Int. J. Agricult. Stat. Sci. Vol. 14, No. 1, pp. 231-237, 2018 ***ISSN : 0973-1903***

ORIGINAL ARTICLE

SMALL AREA ESTIMATION UNDER A SPATIAL MODEL USING DATA FROM TWO SURVEYS

Sadikul Islam, Hukum Chandra*, Kaustav Aditya and Shashi Bhushan Lal

ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110 012, India.
E-mail: hchandra12@gmail.com

Abstract : In this paper, we describe small area estimation (SAE) under a spatial dependent random effects model by combining data from two independent surveys. The spatial dependence is introduced using simultaneous autoregressive (SAR) structure in the random area effects part of the model. We use data from two independent surveys. The first survey, small in sample size, collects both variable of interest as well as auxiliary variables and the second survey, relatively larger in sample size, has some auxiliary variables common to the first survey. Our empirical results, based on simulation studies, show that proposed SAE method using the data from two surveys is efficient as compared to the one based on data from single survey. Use of spatial information further enhances the efficiency of the proposed estimator.

Key words : Independent surveys, Borrowing strength, Empirical predictor, Spatial models, SAR.

1. Introduction

These days widespread demand from policy makers is for estimates at a finer level of geographic detail than the broad regions that were commonly used in the past, known as small domain (or small area). In this regards, SAE methods have achieved significant attention due to its potential to produce robust and reliable estimates at small area (or area) level. The usual design-based approach for these small areas are typically known as direct estimates in the literature. Model-dependent indirect SAE method that 'borrow strength' via statistical models are mostly used to produce reliable small area estimates [Rao and Molina (2015)]. However, sample sizes for small area of interest are typically small or some time even zero, in this situation though model dependent estimators perform better than direct estimators, leading to large sampling variability. Hence, survey data available for small area level, present to be insufficient to produce reliable estimate at small area level and sometimes even use of census data cannot provide satisfactory results for intermediate time point since in most of the countries census is conducted only once in a decade, during this period population characteristics of interest may change markedly over time.

Now a days, it is very common in many countries that different agencies, departments and organizations independently conduct surveys from the same target population for same or different purposes, having some auxiliary variable common to each other. In these scenario, combining data from more than one survey can be beneficial to produce reliable SAE methods. Zieschang (1990), Renssen and Nieuwenbroek (1997), Merkouris (2004) and Wu (2004) discussed the problem of combining data from different surveys to estimate totals at the population and large domain levels.

Kim and Rao (2012) described projection approach for estimation of population total and large domain totals by combining data from two independent surveys. An area-level model approach for SAE by combining information from several sources is discussed by Kim *et al*. (2015). Recently, Islam and Chandra (2017) developed an approach of small area estimation by combining the data from two independent surveys under random effects model. Their development is on fitting a random effects model to the first survey and then generating the proxy values of variable of interest for the second survey. They assumed that random area

****Author for correspondence*** ***Received*** *December 22, 2017* ***Revised*** *March 26, 2018* ***Accepted*** *April 16, 2018*

effects in the working model are independent. In practice most of the small area boundaries are drawn arbitrarily, hence there appears spatial correlation among small areas. In this paper we extend the idea of Islam and Chandra (2017) for combining data from two surveys under spatial dependent random effects model. Our aim is to how much gain achieved, incorporating spatial correlation in SAE using data from two surveys. The paper is organised as follow: In Section 2, we review the estimator developed by Islam and Chandra (2017) for random intercepts model with independent area effects and discuss these methods to account for spatial dependence between small areas. In Section 3, we describe the structure of our simulation study and present empirical results. We use real data set from the 1995-96 Australian Agricultural Grazing Industry Survey (AAGIS) conducted by the Australian Bureau of Agricultural and Resource Economics of 759 farms from 12 regions that make up the wheat-sheep zone for Australian broad acre agriculture. Finally, in Section 4, we provide some concluding remarks and make suggestions for further research.

2. Methodological Development

Let us assume that a finite population U containing N units can be partitioned into D non-overlapping domains U_i (i = 1, .., D) such that $\bigcup_{i=1}^{D} U_i = U$. We assume further that there is a known number N_i of population units in small area (or area) i such that $N = \sum_{i=1}^{D} N_i$. The area-specific mean of variable of interest y for small area i is $\bar{Y}_i = N_i^{-1} \sum_{j=1}^{N_i} y_{ij}$, where y_{ij} denote the value of y for unit j in area i. Let $\mathbf{x}_{ij}$ denote a vector of order p for auxiliary variables values of unit j in area i. We further assume that $\mathbf{x}_{ij}$ contains an intercept term as its first component. It is assumed that two surveys are independently conducted in the same population U. The first and second survey are denoted by the notation A and B, respectively. The samples of size $n_{(A)}$ and $n_{(B)}$ units from A and B are denoted by $s_{(A)}$ and $s_{(B)}$, respectively. The area specific sample sizes for $s_{(A)}$ and $s_{(B)}$ in area i are denoted by $n_{(A)i}$ and $n_{(B)i}$, respectively such that $n_{(A)} = \sum_{i=1}^{D} n_{(A)i}$ and $n_{(B)} = \sum_{i=1}^{D} n_{(B)i}$, where $n_{(A)} << n_{(B)}$. Further, it is assumed that the smaller survey A has collected both variable of interest y as well as set of auxiliary variables $\mathbf{x}$. The larger survey B has not collected data on variable of interest y, but it has collected set of auxiliary variables which are common to the first survey A. Here, (k) is used as a subscript to denote the quantities associated with survey k(k = A, B). The sample and non-sample part of the population U, with respect to k are denoted by $s_{(k)}$ and $r_{(k)}$, respectively. The area specific $n_{(ki)}$ sample and $N_i - n_{(ki)}$ non-sample units, with respect to sample k, are denoted by $s_{(ki)}$ and $r_{(ki)}$, respectively for area i. Let, w_{ij}^d denotes the sample weight of the first survey A for unit j in area i and $w_{ij} = w_{ij}^d \Big/ \sum_{j \in s_{(A)i}} w_{ij}^d$ with $\sum_{j \in s_{(A)i}} w_{ij} = 1$. Then the direct estimator (DIR) of small area i mean $\bar{Y}_i$ is given by $\hat{\bar{Y}}_{(A)i}^{DIR} = \sum_{j \in s_{(A)i}} w_{ij} y_{ij}$. The variance of the direct estimator $\hat{\bar{Y}}_{(A)i}^{DIR}$ can be approximated as

$$var\left(\hat{\bar{Y}}_{(A)i}^{DIR}\right) \approx \sum_{j \in s_{(A)i}} w_{ij}\left(w_{ij} - 1\right)\left(y_{ij} - \hat{\bar{Y}}_{(A)i}^{DIR}\right)^2.$$

The expression for variance estimator of the direct estimator is obtained from Särndal *et al.* (1992), page 43, 185 and 391, with $w_{ij} = 1/\pi_{(A)ij}$, $w_{ij} = 1/\pi_{(A)ij}$, $\pi_{(A)ij,ij} = \pi_{(A)ij}$ and $\pi_{(A)ij}\pi_{(A)ik}$, $j \neq k$, where $\pi_{(A)ij}$ is the first order inclusion probability of unit j in area i in A and $\pi_{(A)ij,ik}$ is the second order inclusion probability of units j and k in area i in A. Under simple random sampling, $w_{ij}^d = N_i n_{(A)i}^{-1}, w_{ij} = n_{(A)i}^{-1}$ and the direct estimator is area specific sample mean, $\hat{\bar{Y}}_{(A)i}^{DIR} = \bar{y}_{s(A)i} = n_{(A)i}^{-1} \sum_{j \in s_{(A)i}} y_{ij}$. Now we assume that population units follow unit level linear mixed model particularly random intercepts model [Chandra *et al.* (2007)] of the form

$$y_{ij} = x_{ij}^T \beta + u_i + e_{ij}. \qquad (1)$$

where, β is a vector of p unknown fixed effects, u_i

and e_{ij} denotes random effect of area i and individual random effect, respectively. It is commonly assumed that the random effects are Gaussian and these effects are mutually independent, both across individuals as well as across areas *i.e.* $u_i \sim N(0,\sigma_u^2)$ and $e_{ij} \sim N(0,\sigma_e^2)$. In the entire article, the sampling design is assumed as noninformative, so model assumed for sample units is also valid for non-sample units of the population.

The population level version of the linear mixed model of (1) is achieved by aggregating this over the whole population as

$$y = x\beta + zu + e \tag{2}$$

Here, $y = (y_1^T, ..., y_i^T, ..., y_D^T)$, $y_i = (y_{i1}, ..., y_{iN_i})^T$, $x = (x_1^T, ..., x_i^T, ..., x_D^T)^T$, $x_i = (x_{i1}, ..., x_{iN_i})^T$ is a $N_i \times p$ matrix, $z = diag(z_i = 1_{N_i}; 1 \le i \le D)$, here 1_{N_i} is the unit vector of length N_i, $u = (u_1, \ldots, u_D)^T$ and $e = (e_1^T, ..., e_D^T)^T$, $e_i = (e_{i1}, ..., e_{iN_i})^T$. Since different areas are independent, the covariance matrix of y has block diagonal structure given by

$$V = diag(V_i = \sigma_e^2 I_{N_i} + \sigma_u^2 z_i z_i^T ; 1 \le i \le D),$$ where I_{N_i} is the identity matrix of order N_i. We assume that x has full column rank p. In practice the variance components that define V are unknown and can be estimated from the sample data using methods described, for example, in Harville (1977).

We denote these estimates by $\hat{\Sigma} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2)^T$ and put a 'hat' on any quantity to denote estimate value of parameter. Thus $\hat{V} = diag(\hat{V}_i = \hat{\sigma}_e^2 I_{N_i} + \hat{\sigma}_u^2 z_i z_i^T ; 1 \le i \le D)$. Given a sample $s_{(A)}$ of size $n_{(A)}$ from this population, without loss of generality, we arrange the vector y so that its first $n_{(A)}$ elements correspond to the sample units, and then partition y, x, z and V according to sample and non-sample units. We can therefore rewrite (2) as follows :

$$y = \begin{bmatrix} y_{s(A)} \\ y_{r(A)} \end{bmatrix} = \begin{bmatrix} x_{s(A)} \\ x_{r(A)} \end{bmatrix} \beta + \begin{bmatrix} z_{s(A)} \\ z_{r(A)} \end{bmatrix} u + \begin{bmatrix} e_{s(A)} \\ e_{r(A)} \end{bmatrix}$$ and

$$V = \begin{bmatrix} v_{s(A)s(A)} & v_{s(A)r(A)} \\ v_{r(A)s(A)} & v_{r(A)r(A)} \end{bmatrix},$$ with

$$v_{s(A)s(A)} = diag\left\{\sigma_u^2 z_{n(A)i} z_{n(A)i}^T + \sigma_e^2 I_{n(A)i} ; i = 1, ..., D\right\}$$

and $V_{s(A)r(A)} = diag\left\{\sigma_u^2 z_{n(A)i} z_{N_i - n(A)i}^T ;\ i = 1, ..., D\right\}$.

The empirical predictor for small area mean in area i (denoted by EPA) is given by

$$\hat{\bar{Y}}_{(A)i}^{EPA} = \sum_{j \in s(A)i} w_{ij} \hat{y}_{ij} = \hat{\bar{x}}_{(A)i}^T \beta + \hat{u}_i , \tag{3}$$

where, $\hat{y}_{ij} = x_{(A)ij}^T \hat{\beta} + \hat{u}_i$, $\hat{u}_i = \hat{\gamma}_i (\bar{y}_{s(A)i} - \bar{x}_{s(A)i}^T \hat{\beta})$,

$\hat{\gamma}_i = \hat{\sigma}_u^2 (\hat{\sigma}_u^2 + \hat{\sigma}_e^2 / n_{(A)i})^{-1}$, $\bar{x}_{s(A)i} = n_{(A)i}^{-1} \sum_{j=1}^{n(A)i} x_{(A)i}$ and

$\hat{\bar{x}}_{(A)i} = \sum_{j=1}^{n(A)i} w_{ij} x_{(A)ij}$. Here, $\hat{\bar{x}}_{(A)i}$ is direct estimate of population mean of area i x_i with $E_d\left(\hat{\bar{x}}_{(A)i}\right) = \bar{x}_i$, where $E_d(.)$ denotes the expectation under a design. Let, $\varphi_{ij} = \varphi_{ij}^d \Big/ \sum_{j \in s_{(B)i}} \varphi_{ij}^d$ is a normalized survey weight of B for unit j in area i with $\sum_{j \in s_{(B)i}} \varphi_{ij} = 1$ and $\varphi_{(B)ij}^d$ is a survey weight of B for unit j in area i. Following, Islam and Chandra (2017), we define an empirical predictor of small area i using the data from both the survey A and B (denoted by EPB) as

$$\hat{\bar{Y}}_{(B)i}^{EPB} = \sum_{j \in s_{(B)i}} \varphi_{ij} \tilde{y}_{ij} = \hat{\bar{x}}_{(B)i}^T \hat{\beta} + \hat{u}_i \tag{4}$$

where, $\hat{\bar{x}}_{(B)i} = \sum_{j \in s_{(B)i}} \varphi_{ij} x_{ij}$ is direct estimate of population mean $\bar{x}_i$ with $E_d\left(\hat{\bar{x}}_{(B)i}\right) = \bar{x}_i$. In case of simple random sampling, $\varphi_{ij}^d = N_i n_{(B)i}^{-1}$, $\varphi_{ij} = n_{(B)i}^{-1}$ $\varphi_{ij} = n_{(B)i}^{-1}$,

$\hat{\bar{x}}_{(B)i} = \hat{\bar{x}}_{s(B)i} = n_{(B)i}^{-1} \sum_{j=1}^{n(B)i} x_{ij}$. The empirical estimator

EPA and EPB both are based on the assumption that small areas are independent. However, in practice areas near to each other are often spatially correlated. So there is a great opportunity to us to incorporate spatial correlation in Equation (2). In this paper, we investigate SAE based on linear models with spatially correlated small area effects. The neighbourhood structure of small areas are described by a contiguity matrix. Now we can re-express (2) by considering spatial dependence between different small areas as

$$y = x\beta + zv + e \tag{5}$$

where, $v = (I - \rho W)^{-1}u$ with $E(v) = 0$ and $Var(v) = \Omega = \sigma_u^2\left[(I_D - \rho W)(I_D - \rho W^T)\right]^{-1}$. We also assume a Simultaneous Autoregressive (SAR) error process [Chandra *et al.* (2007)], where the vector of random area effects $v = (v_i)$; $i = 1, ..., D$ satisfies $v = \rho Wv + u$. Here ρ denote spatial autoregressive coefficient, W denote proximity matrix of order D. The W describes how random effects from neighbouring areas are correlated, whereas ρ defines the strength of this spatial relationship. We define W as a contiguity matrix by considering non-zero values to the elements of W only for those pairs of areas that are adjacent. This matrix is defined in row-standardized form due to ease of interpretation; ρ in this case is called the spatial autocorrelation parameter [Chandra *et al.* (2007)]. The element w_{kl} of a contiguity matrix takes the value 1, if area l shares an edge with area k and 0 otherwise. In row-standardised form this becomes

$$w_{kl} = \begin{cases} d_k^{-1}; \; if \;\; k \, and \;\; 1 \; are \; contiguous \\ 0; \quad otherwise \end{cases}$$

where, d_k is the total number of areas that share an edge with area k (including area k itself). It follows that the covariance matrix of y is $Var(y) = V = z\Omega z^T + \sigma_e^2 I_N$ depends on vector of parameters $\Sigma = \left(\sigma_u^2, \sigma_e^2, \rho\right)^T$, that is practically unknown. Replacing it with maximum likelihood (ML) as well as restricted maximum likelihood (REML) [Chandra *et al.* (2007)] estimator $\hat{\Sigma} = \left(\hat{\sigma}_u^2, \hat{\sigma}_e^2, \hat{\rho}\right)^T$. The approximate value of ML and REML estimators $\hat{\sigma}_u^2, \hat{\sigma}_e^2$ and $\hat{\rho}$ can be obtained via a two-step procedure. The Nelder-Mead algorithm [Chandra *et al.* (2007)] is used at the first step to approximate these estimates and these approximates uses as starting values for a Fisher scoring algorithm. It is necessary because the log-likelihood function has multiple local maxima [Chandra *et al.* (2007)]. D_i is the D-vector $(0,0,0,...,1,...,0,0)^T$ with the 1 in position i, $\hat{v} = \hat{\Omega} z_{s(A)}^T \hat{V}_{s_{(A)}s_{(A)}}^{-1}\left(y_{s_{(A)}} - x_{s_{(A)}}\hat{\beta}^{sp}\right)$,

$\hat{V}_{s(A)s(A)} = \hat{\sigma}_e^2 I_{n(A)} + z_{s(A)}\hat{\sigma}_u^2\left[(I_D - \hat{\rho}W)(I_D - \hat{\rho}W^T)\right]^{-1} z_{s(A)}^T$,

where $\hat{\Omega} = \hat{\sigma}_u^2\left[(I_D - \hat{\rho}W)(I_D - \hat{\rho}W^T)\right]^{-1}$ and

$\hat{\beta}^{sp} = \left(x_{s_{(A)}}^T V_{s_{(A)}s_{(A)}}^{-1} x_{s_{(A)}}\right)^{-1} x_{s_{(A)}}^T \hat{V}_{s_{(A)}s_{(A)}}^{-1} y_{s_{(A)}}$ is the empirical best linear unbiased estimator (EBLUE) of β under (5). Then the empirical predictor of small area i under (5) using small survey A data (denoted by EPA.sp) is defined as

$$\hat{\bar{Y}}_{(A)i}^{EPA.sp} = \sum_{j \in s_{(A)i}} w_{ij}\hat{y}_{ij}^{sp} = \hat{\bar{x}}_{(A)i}^T\hat{\beta}^{sp} + \hat{v}_i. \tag{6}$$

where, $\hat{y}_{ij}^{sp} = x_{ij}^T\hat{\beta}^{sp} + \hat{v}_i$, x_{ij} are collected from small survey A. In a similar fashion, we can define empirical predictor of small area i under (5) using data from both the survey A and B (denoted by EPB.sp) as

$$\hat{\bar{Y}}_{(B)i}^{EPB.sp} = \sum_{j \in s_{(B)i}} \varphi_{ij}\tilde{y}_{ij}^{sp} = \hat{\bar{x}}_{(B)i}^T\hat{\beta}^{sp} + \hat{v}_i. \tag{7}$$

where, $\tilde{y}_{ij}^{sp} = x_{ij}^T\hat{\beta}^{sp} + \hat{v}_i, x_{ij}$ are collected from large survey B.

3. Empirical Evaluations

In this Section, we report the results from simulation study that illustrate the performance of the different estimators of small area means defined in the preceding sections. The estimators considered in simulation studies are

- DIR-Direct estimator based on data from survey A;
- EPA-Empirical predictor (3) based on data from survey A;
- EPB-Empirical predictor (4) based on data from both survey A and B;
- EPA.sp-Empirical predictor (6) based on data from survey A;
- EPB.sp-Empirical predictor (7) based on data

from both survey A and B.

The performance of the different estimators in the simulation studies is evaluated by computing the average percentage relative bias (RB), the average percentage relative root mean squared error (RRMSE) and the average percentage relative efficiency (RE) defined by

$$RB = mean\left\{ \overline{m}_i^{-1} H^{-1} \sum_{h=1}^{H} \left(\hat{m}_{ih} - m_i \right) \right\} \times 100$$

$$RRMSE = mean\left\{ \sqrt{ H^{-1} \sum_{h=1}^{H} \left(\frac{\hat{m}_{ih} - m_i}{m_i} \right)^2 } \right\} \times 100 \text{ and}$$

$$RE(estimator) = \frac{RMSE(DIR)}{RMSE(estimator)} \times 100$$

$$RMSE = mean\left\{ \sqrt{ H^{-1} \sum_{h=1}^{H} \left(\hat{m}_{ih} - m_i \right)^2 } \right\} \times 100.$$

Here, the subscript i indexes for small areas and the subscript h indexes for H Monte Carlo simulations, m_i denoting the true area i mean, with predicted value

$$\hat{m}_{ih}, \overline{m}_i = H^{-1} \sum_{h=1}^{H} m_{ih}.$$

A real data set collected in the 1995-96 Australian Agricultural Grazing Industry Survey (AAGIS) conducted by the Australian Bureau of Agricultural and Resource Economics of 759 farms from 12 regions, which make up the wheat-sheep zone for Australian broad acre agriculture is used for simulation. Here survey variable of interest was TCC-total cash costs (in Australian dollars, A$) and Id-unique identification code to each industry, COORD1-latitude, COORD2-longitude, regions, LANDCL- land area (in hectares), BEEFCL-number of closing beef stock, SHEEPCL-number of closing sheep stock and WHEATQ-wheat quantity harvested (in Kilograms) used as auxiliary variables, respectively. Our aim is to estimate the averages of TCC in each of the 12 regions. Now, the original sample data is used to generate a population of size N = 39,562 farms by sampling N times with replacement, with probability proportional to a farm's sample weight. H = 1000 independent samples are selected from this simulated population using stratified random sampling, with regions are strata with region-specific sample size fixed to be the same as in original sample discussed in Table 1. This sample is treated as the larger samples (second sample, $s_{(B)}$ of size $n_{(B)}$ = (759). The first sample (smaller sample), $S_{(A)}$ of size $n_{(A)}$ = 60 is selected with area specific sample sizes, $n_{(A)i}$ = 5 by using stratified random sampling, with regions are strata and this process is repeated H = 1000 times independently to get 1000 independent samples. Similar to $n_{(A)}$ = 60 two additional sample sizes, $n_{(A)}$ = 120 and 240 with area-specific sample sizes as $n_{(A)}$ = 10 and 20 respectively are also considered for smaller sample, $s_{(A)}$ to get 1000 independent samples. We discuss the performance of the estimators, namely DIR, EPA, EPB, EPA.sp and EPB.sp in simulation study in Table 2. Figs. 1 and 2 demonstrate the boxplots of dispersion of region-specific values of percentage RB and relative RMSE for all the estimators. The results in Table 2 reveal that the EPB.sp has lowest relative bias followed by EPB estimator, except for $n_{(A)i}$ = 10. Fig. 1 shows the boxplots in terms of the dispersion of region-specific values of actual percentage RB for the estimators namely, DIR, EPA, EPB, EPA.sp and EPB.sp. We see that all the empirical estimator EPA, EPA.sp, EPB and EPB.sp does not outperform the DIR estimator though the incorporation of spatial dependence narrow down the dispersion of EPA.sp and EPB.sp than EPA and EPB, respectively. From the Table 2, we observe that combining information from two independent surveys reduce the bias and also shows that incorporation of spatial correlation has significant effect on bias reduction. Again, the EPB.sp outperforms minimum relative RMSE and maximum efficiency followed by EPB estimator for all the sample size combinations. Fig. 2 confirms the observation of Table 2 on relative RMSE in the simulation study. Again, the EPB.sp outperforms DIR, EPA, EPB and EPA.sp in terms of the dispersion of relative RMSE between regions for all the sample sizes of first (*i.e.* smaller) sample. Generally, the results described in Table 2 and Fig. 2 support the conclusion that the combining data from two surveys improves estimates of small areas when consider spatial dependent random effects, with the proposed EPB.sp emerging as the best performing of the methods that we investigated in the empirical evaluations.

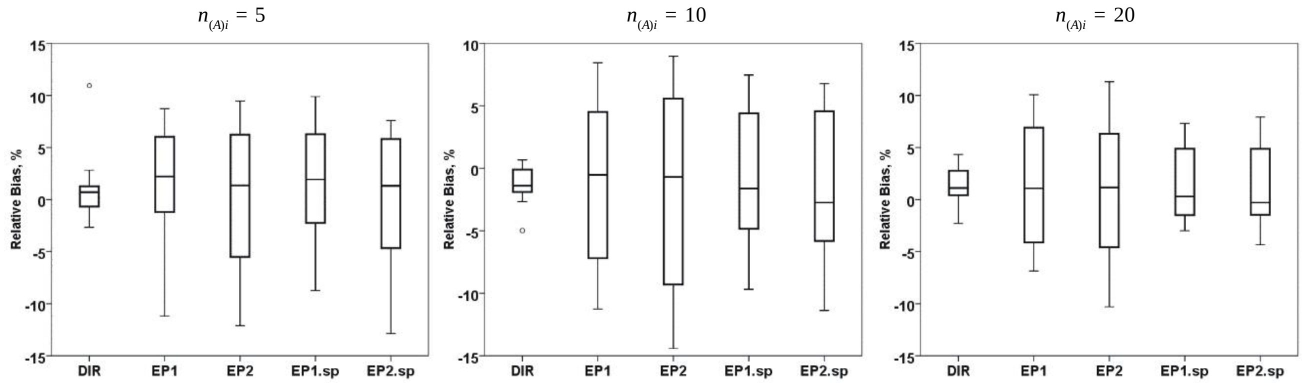

Fig. 1 : Boxplots of region-specific values of relative bias of the different predictors from simulations using the AAGIS data.

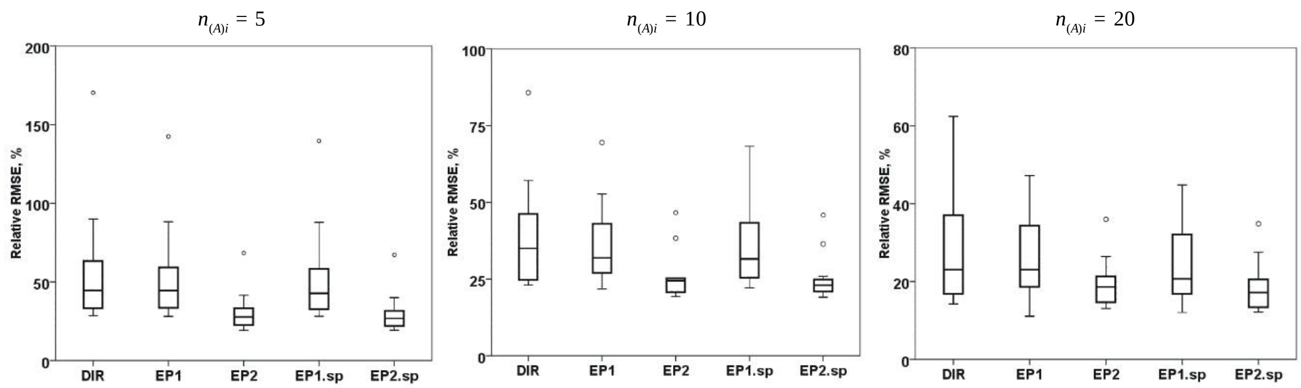

Fig. 2 : Boxplots of region-specific values of relative RMSE of the different predictors from simulations using the AAGIS data.

Table 1 : Region wise population size and corresponding sample size in 12 regions.

Region	1	2	3	4	5	6	7	8	9	10	11	12
N_i	5918	1776	4770	2929	2335	3719	4090	1450	3731	1901	4960	1683
$n_{(B)i}$	88	44	73	54	58	84	63	42	87	47	76	43

Table 2 : Values of percentage relative biases (RB), percentage relative root mean squared errors (RRMSE) and percentage relative efficiencies (RE) of the different estimators from simulations study using the AAGIS data. The values are averaged over 12 regions.

Predictor	$n_{(A)i} = 5$			$n_{(A)i} = 10$			$n_{(A)i} = 20$		
	RB, %	RRMSE, %	RE, %	RB,%	RRMSE,%	RE, %	RB,%	RRMSE, %	RE, %
DIR	1.16	57.6	100	-1.33	38.9	100	1.36	28.2	100
EPA	1.36	54.3	106	-1.14	36.5	106	1.45	25.9	109
EPA.sp	1.57	53.3	108	-0.85	35.6	109	1.35	24.1	117
EPB	- 0.07	31.2	185	-1.60	26.0	149	1.23	19.6	144
EPB.sp	0.05	30.1	192	-1.32	25.3	154	1.17	18.6	152

4. Concluding Remarks

This paper presents result from an initial exploration of SAE of means by combining information from two independent surveys under both spatial independent random effects model as well as spatial dependent random effects model. Our empirical results based on real data show that SAE by combining data from two independent surveys is beneficial in terms of efficiency. It is also observed that incorporation of spatial correlation in the linear mixed model for combing data from independent two surveys indicate significant gains in SAE. The empirical predictor based on spatial model using data from both the survey is the most efficient small area estimator. There is an important issue that still need to be explored in the context of mean squared error (MSE) estimation of the developed empirical predictor of means under a spatial dependent random effects model using data from two surveys and we are working on this.

Acknowledgement

The authors would like to acknowledge the valuable comments and suggestions of the Editor and the referee. The first author gratefully acknowledges the National Fellowship[F./2015-16/NFO-2015-17-OBC-WES-36385/(SA-III/website)] provided by Ministry of Social Justice & Empowerment, Govt. of India, New Delhi for pursuing Ph. D. degree.

References

Chandra, H., N. Salvati and R. Chambers (2007). Small area estimation for spatially correlated population-a comparison of direct and indirect model-based estimators. *Statistics in Transition*, **8**, 331-350.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72(358)**, 320-338.

Islam, S. and H. Chandra (2017). Small area estimation combining data from two independent surveys. (Submitted).

Kim, J. K. and J. N. K. Rao (2012). Combining data from two independent surveys: a model assisted approach. *Biometrika*, **99**, 85-100.

Kim, J., S. Park and S. Kim (2015). Small area estimation combining information from several sources. *Survey Methodology*, **41**, 21-36.

Merkouris, T. (2004). Combining independent regression estimators from multiple surveys. *Journal of American Statistical Association*, **99**, 1131-1139.

Merkouris, T. (2010). Combining information from multiple surveys by using regression for efficient small domain estimation. *Journal of the Royal Statistical Society: Series B*, **72**, 27-48.

Rao, J. N. K. and I. Molina (2015). *Small Area Estimation*. John Wiley & Sons. Inc., New Jersey, 2nd edition.

Renssen, R. H. and N. Nieuwenbroek (1997). Aligning estimates for common variables in two or more sample surveys. *Journal of American Statistical Association*, **92**, 368-375.

Särndal, C. E., B. Swensson and J. H. Wretman (1992). *Model Assisted Survey Sampling*. Springer-Verlag, New York.

Wu, C. (2004). Combining information from multiple surveys through the empirical likelihood method. *Canadian Journal of Statistics*, **32**, 15-26.

Zieschang, K. D. (1990). Sample weighting methods and estimation of totals in the Consumer Expenditure Survey. *Journal of the American Statistical Association*, **85(412)**, 986-1001.