bar{D}}=\frac{1}{N}\sum_{j\in\Omega}\bar{D}'_j\,,\quad \bar{D}'_j=\frac{1}{\bar{R}_{2j}\,\bar{D}_{.j}}\,,\quad \bar{R}_{2j}=\frac{1}{N-1}\sum_{\substack{q\in\Omega\\q\neq j}}\frac{1}{D_{qj}}\,,$$

$$\bar{D}_{.j}=\frac{1}{N-1}\sum_{\substack{q\in\Omega\\q\neq j}}D_{qj}\,,\quad \bar{\gamma}_Y=\frac{m!\,(m-1)!}{N(N-1)\ldots(N-2m+1)}\gamma_Y\,,$$

$\gamma_Y=(\underline{Y}-\bar{Y})'\,\underset{\sim}{\Gamma}\,(\underline{Y}-\bar{Y})$ and $\underset{\sim}{\Gamma}=\binom{N}{m,m}\sum_{i=1}^{m}B_{ii}$ is a symmetric matrix with zeroes on the diagonal which is a function of *N* and *m* only, independent of population values $\underline{\mathbf{Y}}$.

It was observed that since $\overline{\overline{D}} \le 1$, the variance of the SE of population mean under RSS is always lesser than the usual RSS estimator i.e. $V\left(\hat{\hat{Y}}_{SE,RSS}\right) \le V\left(\overline{y}_{RSS}\right)$.

It has been seen that obtaining an unbiased estimator for variance of usual RSS estimator is quite cumbersome. Bootstrap method is one of the resampling techniques which can be used for unbiased variance estimation. Chen *et al.* (2004), Hui *et al.* (2005), Modarres *et al.* (2006) etc. considered methods of bootstrapping under RSS design in the context of infinite population. Under finite population framework, it is always necessary to unbiasedly estimate variance of an estimator. Therefore, the usual applications of the Naive bootstrap procedures (Efron, 1979) may not estimate variance unbiasedly under finite population randomization framework (Wolter, 1985; Rao and Wu, 1988; Ahmad, 1997). Biswas *et al.* (2013, 2018) suggested rescaled Jackknife and bootstrap techniques for unbiased variance estimation of RSS estimator of population mean under finite population framework. It can be observed that the variance expression of the Spatial Estimator (SE) under RSS design is non-linear in nature and, thus, it becomes even more complicated to obtain unbiased estimator of variance of SE under RSS. Keeping this in view, in this present article attempt has been made to use resampling techniques, viz. bootstrap technique, for unbiased variance estimation of the SE under RSS (more precisely Level-2 as per Deshpande *et al.* (2006)) from finite populations. Therefore, two different Rescaled Spatial Bootstrap (RSB) procedures have been proposed in Section 2 for approximately unbiased estimation of variance of the proposed SE under RSS following the approach of Ahmad (1997) and Biswas *et al.* (2018). Finally, statistical properties of the SE and proposed rescaled spatial bootstrap methods for variance estimation were studied empirically through a simulation study and discussed in Section 3 and 4. Concluding remarks are given in Section 5.

2. PROPOSED RESCALED SPATIAL BOOTSTRAP METHODS FOR VARIANCE ESTIMATION OF SPATIAL ESTIMATOR (SE) UNDER RSS

In order to find approximately unbiased estimator of variance for the proposed SE under RSS two different rescaled spatial bootstrap methods are proposed in the subsequent sections. Let, s^* denotes set of sampling units selected in bootstrap sample selected from original RSS samples, and let t^* be the set of remaining sampling units of the population not selected in the bootstrap sample. Proposed methods are based on prediction of values of remaining units in the population belonging to t^* using the resampled units selected from original RSS sample belonging to s^* considering RSS randomization framework. The expression for variance of RSS estimator given by Patil *et al.* (1995) for moderate to large N can be written as

$$V\left(\overline{y}_{RSS}\right) = \frac{1}{mr}\left\{\frac{N-1-mr}{N-1}\sigma_Y^2 - \overline{\gamma}_Y\right\} \cong \frac{1}{mr}\frac{N-1-mr}{N-1}\left\{\sigma_Y^2 - \overline{\gamma}_Y\right\}. \quad (2.1)$$

Using above approximation the variance of the proposed SE under RSS as shown in equation in (1.3) can be rewritten as

$$V\left(\hat{\hat{Y}}_{SE,RSS}\right) \cong \left(\frac{mr + (N-mr)\overline{\overline{D}}}{N}\right)^2 \frac{N-1-mr}{mr\left(N-1\right)}\left(\sigma_Y^2 - \overline{\gamma}_Y\right). \quad (2.2)$$

2.1 Rescaled Spatial Stratified Bootstrap (RSSB) Method

Rescaled Spatial Stratified Bootstrap (RSSB) method is based on different ranks, i.e. implicit population strata, of the RSS sample. It may be noted that in the population, rank based strata can be assumed and observations of each rank are considered as units within a stratum. The RSS sample of size n = mr is composed of r observations for each of the m ranks. Thus, population is assumed to be divided into m strata and sample from each stratum consisting of r units. The steps involved in RSSB Method for estimation of variance of the proposed SE under RSS are given as

a) Draw a sample $\left\{y^*_{(i:m)k}\right\}_{k=1}^{z}$ of size $z(<r)$ by SRSWOR from the observed sample values $y_{(i:m)1}$, $y_{(i:m)2}$, …, $y_{(i:m)r}$ of the i^{th} stratum.
b) Independently implement step (a) for all ranks $i = 1, \ldots, m$ and obtain a bootstrap sample as $\left\{y^*_{(i:m)k}\right\} \in s^*$, $\forall\ i = 1, \ldots, m$ and $k = 1, \ldots, z$.
c) Then compute,

$$\tilde{y}_{(i:m)k} = \overline{y}_{RSS} + f_1^{1/2}\left(y^*_{(i:m)k} - \overline{y}_{RSS}\right);\ \forall k = 1,\ldots,z \quad \text{and}$$

$$\tilde{\tilde{y}}_{s^*,st} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{z}\sum_{k=1}^{z}\tilde{y}_{(i:m)k}, \quad (2.3)$$

where

$$f_1 = \left[\left\{z\left(N-1-mr\right)\right\}\Big/\left\{\left(r-z\right)\left(N-1\right)\right\}\right]$$
$$\left[\left\{mr+(N-mr)\bar{\bar{D}}\right\}\Big/\left\{mz+(N-mz)\bar{\bar{D}}\right\}\right]^2.$$

d) Using these $\tilde{y}_{(i:m)k}$ predict all the sampling units belongs to $t*$ as

$$\tilde{y}_{j,p} = \left(\sum_{k=1}^{z}\sum_{i=1}^{m}(\tilde{y}_{(i:m)k} / d^*_{(i:m)k,j})\right)\Bigg/\left(\sum_{k=1}^{z}\sum_{i=1}^{m}(1/d^*_{(i:m)k,j})\right),$$
$$(i:m)k \in s* \text{ and } j \in t*$$

where $d^*_{(i:m)k,j}$ denotes distances between specific bootstrap sample unit $(i:m)k \in s*$ and j^{th} non-bootstrap sampling unit which belongs to $t*$.

e) Then work out

$$\tilde{\bar{y}}_{t*,st} = \frac{1}{N-mz}\sum_{j\in t^*}\tilde{y}_{j,p}.$$

f) Finally obtain

$$\tilde{T}_{st} = \left[mz\,\tilde{\bar{y}}_{s*,st} + (N-mz)\tilde{\bar{y}}_{t*,st}\right]\Big/N. \tag{2.4}$$

g) Return the resample in the sample and independently replicate step (a) through (f) for a large number, say B, times and calculate the corresponding $\tilde{T}^1_{st}, \tilde{T}^2_{st}, \ldots, \tilde{T}^B_{st}$.

h) The bootstrap variance estimator of $\tilde{T}_{st}$ is given by

$$\hat{\hat{V}}_{b,st} = V_*\left(\tilde{T}_{st}\right) = E_*\left(\tilde{T}_{st} - E_*\,\tilde{T}_{st}\right)^2 \tag{2.5}$$

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

The Monte Carlo estimator $\hat{\hat{V}}_{b,st}(a)$ as an approximation to $\hat{\hat{V}}_{b,st}$ is given by

$$\hat{\hat{V}}_{b,st}(a) = \frac{1}{B-1}\sum_{b=1}^{B}(\tilde{T}^b_{st} - \bar{T}_{st})^2,$$

where $\bar{T}_{st} = \frac{1}{B}\sum_{b=1}^{B}\tilde{T}^b_{st}$.

Now, by taking expectation on $\hat{\hat{V}}_{b,st}$ with respect to the randomization procedure adopted in original RSS sample selection we get

$$E\left(\hat{\hat{V}}_{b,st}\right) = \frac{1}{N^2}\Big[m^2z^2E\left(V_*(\tilde{\bar{y}}_{s*,st})\right) + (N-mz)^2E\left(V_*(\tilde{\bar{y}}_{t*,st})\right) +$$
$$2mz(N-mz)E\left(Cov_*(\tilde{\bar{y}}_{s*,st}, \tilde{\bar{y}}_{t*,st})\right)\Big]. \tag{2.6}$$

Now, since the resamples are taken independently by SRSWOR from all the ranks, then using Equation (2.3) we get

$$V_*(\tilde{\bar{y}}_{s*,st}) = f_1\,\frac{1}{m^2}\sum_{i=1}^{m}V_*\left[\bar{y}^*_{(i)}\right] + V_*\left(\left(1-f_1^{1/2}\right)\bar{y}_{RSS}\right)$$
$$= f_1\,s^2_{st}$$

where, $s^2_{st} = \frac{1}{m^2}\left(\frac{1}{z}-\frac{1}{r}\right)\frac{1}{r-1}\sum_{i=1}^{m}\left(\sum_{k=1}^{r}y^2_{(i:m)k} - r\,\bar{y}_{(i)}^{\ 2}\right)$,

$$\bar{y}^*_{(i)} = \frac{1}{z}\sum_{k=1}^{z}y^*_{(i:m)k} \text{ and } \bar{y}_{(i)} = \frac{1}{r}\sum_{k=1}^{r}y_{(i:m)k}.$$

Now, by taking expectation with respect to randomization of original RSS sample selection procedure on s^2_{st} following approach of Patil *et al.* (1995) we get

$$E\left(s^2_{st}\right) = \frac{1}{m^2}\left(\frac{1}{z}-\frac{1}{r}\right)\frac{1}{r-1}\sum_{i=1}^{m}\left(\sum_{k=1}^{r}E\left(y^2_{(i:m)k}\right) - r.E\left(\bar{y}_{(i)}^{\ 2}\right)\right)$$
$$= \frac{1}{m^2}\left(\frac{1}{z}-\frac{1}{r}\right)\frac{1}{r-1}\sum_{i=1}^{m}\left[\sum_{k=1}^{r}\left\{\sigma^2_{(i:m)} + \mu^2_{(i:m)}\right\} - r\left\{\frac{1}{r}\sigma^2_{(i:m)} + \frac{r-1}{r}C_{ii} + \mu^2_{(i:m)}\right\}\right]$$
$$= \frac{1}{m}\left(\frac{1}{z}-\frac{1}{r}\right)\left[\sigma^2_Y - \bar{\gamma}_Y\right]$$

Since, from Patil *et al.* (1995) following results can be seen

$$E\left(y_{(i:m)k}\right) = \mu_{(i:m)},\ V\left(y_{(i:m)k}\right) = \sigma^2_{(i:m)},\ Cov\left(y_{(i:m)k}, y_{(j:m)k'}\right) = C_{ij},$$

$$\sum_{i=1}^{m}\sigma^2_{(i:m)} = m\sigma^2_Y - \sum_{i=1}^{m}\left(\mu_{(i:m)} - \bar{Y}\right)^2 \text{ and } \bar{\gamma}_Y = \frac{1}{m}\left[\sum_{i=1}^{m}\left(\mu_{(i:m)} - \bar{Y}\right)^2 + \sum_{i=1}^{m}C_{ii}\right].$$

Then, by taking expectation of $V_*(\tilde{\bar{y}}_{s*,st})$ using the results of above equation we get

$$E\left(V_*(\tilde{\bar{y}}_{s*,st})\right) = \frac{N-1-mr}{m\,r\,(N-1)}\left[\frac{mr+(N-mr)\bar{\bar{D}}}{mr_1+(N-mr_1)\bar{\bar{D}}}\right]^2\left[\sigma^2_Y - \bar{\gamma}_Y\right]. \tag{2.7}$$

For the second term of Equation 2.6, in case of large sample sizes, by considering only the contributing terms to the approximate form of $V_*\left(\bar{y}^*_{t*,st}\right)$ and ignoring the higher order terms, we get

$$V_*(\tilde{\bar{y}}_{t*,st}) = f_1\,V_*\left(\bar{y}^*_{t*,st}\right) \cong f_1 V_*\left[\frac{1}{m}\sum_{i=1}^{m}\bar{y}^*_{(i)}\right]\bar{\bar{d}}^{\,2} = f_1\,s^2_{st}\,\bar{\bar{d}}^{\,2},$$

where

$$\bar{y}^*_{t^*,st} = \frac{1}{N-mz}\sum_{j\in t^*}\left[\sum_{k=1}^{z}\sum_{i=1}^{m}(y^*_{(i:m)k} / d^*_{(i:m)k,j}) \Big/ \sum_{k=1}^{z}\sum_{i=1}^{m}(1/d^*_{(i:m)k,j})\right],$$

$$\bar{\bar{d}} = \frac{1}{N}\sum_{j\in\Omega}\bar{d}_j,\ \bar{d}_j = \frac{1}{\bar{r}_{2j,RSS}\,\bar{d}_{.j,RSS}},\ \bar{r}_{2j,RSS} = \frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m}\frac{1}{d_{i(i:m)k,j}}$$

and $\bar{d}_{.j,RSS} = \frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m} d_{i(i:m)k,j}$.

Now, by taking expectation with respect to randomization of original sample selection procedure and ignoring higher order terms we get

$$E\left(s^2_{st}\,\bar{\bar{d}}^{\,2}\right) \cong \frac{1}{m}\left(\frac{1}{r_1}-\frac{1}{r}\right)\left[\sigma^2_Y - \bar{\gamma}_Y\right]\bar{\bar{D}}^2.$$

Detailed derivation for significant terms can be seen in the Appendix.

Now, using this result and taking expectation on $V_*(\tilde{\bar{y}}_{t^*,st})$ we get

$$E\left[V_*(\tilde{\bar{y}}_{t^*,st})\right] \cong \frac{N-1-mr}{m\,r(N-1)}\left[\frac{mr+(N-mr)\bar{\bar{D}}}{mz+(N-mz)\bar{\bar{D}}}\right]^2\left[\sigma^2_Y - \bar{\gamma}_Y\right]\bar{\bar{D}}^2. \tag{2.8}$$

Again, for the third term of Equation 2.6, in case of large sample sizes, by considering only the contributing terms to the approximate form of covariance term and ignoring the higher order terms, we can get

$$Cov_*(\tilde{\bar{y}}_{s^*,st},\ \tilde{\bar{y}}_{t^*,st}) = f_1\ Cov_*\left[\bar{y}^*_{s^*,st},\ \bar{y}^*_{t^*,st}\right] \cong f_1\, s^2_{st}\,\bar{\bar{d}}.$$

Now, by taking expectation and ignoring second and higher order terms we get

$$E\left[Cov_*(\tilde{\bar{y}}_{s^*,st},\ \tilde{\bar{y}}_{t^*,st})\right] \cong \frac{N-1-mr}{m\,r(N-1)}\left[\frac{mr+(N-mr)\bar{\bar{D}}}{mz+(N-mz)\bar{\bar{D}}}\right]^2\left[\sigma^2_Y - \bar{\gamma}_Y\right]\bar{\bar{D}}. \tag{2.9}$$

Finally, replacing the results of expectations from Equation (2.7), (2.8) and (2.9) in the Equation (2.6) we get

$$E\left(\hat{V}_{b,st}\right) \cong \left(\frac{mr+(N-mr)\bar{\bar{D}}}{N}\right)^2 \frac{N-1-mr}{mr(N-1)}\left(\sigma^2_Y - \bar{\gamma}_Y\right) \cong V\left(\hat{\bar{Y}}_{SE,RSS}\right). \tag{2.10}$$

Hence, it can be observed that proposed Rescaled Spatial Stratified Bootstrap (RSSB) Method approximately leads to unbiased variance estimation of proposed Spatial Estimator (SE) under RSS.

2.2 Rescaled Spatial Clustered Bootstrap (RSCB) Method

Rescaled Spatial Clustered Bootstrap (RSCB) method based on observations of different cycles in the RSS sample. The RSS sample of size n = mr comprises r cycles containing one observation from each of the m ranks. Let us consider, r cycles as r clusters consisting of m units, which are from each rank within a cycle. The steps involved in RSCB method for estimation of variance of the proposed SE in case of RSS are given below

a) Draw a sample of $z(<r)$ clusters (cycles) by SRSWOR from the observed r clusters in the observed RSS sample and observe all the units of the selected clusters which creates a bootstrap sample as $\{y^*_{(i:m)k}\} \in s^*,\ \forall\ i=1,\dots,m$ and $k=1,\dots,z$.

b) Then work out,

$$\tilde{y}_{(i:m)k} = \bar{y}_{RSS} + f_2^{1/2}\left(y^*_{(i:m)k} - \bar{y}_{RSS}\right)$$

and
$$\tilde{\bar{y}}_{s^*,cl} = \frac{1}{z}\sum_{k=1}^{z}\tilde{\bar{y}}_k = \frac{1}{z}\sum_{k=1}^{z}\frac{1}{m}\sum_{i=1}^{m}\tilde{y}_{(i:m)k}, \tag{2.11}$$

where

$$f_2 = \left[\{z(N-1-mr)\}\big/\{(r-z)(N-1)\}\right]\left[\left\{mr+(N-mr)\bar{\bar{D}}\right\}\big/\left\{mz+(N-mz)\bar{\bar{D}}\right\}\right]^2.$$

c) Predict all the sampling units belongs to t^* using these $\tilde{y}_{(i:m)k}$ and the corresponding distance, $d^*_{(i:m)k,j}$, with j^{th} non-bootstrap sampling unit belonging to t^* as

$$\tilde{y}_{j,p} = \left(\sum_{k=1}^{z}\sum_{i=1}^{m}(\tilde{y}_{(i:m)k} / d^*_{(i:m)k,j})\right)\Big/\left(\sum_{k=1}^{z}\sum_{i=1}^{m}(1/d^*_{(i:m)k,j})\right),$$

$(i:m)k \in s^*$ and $j \in t^*$.

d) Then work out

$$\tilde{\bar{y}}_{t^*,cl} = \frac{1}{N-mz}\sum_{j\in t^*}\tilde{y}_{j,p}.$$

e) Finally calculate

$$\tilde{T}_{cl} = \left[mz\,\bar{y}_{s^*,cl} + (N-mz)\bar{y}_{t^*,cl}\right]\big/N. \tag{2.12}$$

f) Return the resample in the sample and independently replicate step (a) through (e) for a large number, say B, times and compute the corresponding $\tilde{T}^1_{cl}, \tilde{T}^2_{cl}, \dots, \tilde{T}^B_{cl}$.

g) The bootstrap variance estimator of $\tilde{T}_{cl}$ is given by

$$\hat{\hat{V}}_{b,cl} = V_*\left(\tilde{T}_{cl}\right) = E_*\left(\tilde{T}_{cl} - E_*\,\tilde{T}_{cl}\right)^2. \qquad (2.13)$$

The Monte Carlo approximation estimator, $\hat{\hat{V}}_{b,cl}(a)$, as given by

$$\hat{\hat{V}}_{b,cl}(a) = \frac{1}{B-1}\sum_{b=1}^{B}(\tilde{T}_{cl}^{b} - \overline{T}_{cl})^2,$$

where $\overline{T}_{cl} = \frac{1}{B}\sum_{b=1}^{B}\tilde{T}_{cl}^{b}$.

Now, by taking expectation on $\hat{\hat{V}}_{b,cl}$ with respect to original sample selection we can get

$$E\left(\hat{\hat{V}}_{b,cl}\right) = \frac{1}{N^2}\Big[m^2z^2E\left(V_*(\tilde{\bar{y}}_{s*,cl})\right) + (N-mz)^2E\left(V_*(\tilde{\bar{y}}_{t*,cl})\right) + 2mz(N-mz)E\left(Cov_*(\tilde{\bar{y}}_{s*,cl}, \tilde{\bar{y}}_{t*,cl})\right)\Big]. \qquad (2.14)$$

Since, in this case resamples of z clusters (cycles) are drawn by SRSWOR from r clusters, then using Equation (2.11) we get

$$V_*(\tilde{\bar{y}}_{s*,cl}) = f_2\, V_*\left[\frac{1}{z}\sum_{k=1}^{z}\bar{y}_k^*\right] + \left(1-f_2^{1/2}\right)^2 V_*\left(\,\bar{y}_{RSS}\right) = f_2\, s_{cl}^2,$$

where $s_{cl}^2 = \left(\frac{1}{z}-\frac{1}{r}\right)\frac{1}{r-1}\sum_{k=1}^{r}\left(\bar{y}_k - \bar{y}_{RSS}\right)^2$ and

$$\bar{y}_k^* = \frac{1}{m}\sum_{i=1}^{m}y_{(i:m)k}^* = \bar{y}_k.$$

By taking expectation with respect to randomization of original RSS sample selection procedure on s_{cl}^2 following Patil *et al.* (1995) we get

$$E\left(s_{cl}^2\right) = \left(\frac{1}{z}-\frac{1}{r}\right)\frac{1}{r-1}\left[\sum_{k=1}^{r}\left\{V(\bar{y}_k) + \left(E(\bar{y}_k)\right)^2\right\} - r\left\{V(\bar{y}_{RSS}) + \left(E(\bar{y}_{RSS})\right)^2\right\}\right]$$

$$= \frac{1}{m}\left(\frac{1}{z}-\frac{1}{r}\right)\left[\,\sigma_Y^2 - \bar{\gamma}_Y\right].$$

Therefore, by taking expectation of $V_*(\tilde{\bar{y}}_{s*,cl})$ using above equation we get

$$E\left(V_*(\tilde{\bar{y}}_{s*,cl})\right) = \frac{N-1-mr}{m\,r\,(N-1)}\left[\frac{mr+(N-mr)\bar{\bar{D}}}{mr_1+(N-mr_1)\bar{\bar{D}}}\right]^2\left[\,\sigma_Y^2 - \bar{\gamma}_Y\right]. \qquad (2.15)$$

For the second term of Equation 2.14, in case of large sample sizes, by considering only the contributing terms to the approximate form of $V_*\left(\bar{y}_{t*,cl}^*\right)$ and ignoring the higher order terms, we can get

$$V_*(\tilde{\bar{y}}_{t*,cl}) = f_2\, V_*\left(\bar{y}_{t*,cl}^*\right) \cong f_2\, V_*\left[\frac{1}{z}\sum_{k=1}^{z}\bar{y}_k^*\right]\bar{\bar{d}}^2 = f_2\, s_{cl}^2\,\bar{\bar{d}}^2$$

where,

$$\bar{y}_{t*,cl}^* = \frac{1}{N-mz}\sum_{j\in t^*}\left[\sum_{k=1}^{z}\sum_{i=1}^{m}(y_{(i:m)k}^*/d_{(i:m)k,j}^*)\Big/\sum_{k=1}^{z}\sum_{i=1}^{m}(1/d_{(i:m)k,j}^*)\right].$$

Now, by taking expectation with respect to the randomization at original sample selection stage and after ignoring higher order terms we get

$$E\left[V_*(\tilde{\bar{y}}_{t*,cl})\right] \cong \frac{N-1-mr}{m\,r\,(N-1)}\left[\frac{mr+(N-mr)\bar{\bar{D}}}{mr_1+(N-mr_1)\bar{\bar{D}}}\right]^2\left[\,\sigma_Y^2 - \bar{\gamma}_Y\right]\bar{\bar{D}}^2 \qquad (2.16)$$

Once again, for the third term of Equation 2.14, in case of large sample sizes, by considering only the contributing terms to the approximate form of covariance term and ignoring the higher order terms, we can write

$$Cov_*(\tilde{\bar{y}}_{s*,cl},\, \tilde{\bar{y}}_{t*,cl}) = f_2\, Cov_*\left[\bar{y}_{s*,cl}^*,\ \bar{y}_{t*,cl}^*\right] \cong f_2\, s_{cl}^2\,\bar{\bar{d}}.$$

Ignoring higher order terms while taking expectation we get

$$E\left[Cov_*(\tilde{\bar{y}}_{s*,cl},\, \tilde{\bar{y}}_{t*,cl})\right] \cong \frac{N-1-mr}{m\,r\,(N-1)}\left[\frac{mr+(N-mr)\bar{\bar{D}}}{mz+(N-mz)\bar{\bar{D}}}\right]^2\left[\,\sigma_Y^2 - \bar{\gamma}_Y\right]\bar{\bar{D}}. \qquad (2.17)$$

Finally, by substituting the results of Equation (2.15), (2.16) and (2.17) in the Equation (2.14)

$$E\left(\hat{\hat{V}}_{b,cl}\right) \cong \left(\frac{mr+(N-mr)\bar{\bar{D}}}{N}\right)^2\frac{N-1-mr}{mr(N-1)}\left(\sigma_Y^2 - \bar{\gamma}_Y\right)$$

$$\cong V\left(\hat{\bar{Y}}_{SE,RSS}\right). \qquad (2.18)$$

Hence, it can be seen that the proposed Rescaled Spatial Clustered Bootstrap (RSCB) method is useful for approximately unbiased variance estimation of the proposed SE in case of RSS.

3. SIMULATION STUDY

A simulation study was carried out to study statistical properties of both the proposed Rescaled Spatial Bootstrap (RSB) methods for variance estimation of the Spatial Estimator of population mean (Biswas *et al.*, 2015) under RSS were studied, considering underlying population as finite and spatial in nature. While considering spatially correlated populations, Rai *et al.* (2007) fitted observed wheat yield distribution of 1997-98 in different fields of Rohtak district of Haryana state and found exponential variogram model is the best-fitted model. Therefore in this study, first, a univariate population of 900 areal units was generated using exponential variogram model using 'SIM2D' procedure in SAS in the form of regular grid of size as 1x1 unit^2 in the similar line of Biswas *et al.* (2015). In order to generate this spatial population, value of Moran's spatial correlation coefficient (β) as around 0.7 and value of per cent Coefficient of Variation (%CV) as 20% have been considered. Keeping this in view, the parameters of the generated population were taken as

Parameter	Mean	Scale/Sill	Range	Nugget effect	Angle
Value	30	36.2881	30.6239	0.88	135°

Further, another auxiliary variate (X) which is highly correlated with study variable Y, was generated for ranking purpose of the sampling units for selection of RSS samples with following parameter values of mean of X (μ_X), standard deviation of X (σ_X) and correlation coefficient between X and Y (ρ_{XY}) as

Parameter	μ_X	σ_X	ρ_{XY}
Value	45	8	0.7

Next, in order to study distributional properties of the Spatial Estimator of population mean (Biswas *et al.*, 2015) in case of RSS, 500 samples of different sample sizes, using different combination of number of ranks (m) and number of cycles (r), were selected from this generated population using RSS. After that, 200 bootstrap samples were drawn from each of the selected 500 RSS samples following both the proposed Rescaled Spatial Bootstrap (RSB) methods and estimates of variance of the Spatial Estimator of population mean under RSS were obtained. The Monte Carlo bootstrap estimates of population mean and estimate of variance of SE under RSS were calculated. In order to compare the performance of these proposed RSB methods for variance estimation of SE under RSS, percentage Relative Bias (%RB) and Relative Stability (RS) were calculated using following formulae

$$\%RB = \left[\frac{\frac{1}{s}\sum_s \left\{ \hat{V}_s\left(\hat{\hat{Y}}_{SE,RSS}\right) \right\} - V\left(\hat{\hat{Y}}_{SE,RSS}\right)}{V\left(\hat{\hat{Y}}_{SE,RSS}\right)} \right] \times 100$$

$$\text{and } RS = \frac{\left[\frac{1}{s}\sum_s \left\{ \hat{V}_s\left(\hat{\hat{Y}}_{SE,RSS}\right) - V\left(\hat{\hat{Y}}_{SE,RSS}\right) \right\}^2 \right]^{1/2}}{V\left(\hat{\hat{Y}}_{SE,RSS}\right)}$$

where, $\hat{V}_s\left(\hat{\hat{Y}}_{SE,RSS}\right)$ is the estimates of variance of the proposed SE under RSS at sth sample.

4. RESULTS AND DISCUSSIONS

Biswas *et al.* (2015) shown that the Spatial Estimator (SE) of population mean under RSS scheme is superior with respect to the usual RSS. While comparing with usual SRSWOR estimator, the SE in case of RSS has always given significant amount of gain in efficiency. Results of the simulation study comprising the statistical performance of both the proposed Rescaled Spatial Bootstrap (RSB) methods for variance estimation of the Spatial Estimator of population mean under RSS are presented here. Table 1 and 2 show different statistical properties of the proposed variance estimation methods developed in case of the proposed SE in case of RSS viz. Rescaled Spatial Stratified Bootstrap (RSSB) and Rescaled Spatial Clustered Bootstrap (RSCB) methods respectively, such as Monte Carlo (MC) bootstrap estimate, estimates of variance, percentage Relative Bias (%RB) and Relative Stability (RS) of the estimates of variance for different sample sizes (n = mr) and corresponding bootstrap sample sizes (mz) combinations.

Following points can be noted from Table 1 and 2.

- MC mean in both the proposed bootstrap procedures are close to population mean i.e. 30. This indicates that both bootstrap estimators of population mean are almost unbiased.
- Both the RSSB and RSCB methods result in lower percentage relative bias for estimation of variance of the SE of population mean in case of RSS at different combination of sample and bootstrap sample sizes. It was observed that standard bootstrap procedures i.e. without using the proposed rescaling factors in case of both methods give high amount of percentage relative bias. Thus, proposed rescaling factors were quite effective in reducing

Table 1. Monte Carlo estimates, estimate of variances, percentage Relative Biases (%RB) and Relative Stabilities (RS) of the estimates of variance using Rescaled Spatial Stratified Bootstrap (RSSB) method for different sample size (mr) and bootstrap sample size (mz) combinations

m	r	z	Rescaled Spatial Stratified Bootstrap (RSSB) method			
			MC Mean	Estimate of Variance	%RB	RS
2	30	12	30.36	0.42	3.84	0.21
2	60	24	30.32	0.19	-8.10	0.16
2	90	36	30.38	0.12	7.89	0.18
3	20	8	30.33	0.37	7.21	0.22
3	40	15	30.36	0.16	6.99	0.21
3	60	25	30.37	0.11	2.56	0.15
4	15	6	30.34	0.34	2.98	0.27
4	30	12	30.34	0.16	4.01	0.15
4	45	20	30.34	0.10	5.94	0.19
5	12	5	30.39	0.37	6.63	0.24
5	24	10	30.40	0.14	9.48	0.24
5	36	15	30.38	0.08	8.48	0.21

Table 2. Monte Carlo estimates, estimate of variances, percentage Relative Biases (%RB) and Relative Stabilities (RS) of the estimates of variance using Rescaled Spatial Clustered Bootstrap (RSCB) method for different sample size (mr) and bootstrap sample size (mz) combinations

m	r	z	Rescaled Spatial Clustered Bootstrap (RSCB) method			
			MC Mean	Estimate of Variance	%RB	RS
2	30	12	30.36	0.42	3.92	0.27
2	60	24	30.32	0.19	-6.84	0.18
2	90	36	30.38	0.12	6.34	0.19
3	20	8	30.33	0.37	8.67	0.35
3	40	15	30.36	0.16	6.50	0.27
3	60	25	30.36	0.11	2.28	0.21
4	15	6	30.33	0.34	1.94	0.40
4	30	12	30.33	0.17	8.02	0.30
4	45	20	30.34	0.10	5.46	0.24
5	12	5	30.39	0.33	6.98	0.41
5	24	10	30.40	0.14	7.32	0.35
5	36	15	30.38	0.08	8.13	0.29

%RB considerably as compared to usual RSSB and RSCB methods for variance estimation of the Spatial Estimator of population mean under RSS. Therefore, both the proposed variance estimation methods are almost unbiased as established theoretically and by simulation results.

- Further, while comparing through Relative Stability (RS), it can be seen that RSSB method always provide less amount of RS as compared to RSCB with varying sample and bootstrap sample sizes. Therefore, RSSB method is more stable than RSCB method.
- Therefore, RSSB method can be considered better than RSCB method for variance estimation of the proposed SE in case of RSS.

5. CONCLUSIONS

Due to complexities in estimation of variance of the SE under RSS, in this article, two different Rescaled Spatial Bootstrap (RSB) methods viz. Rescaled Spatial Stratified Bootstrap (RSSB) based on ranks and Rescaled Spatial Clustered Bootstrap (RSCB) methods based on cycles of RSS sample, were developed. Both the methods provide approximately unbiased estimation of the proposed estimator theoretically. Simulation study reveals that both the proposed methods give approximately unbiased estimation of variance of the SE under RSS for different combination of sample and bootstrap sample sizes, but while considering relative stability, RSSB method was found to be more stable. Therefore, RSSB method can be considered better than RSCB method for variance estimation of the SE under RSS in the context of spatially correlated finite populations. However, as far as simplicity of both the methods is concerned, RSCB method can be considered for unbiased variance estimation of the SE under RSS.

REFERENCES

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

Al-Saleh, M.F. and Samawi, H.M. (2007). A note on inclusion probability in ranked set sampling and some of its variations. *Test*, **16**, 198-209.

Al-Saleh, M.F. and Zheng, G. (2002). Estimation of bivariate characteristics using ranked set sampling. *Australian & New Zealand Journal of Statistics*, **44**, 221-232.

Arbia, G. (1993). The use of GIS in spatial statistical surveys. *Int. Stat. Rev.* **61(2)**, 339-359.

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., Ahmad, T. and Sahoo, P.M. (2015). Spatial estimation approach under ranked set sampling from spatial correlated finite population. *International Journal of Agricultural and Statistical Sciences*, **11(2)**, 551-558.

Biswas, A., Rai, A. and Ahmad, T. (2018). Rescaling Bootstrap Technique for Variance Estimation for Ranked Set Samples in

Finite Population. *Communications in Statistics - Simulation and Computation.* DOI: 10.1080/03610918.2018.1527349.

Bocci, C., Petrucci, A. and Rocco, E. (2010). Ranked set sampling allocation models for multiple skewed variables: An application to agricultural data. *Environmental and Ecological Statistics*, **17(3)**, 333-345.

Chen, Z., Bai, Z. and Sinha, B. (2004). *Ranked Set Sampling: Theory and Applications*. Springer, New York.

Cobby, J.M., Ridout, M.S., Bassett, P.J., and Large, R.V. (1985). An investigation into the use of ranked set sampling on grass and grass-clover swards. *Grass and Forage Science*, **40**, 257-263.

Dell, J.R. and Clutter, J.L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, **28**, 545-553.

Deshpande, J.V., Frey, J. and Ozturk, O. (2006). Nonparametric ranked-set sampling confidence intervals for a finite population. *Environmental and Ecological Statistics*, **13**, 25-40.

Donald, S. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 Association for Computing Machinery (ACM) National Conference*, pp. 517-524.

Dong, X., Cui, L. and Liu, F. (2012). A further study on reliable life estimation under ranked set sampling. *Communications in Statistics: Theory and Methods*, **41(21)**, 3888-3902.

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

Frey, J. (2011). Recursive computation of inclusion probabilities in ranked set sampling. *Journal of Statistical Planning and Inference*, **141**, 3632-3639.

Ghitany, M.E. (2005). On reliability estimation based on ranked set sampling. *Communications in Statistics: Theory and Methods*, **34(5)**, 1213-1216.

Gökpinar, F., and Özdemir, Y.A. (2010). Generalization of inclusion probabilities in ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, **39(1)**, 89-95.

Gökpinar, F., and Özdemir, Y.A. (2012). A Horvitz-Thompson estimator of the population mean using inclusion probabilities of ranked set sampling. *Communications in Statistics - Theory and Methods*, **41(6)**, 1029-1039.

Gökpinar, F., and Özdemir, Y.A. (2014). Simple computational formulas for inclusion probabilities in ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, **43(1)**, 117-130.

Halls, L.S. and Dell, T.R. (1985). Trial of ranked set sampling for forage yields. *Forest Science*, **12**, 22-26.

Hedayat, A.S., Rao, C.R. and Stufken, J. (1988). Sampling plans excluding contiguous units. *J. Statist. Plann. Inf.* **19**, 159-170.

Hui, T., Modarres, R. and Zheng, G. (2005). Bootstrap confidence interval estimation of mean via ranked set sampling linear regression. *Journal of Statistical Computation and Simulation*, **75**, 543-553.

Husby, C.E., Stasny, E.A. and Wolfe, D.A. (2005). An application of ranked set sampling for mean and median estimation using USDA crop production data. *Journal of Agricultural Biological & Environmental Statistics*, **10(3)**, 354-373.

Jafari Jozani, M. and Johnson, B.C. (2011). Design based estimation for ranked set sampling in finite population. *Environmental and Ecological Statistics*, **18**, 663-685.

Jafari Jozani, M. and Johnson, B.C. (2012). Randomized nomination sampling in finite populations. *Journal of Statistical Planning and Inference*, **142**, 2103-2115.

Kankure, A.K. and Rai, A. (2008). Spatial ranked set sampling from spatially correlated population. *J. Ind. Soc. Agril. Statist.*, **62(3)**, 221-230.

Krishna, P. (2002). *Some aspects of ranked set sampling from finite population*. M.Sc. Thesis, Indian Agricultural Research Institute, New Delhi-110012.

Kvam, P.H. and Samaniego, F.J. (1994). Nonparametric maximum likelihood estimation based on ranked set samples. *J. Amer. Statist. Assoc.*, **89**, 526-537.

Martin, W., Sharik, T., Oderwald, R. and Smith, D. (1980). *Evaluation of ranked set sampling for estimating shrub phytomass in Appalachian oak forests*. School of Forestry and Wild life Resources (Blacksburg, VA: Virginia Polytechnic Institute and State University).

McIntyre, G.A. (1952). A method of unbiased selective sampling using ranked sets. *Aust. J. Agril. Res.*, **3**, 385-390.

Modarres, R., Hui, T.P. and Zheng, G. (2006). Resampling methods for ranked set samples. *Computational Statistics & Data Analysis*, **51**, 1039-1050.

Nahhas, R.W., Wolfe, D.A. and Chen, H. (2002). Ranked set sampling: Cost and optimal set size. *Biometrics*, 58, 964-971.

Özdemir, Y.A., and Gökpinar, F. (2007). A generalized formula for inclusion probabilities in ranked set sampling. *Hacettepe Journal of Mathematics and Statistics*, **36**, 89-99.

Özdemir, Y.A., and Gökpinar, F. (2008). A new formula for inclusion probabilities in median ranked set sampling. *Communications in Statistics - Theory and Methods*, **37(13)**, 2022-2033.

Ozturk, O. (2014). Estimation of population mean and total in finite population setting using multiple auxiliary variables. *Journal of Agricultural, Biological and Environmental Statistics*, **19**, 161-184.

Ozturk, O. (2016a). Estimation of a finite population mean and total using population ranks of sample units. *Journal of Agricultural, Biological and Environmental Statistics,* **21(1)**, 181-202.

Ozturk, O. (2016b). Statistical inference based on judgment post-stratified samples in finite population. *Survey Methodology*, **42(2)**, 239-262.

Ozturk, O., and Jafari Jozani, M. (2013). Inclusion probabilities in partially rank ordered set sampling. *Computational Statistics and Data Analysis*, **69**, 122-132.

Ozturk, O., Bilgin, O.C. and Wolfe, D.A. (2005). Estimation of population mean and variance in flock management: A ranked set sampling approach in a finite population setting. *Journal of Statistical Computation and Simulation*, **75(11)**, 905-919.

Patil, G.P., Sinha, A.K. and Taillie, C. (1994). Ranked set sampling. Handbook of Statistics, **12**, (eds. G.P. Patil, and C.R. Rao), pp. 167-198, North-Holland, Amsterdam.

Patil, G.P., Sinha, A.K. and Taillie, C. (1995). Finite population corrections for ranked set sampling. *Annals of Institute of Statistical Mathematics*, **47**, 621-636.

Rai, A. and Krishna, P. (2013). Ranked set sampling from finite population under randomization framework. *J. Ind. Soc. Agril. Statist.*, **67(3)**, 363-369.

Rai, A., Gupta, N. K. and Singh, R. (2007). Small area estimation of crop production using spatial models. *Model Assist. Stat. Appl.* **2(2)**, 89-98.

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

Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika.* **57**, 377-387.

Sahoo, P. M., Singh, R., Rai, A. (2006). Spatial sampling procedure for agricultural surveys using geographical information system. *J. Ind. Soc. Agricult. Statist.* **60(2)**, 134-143.

Samawi, H.M. and Al-Sagheer, O.A.M. (2001). On the estimation of distribution function using extreme and median ranked set sampling. *Biometrical Journal*, **43**, 357-373.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer, New York.

Sud, U.C. and Mishra, D.C. (2006). Estimation of finite population mean using ranked set two stage sampling designs. *J. Ind. Soc. Agricult. Statist.* **60(2)**, 108-117.

Sud, U.C. and Mishra, D.C. (2007). Estimation of finite population mean using double ranked set sampling. *J. Ind. Soc. Agricult. Statist.* **61(3)**, 389-399.

Takahasi, K. and Futatsuya, M. (1988). Ranked set sampling from a finite population. *Proceedings of the Institute of Statistical Mathematics*, **36**, 55-68.

Takahasi, K. and Futatsuya, M. (1998). Dependence between order statistics in samples from finite population and its application to ranked set sampling. *Annals of the Institute of Statistical Mathematics,* **50(1)**, 49-70.

Wolfe, D.A. (2012). Ranked set sampling: Its relevance and impact on statistical inference. *ISRN Probability and Statistics*, DOI:10.5402/2012/568385.

Wolter, K. M. (1985). *Introduction to Variance estimation*, Springer, New York.

Zhang, Z. and Griffith, D.A. (2000). Integrating GIS components and spatial statistical analysis in DBMSs. *Int. Jour. G.I.S.* **14(6)**, 543-566.

APPENDIX

In Section 2, it can be seen that at the time of predicting all the non-bootstrap sample units from the population using the bootstrap sample units, there will be $\bar{\bar{d}}$ in the approximate variance expression which is given by

$$\bar{\bar{d}} = \frac{1}{N}\sum_{j\in\Omega} \bar{d}_j, \tag{A1}$$

where

$$\bar{d}_j = \frac{1}{\bar{r}_{2j,RSS}\,\bar{d}_{.j,RSS}},\quad \bar{r}_{2j,RSS} = \frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m}\frac{1}{d_{i(i:m)k,j}} \text{ and}$$

$$\bar{d}_{.j,RSS} = \frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m} d_{i(i:m)k,j}.$$

From Biswas *et al.* (2015), we get $E_s\left[\bar{d}_{.j,RSS}\right] = \bar{D}_{.j}$ and $E_s\left(\bar{r}_{2j,RSS}\right) = \bar{R}_{2j}$ respectively, E_s denotes expectation due randomization at sampling stage over all sample units belonging to set s.

Now, let

$$\bar{r}_{2j,RSS} = \bar{R}_{2j} + \varepsilon_1; \qquad \text{Such that } E_s\left(\varepsilon_1\right) = 0$$

$$\bar{d}_{.j,RSS} = \bar{D}_{.j} + \varepsilon_2; \qquad \text{Such that } E_s\left(\varepsilon_2\right) = 0$$

Then, $\bar{d}_j$ can be rewritten as

$$\bar{d}_j = \frac{1}{\bar{R}_{2j}\bar{D}_{.j}}\left(1+\frac{\varepsilon_1}{\bar{R}_{2j}}\right)^{-1}\left(1+\frac{\varepsilon_2}{\bar{D}_{.j}}\right)^{-1}$$

$$= \frac{1}{\bar{R}_{2j}\bar{D}_{.j}}\left[1-\frac{\varepsilon_1}{\bar{R}_{2j}}-\frac{\varepsilon_2}{\bar{D}_{.j}}+\frac{\varepsilon_1\varepsilon_2}{\bar{R}_{2j}\bar{D}_{.j}}+\ldots\right]. \tag{A2}$$

Then, by taking the expectation of sampling stage on $\bar{\bar{d}}$ and retaining first order terms we get

$$E\left(\bar{\bar{d}}\right) = E_{\bar{s}}\,E_s\left(\bar{\bar{d}}\right) = E_{\bar{s}}\left(\frac{1}{N}\sum_{j=1}^{N}E_s\left(\bar{d}_j\right)\right) \cong$$

$$E_{\bar{s}}\left(\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\bar{R}_{2j}\bar{D}_{.j}}\right) = \frac{1}{N}\sum_{j=1}^{N}\bar{D}'_j = \bar{\bar{D}}. \tag{A3}$$

Next, the variance is given by

$$V\left(\bar{\bar{d}}\right) = V\left(\frac{1}{N}\sum_{j\in\Omega}\bar{d}_j\right) = \frac{1}{N^2}\left[\sum_{j\in\Omega}V\left(\bar{d}_j\right)+\sum\sum_{j\neq k\in\Omega} Cov\left(\bar{d}_j,\bar{d}_k\right)\right].$$

Now, proceeding with the variance of $\bar{d}_j$ we get

$$V\left(\bar{d}_j\right) = V_{\bar{s}}E_s\left(\bar{d}_j\right) + E_{\bar{s}}V_s\left(\bar{d}_j\right). \tag{A4}$$

Now, first term can be expanded as

$$V_{\bar{s}}E_s\left(\bar{d}_j\right) = V_{\bar{s}}\left(\bar{D}'_j\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\bar{D}'_j - \bar{\bar{D}}\right)^2 = \sigma_D^2.$$

Now using the results of Equation(B2) we get

$$V_s\left(\bar{d}_j\right) = E_s\left(\bar{d}_j - \bar{D}'_j\right)^2 \cong \bar{D}'^2_j\left[\frac{E_s\left(\varepsilon_1^2\right)}{\bar{R}_{2j}^2}+\frac{E_s\left(\varepsilon_2^2\right)}{\bar{D}_{.j}^2}+\frac{2E_s\left(\varepsilon_1\varepsilon_2\right)}{\bar{R}_{2j}\bar{D}_{.j}}+\ldots\right]. \tag{A5}$$

Then

$$E_s\left(\varepsilon_1^2\right) = V_s\left[\frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m}\left\{\left(\frac{1-\bar{R}_{2j}.d_{i(i:m)k,j}}{\bar{D}_{.j}}\right)\left(1+\frac{d_{i(i:m)k,j}-\bar{D}_{.j}}{\bar{D}_{.j}}\right)^{-1}\right\}\right]$$

$$\cong \frac{1}{\bar{D}_{.j}^{2}} V_s\left(1-\bar{R}_{2j}\bar{d}_{.j,RSS}\right)$$

$$= \frac{\bar{R}_{2j}^{2}}{\bar{D}_{.j}^{2}} V_s\left(\bar{d}_{.j,RSS}\right)$$

Proceeding in similar way we get

$$E_s\left(\varepsilon_2^2\right) = V_s\left(\bar{d}_{.j,RSS}\right)$$

$$E_s\left(\varepsilon_1\varepsilon_2\right) \cong \frac{1}{\bar{D}_{.j}} \operatorname{cov}_2\left(1-\bar{R}_{2j}\bar{d}_{.j,RSS}\,,\ \bar{d}_{.j,RSS}\right) = -\frac{\bar{R}_{2j}}{\bar{D}_{.j}} V_2\left(\bar{d}_{.j,RSS}\right)$$

Finally putting these results in Equation (A5) we get

$$V_s\left(\bar{d}_j\right) \cong \bar{D}_j'^2\left[\frac{1}{\bar{D}_{.j}^2} V_s\left(\bar{d}_{.j,RSS}\right) + \frac{1}{\bar{D}_{.j}^2} V_s\left(\bar{d}_{.j,RSS}\right) - 2\frac{1}{\bar{D}_{.j}^2} V_s\left(\bar{d}_{.j,RSS}\right)\right] = 0.$$

Accordingly from the Equation(A4) we get

$$V\left(\bar{d}_j\right) = \sigma_D^2\,. \tag{A6}$$

The covariance term is given by

$$\operatorname{Cov}\left(\bar{d}_j,\bar{d}_k\right) = \operatorname{Cov}_{\bar{s}} E_s\left(\bar{d}_j,\bar{d}_k\right) + E_{\bar{s}} \operatorname{Cov}_s\left(\bar{d}_j,\bar{d}_k\right). \tag{A7}$$

Now

$$\operatorname{Cov}_{\bar{s}} E_s\left(\bar{d}_j,\bar{d}_k\right) = \operatorname{Cov}_{\bar{s}}\left(\bar{D}'_j,\bar{D}'_k\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\bar{D}'_j - \bar{\bar{D}}\right)\left(\bar{D}'_k - \bar{\bar{D}}\right) = -\frac{\sigma_D^2}{N-1}.$$

Proceeding in similar way as in variance term for the second term in Equation (A7) we get

$$E_{\bar{s}} \operatorname{Cov}_s\left(\bar{d}_j,\bar{d}_k\right) = 0\,.$$

Then putting these results in Equation (A7) we get

$$\operatorname{Cov}\left(\bar{d}_j,\bar{d}_k\right) = -\frac{\sigma_D^2}{N-1}\,. \tag{A8}$$

Therefore the variance of $\bar{\bar{d}}$ is given by

$$V\left(\bar{\bar{d}}\right) = \frac{1}{N^2}\left[\sum_{j=1}^{N} V\left(\bar{d}_j\right) + \sum\sum_{j\neq k}^{N} \operatorname{Cov}\left(\bar{d}_j,\bar{d}_k\right)\right]$$

$$= \frac{1}{N^2}\left[N\sigma_D^2 + N(N-1)\left(-\frac{\sigma_D^2}{N-1}\right)\right]$$

$$= 0 \tag{A9}$$