Int. J. Agricult. Stat. Sci. Vol. 11, No. 2, pp. 551-558, 2015 ***ISSN : 0973-1903***

ORIGINAL ARTICLE

SPATIAL ESTIMATION APPROACH UNDER RANKED SET SAMPLING FROM SPATIAL CORRELATED FINITE POPULATION

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

Indian Agricultural Statistics Research Institute (IASRI), Library Avenue, Pusa, New Delhi - 110 012, India.
E-mail: ankur@iasri.res.in

Abstract : 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. While working with spatial population, classical statistical methods fail to capture the dependency present in the underlying data. In this article, an attempt was made to develop efficient estimation procedure through RSS sampling design incorporating spatial dependency among sampling units of a spatial finite population. Distance between spatial units was taken as measure of spatial dependency. The properties of the proposed Spatial Estimator (SE) were further studied empirically through a simulation study. The proposed Spatial Estimator (SE) under RSS of population mean from spatial data was found to be better than usual RSS estimator.

Key words : Ranked set sampling, Prediction approach, Inverse distance weighting, Spatial estimator.

1. Introduction

In agricultural and environmental surveys often the parameters of interest are spatial in nature, in which neighbouring sampling units are likely to have similar attribute values as compared to distant units. This implies spatial dependency in the underlying data. Classical statistical methods when applied to spatial population structure fail to capture the dependency present in the underlying data [Zhang and Griffith (2000)]. Several authors [Hedayat *et al.* (1988), Arbia (1993), Sahoo *et al*. (2006), Kankure and Rai (2008)] have considered these types 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. Spatial estimator based on the spatial dependency of the sampling units is likely to improve upon the usual estimator of target parameter by traditional aerial sampling schemes. Prediction approach [Royall (1970)] is one of useful techniques in which spatial dependency of the finite population can be incorporated to predict unobserved population units based on their distances with observed sampling units. One way to predict the unobserved unknown population units is through Inverse Distance Weighting (IDW) method as given by Donald (1968).

Ranked Set Sampling (RSS), a well-known sampling scheme for estimation of population parameter of interest, first introduced by McIntyre (1952) and used to estimate mean pasture yield. It 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. Patil *et al*. (1995), Sud and Mishra (2006, 2007), Kankure and Rai (2008), Rai and Krishna (2013) tried to extend the theory of RSS without replacement in the context of finite population. In order to select a sample through RSS design, m^2 units are randomly drawn by SRSWOR from a finite population of size N with mean $\overline{Y}$ and finite variance σ_Y^2 and randomly divided into m sets of size m. Then, the units within each set are ranked on the basis of some ancillary 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 *m*th set. The entire cycle is replicated r times until in total n = mr observations are quantified. These n quantified units constitute a ranked set sample. Patil *et al*. (1995)

****Author for correspondence** **Received** March 11, 2015 **Revised** July 09, 2015 **Accepted** October 18, 2015*

showed that ranked set estimator of the population mean $\overline{Y}$ given by following expression

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

is an unbiased estimator of the population mean $\overline{Y}$ and the variance of the RSS estimator $\overline{y}_{RSS}$ is given by

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

Where, $\sigma_Y^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \overline{Y})^2$, $\overline{\gamma}_Y = \frac{m!(m-1)!}{N(N-1)...(N-2m+1)}\gamma_Y$,

$$\gamma_Y = (\underline{Y} - \overline{Y})' \underset{\sim}{\Gamma} (\underline{Y} - \overline{Y})$$

$\underline{Y} = [Y_1, Y_2, ..., Y_N]$ and $\underset{\sim}{\Gamma} = \binom{N}{m,m}\sum_{i=1}^{m} B_{ii}$ is a symmetric matrix with zeroes on the diagonal and a function of N and m only and it does not depend on population values $\underline{Y}$.

In this study, we have considered RSS randomization framework to select samples from a spatial finite population and a corresponding Spatial Estimator (SE) has been proposed for estimation of population mean using IDW method through prediction approach. This estimator accounts for spatial relationship among the sampling units of the population. Details are discussed in Section 2. Variance expression and corresponding expression for relative precision were developed for the proposed SE in case of RSS. Finally, statistical properties of the proposed SE were studied empirically through a simulation study and discussed in Section 3.

2. Proposed Spatial Estimator (SE) in case of RSS

Let, there be a spatial finite population of size N consisting of units U_i, $i \in \Omega = \{1, 2, ..., N\}$. Suppose, Y be the character of interest and without loss of generality, the respective population values are $Y_1 \le Y_2 \le Y_2 ... \le Y_N$. Let the parameter of interest be linear in nature *i.e.* population mean $\overline{Y} = \frac{1}{N}\sum_{q\in\Omega} Y_q$. Let D_{qj} denotes distance between the population units U_q and U_j, $q, j \in \Omega$. Let, a RSS sample of size n = mr units are drawn from N population units as described in earlier section. Let, s denotes the set of all sampled units. Let, the sample mean be $\overline{y}_{RSS} = \frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m} y_{(i:m)k}$, where $y_{(i:m)k}$ denotes the value of the *i*th ranked unit in the *i*th set of size m in the *k*th cycle of the sample. Let $d_{(i:m)k,j}$ is the distance between the *(i:m)k*th sampled unit and any other non-sampled unit U_j, $j \in \overline{s}$, where $\overline{s}$ is the set of all non-sample units. It may be noted that values of $d_{(i:m)k,j}$'s are random as it depends on selection of *(i:m)k*th sampled unit. Thus, the predicted value of *j*th non-sampled unit, Y_j, $j \in \overline{s}$, following Donald (1968) can be given as

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

$$(i:m)k \in s \text{ and } j \in \overline{s}. \tag{3}$$

Combining these predicted values of (N - mr) non-sampled units of the population, we can get the mean of non-sampled population units as

$$\overline{y}_{p,RSS} = \frac{1}{N-mr}\sum_{j\in\overline{s}} y_{j,p}.$$

Following the prediction approach of Royall (1970), the proposed Spatial Estimator (SE) in case of RSS for estimation of population mean, $\overline{Y}$, is given as

$$\hat{\overline{Y}}_{SE,RSS} = \left[mr\overline{y}_{RSS} + (N-mr)\overline{y}_{p,RSS}\right] \Big/ N. \tag{4}$$

Now, by taking expectation on $\hat{\overline{Y}}_{SE,RSS}$, we get

$$E\left(\hat{\overline{Y}}_{SE,RSS}\right) = E_{\overline{s}}E_s\left(\hat{\overline{Y}}_{SE,RSS}\right) = \frac{mr}{N}\overline{Y} + \frac{N-mr}{N}E_{\overline{s}}\left[\frac{1}{N-mr}\sum_{j\in\overline{s}} E_s\left(y_{j,p}\right)\right] \tag{5}$$

Where, E_s is the expectation over all sampled units $(i:m)k \in s$, whereas $E_{\overline{s}}$ is the expectation over all non-sampled units $j \in \overline{s}$.

Now, from Equation (3), it can be observed that the expression of $y_{j,p}$ is a ratio. Let,

$$y_{j,p} = \overline{z}_{1j} / \overline{z}_{2j} \tag{6}$$

Where, $\overline{z}_{1j} = \frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m}\left(y_{(i:m)k} / d_{(i:m)k,j}\right)$ and

$$\overline{z}_{2j} = \frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m}\left(1/d_{(i:m)k,j}\right).$$

Following Patil *et al*. (1995), it is notable that

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

$$= \frac{1}{mr}\sum_{k=1}^{r}\sum_{i=1}^{m}\left[\sum_{\substack{q\in\Omega\\ q<j}} D_{qj}\frac{\binom{q-1}{i-1}\binom{N-1-q}{m-i}}{\binom{N-1}{m}} + \sum_{\substack{q\in\Omega\\ q>j}} D_{qj}\frac{\binom{q-2}{i-1}\binom{N-q}{m-i}}{\binom{N-1}{m}}\right]$$

$$= \frac{1}{N-1}\sum_{\substack{q\in\Omega\\ q\neq j}} D_{qj} = \bar{D}_{.j}, \text{ say} \qquad (7)$$

since, $D_{qj} = 0$, if $q = j$.

Now, in order to take expectation on $y_{j,p}$, let us consider the following two terms by proceeding in similar fashion of Equation (7).

$$E_s\left(\bar{z}_{1j}\right) = \frac{1}{N-1}\sum_{\substack{q\in\Omega\\ q\neq j}}\frac{Y_q}{D_{qj}} = \bar{R}_{1j} \text{ and}$$

$$E_s\left(\bar{z}_{2j}\right) = \frac{1}{N-1}\sum_{\substack{q\in\Omega\\ q\neq j}}\frac{1}{D_{qj}} = \bar{R}_{2j} \text{ (say)} \qquad (8)$$

Then, by rearranging the terms of $y_{j,p}$ and expanding through Taylor series expansion, we get

$$y_{j,p} = \bar{R}_j\left[1 + \frac{\bar{z}_{1j}-\bar{R}_{1j}}{\bar{R}_{1j}} - \frac{\bar{z}_{2j}-\bar{R}_{2j}}{\bar{R}_{2j}} - \frac{\left(\bar{z}_{1j}-\bar{R}_{1j}\right)\left(\bar{z}_{2j}-\bar{R}_{2j}\right)}{\bar{R}_{1j}\bar{R}_{2j}} + \left(\frac{\bar{z}_{2j}-\bar{R}_{2j}}{\bar{R}_{2j}}\right)^2 \ldots\right] \qquad (9)$$

Where, $\bar{R}_j = \bar{R}_{1j}/\bar{R}_{2j}$.

Then, taking expectation over sampled units (E_s) and retaining first order terms, we get

$$E_s\left(y_{j,p}\right) \cong \bar{R}_j. \qquad (10)$$

Now, using this result in Equation (4) and proceeding in similar fashion as Equation (28) in Appendix A, we get

$$E\left(\hat{\bar{Y}}_{SE,RSS}\right) \cong \frac{mr}{N}\bar{Y} + \frac{N-mr}{N}E_{\bar{s}}\left\{\frac{1}{N-mr}\sum_{j\in\bar{s}} R_j\right\}$$

$$= \frac{mr}{N}\bar{Y} + \frac{N-mr}{N}\bar{\bar{R}}. \qquad (11)$$

Where, $\bar{\bar{R}} = \frac{1}{N}\sum_{j\in\Omega}\bar{R}_j$.

The bias of $\hat{\bar{Y}}_{SE,RSS}$ upto first order can be written as

$$\text{Bias}\left(\hat{\bar{Y}}_{SE,RSS}\right) = E\left(\hat{\bar{Y}}_{SE,RSS}\right) - \bar{Y} \cong \frac{(N-mr)}{N}\left(\bar{\bar{R}} - \bar{Y}\right). \qquad (12)$$

Here, it can be noted that, $\bar{Y}$ and $\bar{\bar{R}}$ are simple average and inverse distance weighted average of all population y-values.

The variance of the estimator, $\hat{\bar{Y}}_{SE,RSS}$, is given by

$$V\left(\hat{\bar{Y}}_{SE,RSS}\right) = \frac{1}{N^2}\left[m^2r^2V\left(\bar{y}_{RSS}\right) + (N-mr)^2V\left(\bar{y}_{p,RSS}\right) + 2mr(N-mr)\text{Cov}\left(\bar{y}_{RSS}, \bar{y}_{p,RSS}\right)\right]. \qquad (13)$$

The second term in (13) can be written as

$$V\left(\bar{y}_{p,RSS}\right) = E_{\bar{s}}V_s\left\{\bar{y}_{p,RSS}\right\} + V_{\bar{s}}E_s\left\{\bar{y}_{p,RSS}\right\} \qquad (14)$$

From Equation (9) the variance of $y_{j,p}$ over sampled units (V_s) upto the second order of approximation can be obtained as

$$V_s\left(y_{j,p}\right) \cong \bar{R}_j\left[\frac{1}{\bar{R}_{1j}^2}V_s\left(\bar{z}_{1j}-\bar{R}_{1j}\right) + \frac{1}{\bar{R}_{2j}^2}V_s\left(\bar{z}_{2j}-\bar{R}_{2j}\right) - \frac{2}{\bar{R}_{1j}\bar{R}_{2j}}\text{Cov}_s\left(\bar{z}_{1j}-\bar{R}_{1j}, \bar{z}_{2j}-\bar{R}_{2j}\right)\right] \qquad (15)$$

Where, V_s and Cov_s are variance and covariance based on sample units.

Now, expanding $\left(\bar{z}_{1j}-\bar{R}_{1j}\right)$ and $\left(\bar{z}_{2j}-\bar{R}_{2j}\right)$ through Taylor series expansion and taking variance and covariance for sampled units, we get

$$V_s\left(\bar{z}_{1j} - \bar{R}_{1j}\right) \cong \frac{1}{\bar{D}_{.j}^2}\left[V_s\left(\bar{y}_{RSS}\right) + \bar{R}_{1j}^2 V_s\left(\bar{d}_{.j,RSS}\right) - 2\bar{R}_{1j}\mathrm{Cov}_s\left(\bar{y}_{RSS}, \bar{d}_{.j,RSS}\right)\right],$$

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

$$\mathrm{Cov}_s\left[\left(\bar{z}_{1j} - \bar{R}_{1j}\right),\left(\bar{z}_{2j} - \bar{R}_{2j}\right)\right] \cong \frac{1}{\bar{D}_{.j}^2}\left[\bar{R}_{1j}\bar{R}_{2j} V_s\left(\bar{d}_{.j,RSS}\right) - \bar{R}_{2j}\mathrm{Cov}_s\left(\bar{y}_{RSS}, \bar{d}_{.j,RSS}\right)\right] \quad (16)$$

Substituting the results of Equation (16) in Equation (15), we get

$$V_s\left(y_{j,p}\right) \cong \frac{1}{\bar{R}_{2j}^2 \bar{D}_{.j}^2} V_s\left(\bar{y}_{RSS}\right) = \frac{1}{mr}\left[\frac{N-1-mr}{N-1}\sigma_Y^2 - \bar{\gamma}_Y\right]\bar{D}_j^{'2} \quad (17)$$

Where, $\bar{D}_j^{'} = \dfrac{1}{\bar{R}_{2j}\bar{D}_{.j}}$.

Proceeding in the same way, following expression of the covariance (Cov_s) upto the second order of approximation can be obtained as

$$\mathrm{Cov}_s(y_{j,p}, y_{k,p}) \cong \bar{D}_j^{'}\bar{D}_k^{'} V_s\left(\bar{y}_{RSS}\right) = \frac{1}{mr}\left[\frac{N-1-mr}{N-1}\sigma_Y^2 - \bar{\gamma}_Y\right]\bar{D}_j^{'}\bar{D}_k^{'}. \quad (18)$$

Now after combining Equation (17) and (18), we get

$$V_s\left(\bar{y}_{p,RSS}\right) \cong \frac{1}{mr}\left[\frac{N-1-mr}{N-1}\sigma_Y^2 - \bar{\gamma}_Y\right]\left(\frac{1}{N-mr}\sum_{j\in\bar{s}}\bar{D}_j^{'}\right)^2. \quad (19)$$

By using Equations (29) and (30) in Appendix A

$$E_{\bar{s}}\left(\frac{1}{N-mr}\sum_{j\in\bar{s}}\bar{D}_j^{'}\right) = \frac{1}{N}\sum_{j\in\Omega}\bar{D}_j^{'} = \bar{\bar{D}} \text{ and}$$

$$V_{\bar{s}}\left(\frac{1}{N-mr}\sum_{j\in\bar{s}}\bar{D}_j^{'}\right) = \frac{mr}{(N-mr)^2}\left(\frac{N-1-mr}{N-1}\sigma_D^2 - \bar{\gamma}_D\right) \quad (20)$$

Where, $\sigma_D^2 = \dfrac{1}{N}\sum_{j\in\Omega}\left(\bar{D}_j^{'} - \bar{\bar{D}}\right)^2$ and $\bar{\gamma}_D$ is defined in analogue to forms of $\bar{\gamma}_Y$ in Equation (2).

Now, by taking expectation over non-sampled units of the population $\left(E_{\bar{s}}\right)$ on the result of Equation (19) utilizing the results of Equation (20), we get

$$E_{\bar{s}}V_s\left(\bar{y}_{p,RSS}\right) = \frac{1}{mr}\left[\frac{N-1-mr}{N-1}\sigma_Y^2 - \bar{\gamma}_Y\right] \times\left[\bar{\bar{D}}^2 + \frac{mr}{(N-mr)^2}\left\{\frac{N-1-mr}{N-1}\sigma_D^2 - \bar{\gamma}_D\right\}\right]. \quad (21)$$

Through, Equation (29) in Appendix A in the second part of Equation (14), we get

$$V_{\bar{s}}E_s\left(\bar{y}_{p,RSS}\right) \cong V_{\bar{s}}\left[\frac{1}{N-mr}\sum_{j\in\bar{s}}\bar{R}_j\right] = \frac{mr}{(N-mr)^2}\left[\frac{N-1-mr}{N-1}\sigma_R^2 - \bar{\gamma}_R\right] \quad (22)$$

Where, $\sigma_R^2 = \dfrac{1}{N}\sum_{j\in\Omega}\left(\bar{R}_j - \bar{\bar{R}}\right)^2$ and $\bar{\gamma}_R$ is defined in analogue to forms of $\bar{\gamma}_Y$ in Equation (2).

Hence, combining Equation (21) and (22) in Equation (13), we get

$$V\left(\bar{y}_{p,RSS}\right) \cong \frac{1}{mr}\left(\frac{N-1-mr}{N-1}\sigma_R^2 - \bar{\gamma}_Y\right) \times\left[\bar{\bar{D}}^2 + \frac{mr}{(N-mr)^2}\left\{\frac{N-1-mr}{N-1}\sigma_D^2 - \bar{\gamma}_D\right\}\right] + \frac{mr}{(N-mr)^2}\left\{\frac{N-1-mr}{N-1}\sigma_R^2 - \bar{\gamma}_R\right\}. \quad (23)$$

Following the similar approach, for the third term in Equation (12), we get

$$\mathrm{Cov}\left(\bar{y}_{RSS}, \bar{y}_{p,RSS}\right) \cong \frac{1}{mr}\left(\frac{N-1-mr}{N-1}\sigma_Y^2 - \bar{\gamma}_Y\right)\bar{\bar{D}}. \quad (24)$$

Finally, by using the Equations (2), (23) and (24) in Equation (13), we get

$$V\left(\hat{\bar{Y}}_{SE,RSS}\right) = \left(\frac{mr + (N-mr)\bar{\bar{D}}}{N}\right)^2 \frac{1}{mr}\left(\frac{N-1-mr}{N-1}\sigma_Y^2 - \bar{\gamma}_Y\right) + \frac{mr}{N^2}\left(\frac{N-1-mr}{N-1}\sigma_R^2 - \bar{\gamma}_R\right) + \frac{1}{N^2}\left(\frac{N-1-mr}{N-1}\sigma_Y^2 - \bar{\gamma}_Y\right)\left(\frac{N-1-mr}{N-1}\sigma_D^2 - \bar{\gamma}_D\right) \quad (25)$$

In case of large population size *i.e.* for large N, we can approximate the variance of proposed Spatial Estimator (SE) under RSS as

$$V\left(\hat{\bar{Y}}_{SE,RSS}\right) \cong \left(\frac{mr + (N - mr)\bar{\bar{D}}}{N}\right)^2 \frac{1}{mr}\left(\frac{N-1-mr}{N-1}\sigma_Y^2 - \bar{\gamma}_Y\right). \quad (26)$$

It can be seen that, $\bar{\bar{D}} = \frac{1}{N}\sum_{j\in\Omega}\frac{1}{\bar{R}_{2j}\bar{D}_{.j}} \le 1$, since $\frac{1}{\bar{R}_{2j}\bar{D}_{.j}} \le 1,\ \forall\ j \in \bar{s}$. Therefore, the proposed SE in case of RSS sampling design for estimation of population mean is more efficient than the usual RSS estimator since, $V\left(\hat{\bar{Y}}_{SE,RSS}\right) \le V\left(\bar{y}_{RSS}\right)$. Relative Precision (RP) of the proposed SE $\hat{\bar{Y}}_{SE,RSS}$ in case of RSS with respect to the usual RSS estimator $\left(\bar{y}_{RSS}\right)$ can be approximately given as

$$RP = \frac{V\left(\bar{y}_{RSS}\right)}{V\left(\hat{\bar{Y}}_{SE,RSS}\right)} \cong \left(\frac{N}{mr + (N - mr)\bar{\bar{D}}}\right)^2. \quad (27)$$

Since, $\bar{\bar{D}} \le 1$, therefore, RP ≥ 1.

Further, statistical properties of the proposed SE in case of RSS are evaluated empirically through a simulation study in the next section.

3. Simulation Study

A simulation study was carried out to study statistical behaviour of the proposed Spatial Estimator (SE) in case of RSS with respect to usual RSS estimator empirically. Also, properties of proposed Rescaled Spatial Bootstrap (RSB) methods for variance estimation of the proposed SE 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 that 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 1 × 1 unit^2.

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 sample with following parameter values of mean of X variate (μ_X), standard deviation of X (σ_X) and correlation coefficient between X and study variate 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 proposed SE 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. Estimates of the proposed SE in case of RSS as well as usual RSS estimator of population mean were obtained from each of the sample. Percentage relative bias of the SE was calculated using following formula as

$$\%\,Bias = \left[\frac{\hat{\bar{Y}}_{SE,RSS} - \bar{Y}}{\bar{Y}}\right] \times 100$$

Where, $\bar{Y}$ and $\hat{\bar{Y}}_{SE,RSS}$ are the population mean and the estimate of population mean based on proposed SE in case of RSS, respectively. Further, variance, % CV, skewness and kurtosis of the proposed SE under RSS were also calculated using the estimates from different samples. Then, percentage gain in efficiency of the proposed SE with respect to usual RSS as well as SRSWOR estimator of population mean for same sample size were obtained using following formulae:

$$GE_a = \left[\frac{V\left(\bar{y}_a\right) - V\left(\hat{\bar{Y}}_{SE,RSS}\right)}{V\left(\hat{\bar{Y}}_{SE,RSS}\right)}\right] \times 100$$

Where, a = RSS or SRS; $V\left(\hat{\bar{Y}}_{SE,RSS}\right)$, $V\left(\bar{y}_{RSS}\right)$

and $V(\bar{y}_{SRS})$ are the approximated variance of the proposed SE, usual RSS and SRSWOR estimator respectively obtained using 500 samples.

SAS codes were written for selection of ranked set samples and empirical evaluation of the proposed SE in case of RSS.

4. Results and Discussion

Different statistical properties of the proposed Spatial Estimator (SE) under RSS for estimation of population mean, namely, variance, %Bias, %CV, skewness, kurtosis and percentage gain in efficiency (GE) of the proposed SE with respect to usual RSS and SRSWOR estimator of population mean for different sample sizes (n) *i.e.* 60, 120 and 180, with different combination of number of ranks (m) and cycles (r) were calculated and presented in Table 1.

Following points can be observed from Table 1.

It can be seen that, although RSS results in unbiased estimator of population mean, the proposed SE under RSS gives negligible negative bias. This bias shows decreasing trend with the increase in sample size (n).

Expected value of % CV decreases with the increase in sample size (n). Thus, the proposed SE under RSS is more stable with the increase in sample size. Furthermore, with the increase in set size (m) with a fixed sample size (n) also increases the stability of the SE.

As in case of usual RSS estimator of population mean, with increase in sample size (n) as well as increase in set size (m) keeping sample size (n) fixed, the variance of the proposed SE under RSS always decreases. This ensures the consistency of the proposed SE under RSS of the population mean.

The value of skewness and kurtosis of the proposed SE under RSS is near to zero. Therefore, it can be concluded that the distribution of the proposed SE is symmetric and mesokurtic in nature.

In terms of gain in efficiency, it can be seen that there is reasonable gain in efficiency (approximately 13-14%) in case of the proposed SE under RSS as compared to usual RSS estimator. Therefore, empirical results clearly support theoretical findings as it is already shown theoretically that the proposed SE is superior with respect to the usual RSS estimator of population mean. While comparing with SRSWOR estimator, the proposed SE in case of RSS always provides remarkable gain in efficiency and it increases with the increase in set size (m) for a fixed sample size (n).

5. Conclusion

In this article, a spatial estimation procedure for estimation of finite population mean was developed under RSS randomization framework. The proposed Spatial Estimator (SE) in case of RSS was developed through prediction approach based on IDW method. This estimator accounts for the spatial dependencies in population using the distances between sampling units. It has been shown that the proposed SE performs better than usual RSS estimator of population mean. However, simulation results show the proposed SE produces negligible bias in estimation of finite population mean.

Table 1 : Variance, % Bias, % CV, Skewness, Kurtosis and GE of the proposed SE under RSS with respect to usual RSS and SRS estimator for varying sample sizes (n) with different combination of set size (m) and number of cycles (r).

n	m	r	SE	Variance	% Bias	% CV	Skewness	Kurtosis	GE_{RSS}	GE_{SRS}
60	2	30	30.27	0.37	-0.28	1.99	-0.38	0.93	9.55	42.18
120	2	60	30.24	0.18	-0.39	1.41	0.27	0.59	13.52	21.06
180	2	90	30.30	0.10	-0.18	1.02	-0.38	0.62	14.36	50.46
60	3	20	30.23	0.31	-0.42	1.86	-0.06	-0.36	8.93	63.87
120	3	40	30.27	0.13	-0.31	1.21	-0.09	-0.48	13.28	64.28
180	3	60	30.28	0.09	-0.26	1.00	0.25	-0.01	13.68	59.77
60	4	15	30.23	0.30	-0.42	1.82	-0.16	-0.05	9.50	71.51
120	4	30	30.25	0.14	-0.37	1.22	-0.38	0.20	14.75	61.68
180	4	45	30.26	0.08	-0.32	0.94	-0.00	0.26	13.99	78.72
60	5	12	30.28	0.28	-0.25	1.75	-0.18	0.08	9.54	85.51
120	5	24	30.31	0.11	-0.17	1.11	-0.23	-0.03	11.76	93.89
180	5	36	30.30	0.06	-0.20	0.84	-0.13	0.17	16.02	126.12

It was also found to be more efficient than the usual RSS estimator. Thus, while sampling from a spatially correlated finite population using RSS design, it is suggested to use the proposed Spatial Estimation approach for estimation of population mean.

A. Appendix

For detailed derivations in Section 2 expectation over non-sampled units of the population $\left(E_{\bar{s}}\right)$ has to be taken on random variables $\overline{R}_j$ and $\overline{D}'_j$. For this purpose first the expectation over non-sampled units of the population $\left(E_{\bar{s}}\right)$ was obtained for the mean of Y's in the non-sample part as follows. These results can be replicated to the random variables $\overline{R}_j$ and $\overline{D}'_j$.

In the non-sample part in case of RSS, there are clearly two disjoint subsets. Suppose,

$\bar{s}$ = the set of all units in the population falling in the non-sample part of RSS design.

$= \bar{s}_1 \cup \bar{s}_2$

Where,

$\bar{s}_1$ = the subset containing all the units which were initially selected in the SRSWOR sample of m^2r units but further were not selected in the final RSS sample of size mr,

$\bar{s}_2$ = the subset containing all the units which were not selected in the initial SRSWOR sample of m^2r units,

and $\bar{s}_1 \cap \bar{s}_2 = \phi$.

Let us define, the mean of Y's in the non-sample part as

$$\bar{y}_{Non,RSS} = \frac{1}{N-mr}\sum_{j\in\bar{s}} y_j = \frac{1}{N-mr} \times\left[\sum_{k=1}^{r}\sum_{i'=1}^{m}\sum_{\substack{i=1\\ i\neq i'}}^{m} y_{(i:m)i'k} + \sum_{j\in\bar{s}_2} y_j\right] \tag{28}$$

Now, taking the expectation over non-sampled units of the population $\left(E_{\bar{s}}\right)$ on $\bar{y}_{Non,RSS}$,

$$E_{\bar{s}}\left(\bar{y}_{Non,RSS}\right) = \frac{1}{N-mr}\left[\sum_{k=1}^{r}\sum_{i'=1}^{m}\sum_{\substack{i=1\\ i\neq i'}}^{m}\sum_{j=1}^{N} Y_j P\left[y_{(i:m)i'k} = Y_j\right] + \sum_{j\in\bar{s}_2}\frac{1}{N}\sum_{j=1}^{N} Y_j\right]$$

$$= \frac{1}{N-mr}\left[\sum_{k=1}^{r}\sum_{i'=1}^{m}\sum_{\substack{i=1\\ i\neq i'}}^{m}\sum_{j=1}^{N} Y_j \frac{\binom{j-1}{i-1}\binom{N-j}{m-i}}{\binom{N}{m}} + \left(N-m^2r\right)\frac{1}{N}\sum_{j=1}^{N} Y_j\right]$$

$$= \frac{1}{N}\sum_{j=1}^{N} Y_j = \overline{Y} \tag{29}$$

For obtaining the variance of $\bar{y}_{Non,RSS}$, we have taken help of the following heuristic approach as

$$V\left(\bar{y}_{Non,RSS}\right) = \frac{1}{(N-mr)^2} V\left[N\overline{Y} - mr\,\bar{y}_{RSS}\right]$$

$$= \frac{1}{(N-mr)^2}\left[0 + m^2r^2 V\left(\bar{y}_{RSS}\right)\right]$$

$$= \frac{mr}{(N-mr)^2}\left\{\frac{N-1-mr}{N-1}\sigma_Y^2 - \bar{\gamma}_Y\right\} \tag{30}$$

References

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

Cochran, W. G. (1977). *Sampling Techniques.* New York, Wiley.

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

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

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

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

Patil, G. P., A. K. Sinha and C. Taillie (1995). Finite population corrections for ranked set sampling. *Ann. Inst. Stat. Math*., **47**, 621-636.

Rai, A. and P. Krishna (2013). Ranked Set Sampling from Finite Population under Randomization Framework. *J. Ind. Soc. Agricult. Statist*., **67(3)**, 108-117.

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

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

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

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

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