*Journal of Statistical Research* *ISSN 0256 - 422 X*
*2011, Vol. 45, No. 1, pp. 59-70*
*Bangladesh*

# A COMPARATIVE STUDY OF COMPOSITE ESTIMATORS FOR SMALL DOMAINS UNDER LAHIRI-MIDZUNO SAMPLING SCHEME

G. C. TIKKIWAL
*Department of Mathematics and Statistics, JNV University, Jodhpur 342 005, India*
*Email: gctikkiwal@yahoo.com*

T. JURKIEWICZ
*Department of Statistics, University of Gdansk, Poland*
*Email: t.jurkiewicz@zr.univ.gda.pl*

SUMMARY

This paper proposes three composite estimators for small domains, which are the weighted sum of direct and synthetic ratio estimators under Lahiri-Midzuno sampling scheme. Further, it compares performance of the composite estimators empirically for estimating crop acreage for small domains. The study shows that the composite estimator, which is a weighted sum of the direct ratio and synthetic ratio estimators, performs better than the other two estimators under certain conditions.

*Keywords and phrases:* Composite estimators, Synthetic ratio estimators, Small domains, Lahiri-Midzuno sampling scheme, SICURE model.

*AMS Classification:* Place Classification here. Leave as is, if there is no classification

# 1 Introduction

Gonzalez and Wakesberg (1973) and Schaible et al. (1977) compare errors of synthetic and direct estimators for standard Metropolitan Statistical Areas and counties of U.S.A. The authors of both the papers conclude that when in small domains sample sizes are relatively small the synthetic estimator out performs the simple direct, whereas, when sample sizes are large the direct outperforms the synthetic. These results suggest that a weighted sum of these two estimators, known as composite estimator, can provide an alternative to choosing one over the other. Tikkiwal and Tikkiwal (1998) and Tikkiwal and Ghiya (2004) define a generalized class of composite estimators for small domains using auxiliary variable, under simple random sampling and stratified random sampling schemes. Further, the authors compare the relative performance of the estimators belonging to the generalized class with the corresponding direct and synthetic estimators. The study suggests the use of composite estimator, combining direct and synthetic ratio estimators, as it has smaller relative bias and standard error. Tikkiwal and Pandey (2007) under Lahiri-Midzuno (L.M) Scheme confirm these results.

In this paper we study the performance of three composite estimators for small domains, under Lahiri-Midzuno scheme of sampling. These estimators are weighted sum of synthetic ratio and a direct: unbiased, almost unbiased and ratio estimator respectively.

## 2 Notations

Suppose that a finite population $U = (1, \ldots, i, \ldots, N)$ is divided into '$A$' non overlapping small domains $U_a$ of given size $N_a(a = 1, \ldots, A)$ for which estimates are required. We denote that characteristic under study by '$y$'. We further assume that the auxiliary information is available and denote this by '$x$'. A random sample $s$ of size $n$ is selected through Larihi-Midzuno sampling scheme (1951, 52) from population $U$ such that $n_a$ units in the sample '$s$' comes from small domain $U_a(a = 1, \ldots, A)$. Consequently, $\sum_{a=1}^{A} N_a = N$ and $\sum_{a=1}^{A} n_a = n$ We denote the various population and sample means for characteristics $Z = X, Y$ by

$$\begin{aligned} \bar{Z} &= \text{mean of the population based on } N \text{ observations.} \\ \bar{Z}_a &= \text{population mean of domain `}a\text{' basted on } N_a \text{ observations.} \\ \bar{z} &= \text{mean of the sample `}s\text{' based on } n \text{ observations} \\ \bar{z}_a &= \text{sample mean of domain `}a\text{' based on } n_a \text{ observations.} \end{aligned}$$

Also, the various mean squares and coefficient of variations of the population '$U$' for characteristics $Z$ are denoted by

$$S_z^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(z_i - \bar{Z}\right)^2, \qquad C_z = \frac{S_z}{\bar{Z}}$$

The coefficient of covariance between $X$ and $Y$ is denoted by $C_{xy} = S_{xy}/(\bar{X}\bar{Y})$, where

$$S_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} \left(y_i - \bar{Y}\right) \left(x_i - \bar{X}\right).$$

The corresponding various mean squares and coefficient of variations of small domains $U_a$ are denoted by

$$S_{z_a}^2 = \frac{1}{N_a - 1} \sum_{i=1}^{N_a} \left(Z_{a_i} - \bar{Z}_a\right)^2, C_{z_a} = \frac{S_{z_a}}{\bar{Z}_a} \; and \; C_{x_a y_a} = \frac{S_{x_a y_a}}{\bar{X}_a \bar{Y}_a},$$

where

$$S_{x_a y_a} = \frac{1}{N_a - 1} \sum_{i=1}^{N_a} \left(y_{a_i} - \bar{Y}_a\right) \left(x_{a_i} - \bar{X}_a\right)$$

and $z_{ai}$ $(a = 1, \ldots, A; i = 1, \ldots, N_a)$ denote the $i^{\text{th}}$ observation of the small domain '$a$' for the characteristic $Z = X, Y$.

## 3 Synthetic ratio estimator

We consider here synthetic ratio estimator of population mean $\bar{Y}_a$ , based on auxiliary information under Lahiri-Midzuno sampling scheme, as described in previous section. The synthetic ratio estimator of population mean $\bar{Y}_a$ of small area 'a' is defined as follows:

$$\bar{Y}_{syn,a} = \frac{\bar{y}}{\bar{x}}\bar{X}_a. \tag{3.1}$$

This estimator may be heavily biased unless the following assumption

$$\left(\bar{Y}_a / \ \bar{X}_a\right) \approx \left(\bar{Y} \ / \ \bar{X}\right) \qquad \text{for all } a \in \ A. \tag{3.2}$$

is satisfied.

## 3.1 Bias and mean square error

Under Lahiri-Midzuno sampling scheme

$$E\left(\bar{y}_{syn,a}\right) = E\left(\frac{\bar{y}}{\bar{x}}\bar{X}_a\right) = \frac{\bar{X}_a}{\bar{X}}E\left(\frac{\bar{y}}{\bar{x}}\bar{X}\right) = \frac{\bar{X}_a}{\bar{X}}\bar{Y}. \tag{3.3}$$

Therefore, design bias of $\bar{y}_{syn,a}$ is

$$B\left(\bar{y}_{syn,a}\right) = \left(\frac{\bar{Y}}{\bar{X}}\bar{X}_a - \bar{Y}_a\right) = B_1(say). \tag{3.4}$$

The mean square error of $\bar{y}_{syn,a}$ is given by

$$MSE\left(\bar{y}_{syn,a}\right) = \frac{\bar{X}_a^2}{\bar{X}^2}V\left(\frac{\bar{y}}{\bar{x}}\bar{X}\right) + B_1^2 = \frac{\bar{X}_a^2}{\bar{X}^2}\left[\frac{1}{\binom{N}{n}}\sum_c \left(\frac{\bar{y}^2}{\bar{x}}\right)_c - \bar{Y}^2\right] + B_1^2, \tag{3.5}$$

where $\sum_c$ stands for summation over all possible samples.

*Remark* 1. The above expression of $MSE\left(\bar{y}_{syn,a}\right)$ is not in analytical form

## 3.2 Estimation of mean square error

The $MSE\left(\bar{y}_{syn,a}\right)$ can be estimated by the following expression

$$\begin{aligned} mse\left(\bar{y}_{syn,a}\right) &= \frac{\bar{X}_a^2}{\bar{X}^2}\nu\left(\bar{y}_R\right) + \hat{B}_1^2 \\ &= \frac{\bar{X}_a^2}{\bar{X}^2}\left[\bar{y}_R^2 - \frac{\bar{X}}{\bar{x}}\left\{\bar{y} - \left(\frac{1}{n} - \frac{1}{N}\right)s_y^2\right\}\right] + \left(\frac{\bar{X}_a}{\bar{X}}\bar{y} - \bar{y}_a\right)^2. \end{aligned} \tag{3.6}$$

Further, if the synthetic assumption given in equations (3.2) satisfies then $B_1 = B\left(\bar{y}_{syn,a}\right) = 0$ and hence consistent estimator of $MSE\left(\bar{y}_{syn,a}\right)$ is

$$mse\left(\bar{y}_{syn,a}\right) = \frac{\bar{X}_a^2}{\bar{X}^2}v\left(\bar{y}_R\right) = \frac{\bar{X}_a^2}{\bar{X}^2}\left[\bar{y}_R^2 - \frac{\bar{X}}{\bar{x}}\left\{\bar{y} - \left(\frac{1}{n} - \frac{1}{N}\right)s_y^2\right\}\right], \tag{3.7}$$

where $\bar{y}_R = \frac{\bar{y}}{\bar{x}}\bar{X}$

### 3.3 Comparison under SRSWOR

The expressions of Bias and Mean Square Error of synthetic ratio estimator under SRSWOR scheme is given by Tikkiwal and Ghiya (2000), while discussing the properties of generalized class of synthetic estimator, as under

$$B_2 = B\left(\bar{y}_{syn,a}\right) = \frac{\bar{Y}}{\bar{X}}\bar{X}_a\left[1 + \frac{N-n}{Nn}\left(C_x^2 - C_{xy}\right)\right] - \bar{Y}_a \tag{3.8}$$

and

$$\begin{aligned} MSE\left(\bar{y}_{syn,a}\right) = &\left(\frac{\bar{Y}}{\bar{X}}\bar{X}_a\right)^2\left[1 + \frac{N-n}{Nn}\left(3C_x^2 + C_y^2 - 4C_{xy}\right)\right] \\ &- 2\bar{Y}_a\left(\frac{\bar{Y}}{\bar{X}}\bar{X}_a\right)\left[1 + \frac{N-n}{Nn}\left(C_x^2 - C_{xy}\right)\right] + \bar{Y}_a^2. \end{aligned} \tag{3.9}$$

Comparing the expression of biases $B_1$ and $B_2$ of $\left(\bar{y}_{syn,a}\right)$ under L-M and SRSWOR schemes, we get from equations (3.4) and (3.8)

$$B_2 - B_1 = \frac{N-n}{Nn}\frac{\bar{Y}}{\bar{X}}\bar{X}_a\left(C_x^2 - C_{xy}\right)$$

So, $B_2 \geq B_1$ if $C_x^2 - C_{xy} \geq 0 \Rightarrow \rho\frac{C_y}{C_x} \leq 1$.

*Remark* 2. If the synthetic assumption given in Eq. (3.2) satisfies then the expression of bias $B_2$ given in Eq. (3.8) reduces to

$$B_2 = \frac{N-n}{Nn}\left(C_x^2 - C_{xy}\right)\bar{Y}_a. \tag{3.10}$$

That is, $B_2 \neq 0$ even if synthetic assumption is satisfied. Whereas under this condition $B_1 = 0$

*Remark* 3. If the synthetic assumption is satisfied then the expressions of $MSE\left(\bar{y}_{syn,a}\right)$ given in equations (3.5) and (3.9) reduces respectively to

$$M_1 = MSE\left(\bar{y}_{syn,a}\right) = \frac{\bar{X}_a^2}{\bar{X}^2}\left[\frac{1}{\binom{N}{n}}\sum_c\left(\frac{\bar{y}^2}{\bar{x}}\right)_c - \bar{Y}^2\right] \tag{3.11}$$

and

$$M_2 = MSE\left(\bar{y}_{syn,a}\right) = \frac{N-n}{Nn}\left(C_x^2 + C_y^2 - 2C_{xy}\right)\bar{Y}_a^2. \tag{3.12}$$

Tikkiwal and Pandey (2007) compares the performance of $\left(\bar{y}_{syn,a}\right)$ under the two schemes and noted that the synthetic ratio estimator performs better under L-M scheme.

## 4 Composite estimators

A natural way to balance the potential bias of synthetic estimator $\left(\bar{y}_{syn,a}\right)$ against the instability of a direct estimator, say $\bar{y}_{d,a}$, is to take a weighted average of $\bar{y}_{d,a}$ and $\bar{y}_{syn,a}$. Such composite estimators of population mean $\bar{Y}_a$ of small area $a$ may be written as

$$\bar{y}_{c,a} = w_a\bar{y}_{d,a} + \left(1 - w_a\right)\bar{y}_{syn,a} \tag{4.1}$$

for a suitably chosen weights $w_a$ $(0 \leq w_a \leq 1)$. Many of the estimators proposed in the literature, both design-based and model-based, have the composite form (3.13). We in this section propose three different composite estimators and study them under L-M design.

Let $U$ be a population with $N$ elements, and let $U_a$ be a domain of $U$, that is, a sub-set of $U$, with $Na < N$. We assume that the elements belonging to $U_a$ can not be identified before hand but the domain size $N_a$ is known. The population parameter of interest here is $\bar{Y}_a$, the mean per element in the domain $U_a$. Further, let a sample $s$ of size $n$ is selected from the whole of $U$ through Lahiri-Midzuno sampling design (1951, 52). Now, $s$ will typically contain some elements from $U_a$, and some from rest of $U$. Let $s_a$ be the sub-set of $s$ consisting of the elements from $U_a$, that is $s_a = s \bigcap U_a$. The number of elements in $s_a$ is random; the probability that $s_a$ being empty is assumed to be negligible.

## 4.1 Composite estimator -1

We define first composite estimator $\bar{y}_{1a}$ by taking $\bar{y}_{d,a} = \frac{1}{N_a} \sum_{k \in s_a} (y_k/\Pi_k) = \bar{y}_{a,\Pi}$ in the general form of composite estimator (4.1). That is,

$$\bar{y}_{1a} = w_a \left\{ \frac{1}{N_a} \sum_{k \in s_a} (y_k/\Pi_k) \right\} + (1 - w_a) \frac{\bar{y}}{\bar{x}} \bar{X}_a, \tag{4.2}$$

where the direct estimator $\bar{y}_{a,\pi}$ is unbiased in general; and the synthetic estimator $\bar{y}_{syn,a} = (\bar{y}/\bar{x})\, \bar{X}_a$ is design biased as discussed in Section 3. Further, the variance of $\bar{y}_{a,\pi}$ is

$$V(\bar{y}_{a,\Pi}) = \frac{1}{N_a^2} \sum_{k \in U_a} \sum_{l \in U_a} \Delta_{kl} \left( \frac{y_k}{\Pi_k} \right) \left( \frac{y_l}{\Pi_l} \right) \tag{4.3}$$

and the unbiased variance estimator is

$$v(\bar{y}_{a,\Pi}) = \frac{1}{N_a^2} \sum_{k \in s_a} \sum_{l \in s_a} \frac{\Delta_{kl}}{\Pi_{kl}} \frac{y_k}{\Pi_k} \frac{y_l}{\Pi_l}, \tag{4.4}$$

where

$$\Delta_{kl} = \Pi_{kl} - \Pi_k \Pi_l \qquad \text{for all } k, l$$

under the L - M design

$$\Pi_k = \frac{N-n}{N-1} p_k + \frac{n-1}{N-1} \text{ for all } k \tag{4.5}$$

and

$$\Pi_{kl} = \begin{cases} \frac{(n-1)}{(N-1)} \left\{ \left( \frac{N-n}{N-2} \right) (p_i + p_j) + \left( \frac{n-2}{N-2} \right) \right\}, & \text{for all } k \neq l \\ \frac{(N-n)}{(N-1)} p_k + \frac{(n-1)}{(N-1)} & \text{for all, } k = l. \end{cases} \tag{4.6}$$

### 4.2 Composite estimator - 2

The second composite estimator $\bar{y}_{2a}$ of $\bar{Y}_a$ we define by taking

$$\bar{y}_{d,a} = \frac{\sum_{k \in s_a} (y_k / \Pi_k)}{\sum_{k \in s_a} (1/\Pi_k)} = \tilde{y}_{s,a}$$

in the form (4.1). That is,

$$\bar{y}_{2a} = w_a \tilde{y}_{s,a} + (1 - w_a)\, \bar{y}_{syn,a}. \tag{4.7}$$

Hence, the direct estimator $\tilde{y}_{s,a}$ is approximately unbiased and its approximately variance is

$$AV\left(\hat{y}_{s,a}\right) = \frac{1}{N_a^2} \sum_{k \in U_a} \sum_{l \in U_a} \Delta_{kl} \left( \frac{y_k - \bar{y}_a}{\Pi_k} \right) \left( \frac{y_l - \bar{y}_a}{\Pi_l} \right) \tag{4.8}$$

and the variance estimator is

$$v\left(\hat{y}_{s,a}\right) = \frac{1}{N_a^2} \sum_{k \in s_a} \sum_{l \in s_a} \frac{\Delta_{kl}}{\Pi_{kl}} \left( \frac{y_k - \tilde{y}_{s_a}}{\Pi_k} \right) \left( \frac{y_l - \tilde{y}_{s_a}}{\Pi_l} \right). \tag{4.9}$$

### 4.3 Composite estimator - 3

The third composite estimator $\bar{y}_{3a}$ of $\bar{Y}_a$ by taking $\bar{y}_{d,a} = (\bar{y}_a / \bar{x}_a)\, \bar{X}_a = \bar{y}_{ar,}$ a design biased estimator of $\bar{Y}_a$. That is,

$$\bar{y}_{3a} = w_a \bar{y}_{ar} + (1 - w_a)\, \bar{y}_{syn,a}, \tag{4.10}$$

where an estimator of the approximate variance of $\bar{y}_{ar}$ is

$$v\left(\bar{y}_{ar}\right) = \left\{ \frac{\bar{X}_a}{\sum_{k \in s_a} (x_k / \Pi_k)} \right\}^2 \sum_{k \in s_a,} \sum_{l \in s_a} \frac{\Delta_{kl}}{\Pi_{kl}} \left( \frac{y_k - \hat{B}_a x_k}{\Pi_k} \right) \left( \frac{y_l - \hat{B}_a x_l}{\Pi_l} \right). \tag{4.11}$$

The values of $\Pi_k$ and $\pi_{kl}$ for all k,l are given by equations (4.5) and (4.6).

**Note:** The justification of considering the three composite estimators is that under L-M design the synthetic estimator $\bar{y}_{syn,a}$ under synthetic assumption is design unbiased. Further, the direct estimators $\bar{y}_{a,\Pi}$ is unbiased and the two other direct estimators $\tilde{y}_{s,a}$ and $\bar{y}_{ar}$ are approximately unbiased.

### 4.4 Estimation of weights

The optimum values $w_a'$ of $w_a$ may be obtained by minimizing the mean square error of $\bar{y}_{c,a}$ with respect to $w_a$ and it is given by

$$w_a' = \frac{MSE\left((\bar{y}_{syn,a}) - E\left(\bar{y}_{d,a} - \bar{Y}_a\right)\ \left(\bar{y}_{syn,a} - \bar{Y}_a\right)\right.}{MSE\left(\bar{y}_{d,a}\right) + MSE\left(\bar{y}_{syn,a}\right) - 2E\left(\bar{y}_{d,a} - \bar{Y}_a\right)\ \left(\bar{y}_{syn,a} - \bar{Y}_a\right)}.$$

Under the assumption that $E\left(\bar{Y}_{d,a}-\bar{Y}_a\right)\ \left(\bar{Y}_{syn,a}-\bar{Y}_a\right)$ is small relative to $MSE\left(\bar{y}_{syn,a}\right)$, the $w_a'$ reduces to

$$w_a^* = \frac{MSE\left(\left(\bar{y}_{syn,a}\right)\right.}{MSE\left(\bar{y}_{d,a}\right)+MSE\left(\bar{y}_{syn,a}\right)}. \tag{4.12}$$

Since $\bar{y}_{syn,a}$ is not an unbiased estimator, therefore, an unbiased estimator of $MSE\left(\bar{y}_{syn,a}\right)$ under the assumption that $cov\left(\bar{y}_{d,a},\bar{y}_{syn,a}\right)=0$, is given by [cf.Rao (2003), equation (4.2.12)].

$$mse\left(\bar{y}_{syn,a}\right)=\left(\bar{y}_{syn,a}-\bar{y}_{d,a}\right)^2-v\left(\bar{y}_{d,a}\right) \tag{4.13}$$

Now, using equation (4.12) the weights $w_a^*$ can be estimated as follows:

$$\hat{w}_a^* = \frac{mse\left(\bar{y}_{syn,a}\right)}{\left(\bar{y}_{syn,a}-\bar{y}_{d,a}\right)^2} \tag{4.14}$$

But this estimator of $w_a^*$ can be very unstable. Schaible (1978) proposes an average weighting scheme based on several variables or "similar" areas or both, to overcome this difficulty.

In our empirical study, presented in next section, we use both the methods for estimating the weights.

# 5 Crop acreage estimation for small domains - A simulation study

In this section we compare the relative performance of the three composite estimators $\bar{y}_{1a}$, $\bar{y}_{2a}$ and $\bar{y}_{3a}$ of $\bar{Y}_a$ under L-M sampling scheme, through a simulation study, as the mean square errors of $\bar{y}_{ka}\,(k=1,2,3)$ are not in analytical form. This we do by taking up crop acreage data from the State of Rajasthan, one of the States in India, for our case study.

## 5.1 Existing methodology for estimation

In order to improve timelines and quality of crop acreage statistics, a scheme known as Timely Reporting Scheme (TRS) has been in vogue since early seventies in most of the States of India. The TRS has the objective of providing quick and reliable estimates of crop acreage statistics and thereby production of the principle crops during each agricultural season. Under the scheme the Patwari (Village Accountant) is required to collect acreage statistics on a priority basis in a 20 percent sample of villages. These statistics are further used to provide state level estimates using direct estimators viz. unbiased (based on sample mean) and ratio estimators.

The performance of both the direct estimators in the State of Rajasthan, like in other states, is satisfactory at state level, as the sampling error is within 5 percent. However, the sampling error of both the direct estimators increases considerably, when they are used for estimating acreage statistics of various principle crops even at district level, what to speak of levels lower than a district. For example, the sampling error of direct ratio estimator for Kharif crops (the crop sown in June-July and harvested in October-November every year) of Jodhpur district (of Rajasthan State) for the agricultural season 1991-92 varies approximately between 6 to 68 percent. Therefore, there is need to use indirect estimators at district and lower levels for decentralized planning and other purposes like crop insurance, bank loan to farmers.

## 5.2 Details of the simulation study

For collection of revenue and for administrative purposes, the State of Rajasthan, like most of the other States of India, is divided into a number of districts.

Further, each district is divided into a number of Tehsils and each Tehsil is also divided into a number of Inspector Land Revenue Circles (ILRCs). Each ILRC consists of a number of villages. For the present study, we take ILRCs as small domains.

In the simulation study, we undertake the problem of crop acreage estimation for all Inspector Land Revenue Circles (ILRCs) of Jodhpur Tehsil of Rajasthan. They are seven in number and these ILRCs contain respectively 29, 44, 32, 30, 33, 40 and 44 villages. These ILRCs are small domains from the TRS point of view. The crop under consideration is Bajra (Indian corn or millet) for the agriculture season 1993-94. The bajra crop acreage for agriculture season 1992-93 is taken as the auxiliary characteristic x.

To assess the relative performance of the estimators $\bar{y}_{1a}$, $\bar{y}_{2a}$ and $\bar{y}_{3a}$ under L-M sampling scheme, their Absolute Relative Bias (ARB) and Simulated relative standard error (Srse) are calculated for each ILRC as follows:

$$ARB\left(\bar{y}_{ka}\right) = \frac{\left|\frac{1}{10000}\sum_{t=1}^{10000} \bar{y}_{ka}^{t} - \bar{Y}_a\right|}{\bar{Y}_a} \quad \text{and} \quad Srse\left(\bar{y}_{ka}\right) = \frac{\sqrt{SMSE\left(\bar{y}_{ka}\right)}}{\bar{Y}_a}, \tag{5.1}$$

where

$$SMSE\left(\bar{y}_{ka}\right) = \frac{1}{10000}\sum_{t=1}^{10000}\left(\bar{y}_{ka}^{t} - \bar{Y}_a\right)^2, \text{for } k = 1, 2, 3 \text{ and } a = 1, \ldots, 7. \tag{5.2}$$

## 5.3 Results

We present the results of ARB and Srse of composite estimators $\bar{y}_{1a}$, $\bar{y}_{2a}$ and $\bar{y}_{3a}$ in Table 2 and Table 3. The total number of villages in Jodhpur Tehsil is 252. We take n = 25, 50, 63 and 100 i.e. samples, approximately, of 10%, 20%, 25% and 40% villages. It may be noted that a sample of 20% villages are presently adopted in TRS. Before simulation, we first examined the validity of synthetic assumption given in Eq. (3.2). The results of these are presented in Table 1. From this we note that the assumption closely meets for ILRCs (3), (4) and (6). Where as, the assumption deviate moderately for ILRC (7), and deviate considerably for ILRCs (1), (2), (5). In case of composite estimators, we estimate the weights for each small domain using Eq. (4.14) but for estimating $\bar{Y}_a$ of small domains of ILRCs (3), (4) and (6) we take average of $\hat{w}_a^*$ over these domains, being “similar”.

1. We observe from both the Tables (Table 5.2 and Table 5.3) that Srse of the composite estimator $\bar{y}_{3a}$ is around 5% (except for n = 25 where it is around 7%) and considerably smaller than Srse of $\bar{y}_{1a}$ and also of Srse of $\bar{y}_{2a}$ for the ILRCs 3, 4 and 6; where synthetic assumption

closely meets. This is some what true for ILRC 7 also but the Srse is much larger than 5% in this case, where the deviation from the synthetic assumption is not much. This is even true for other ILRCs also.

The ARB $(\bar{y}_{3a})$ is considerably small for the three ILRCs 3,4 and 6 but not the smallest when we compare it with ARB $(\bar{y}_{1\,a})$ and ARB $(\bar{y}_{2\,a})$. But the difference in the percentage of ARBs is not much.

Further for n = 50 (i.e., a sample of 20% villages that is being selected under TRS scheme) the ARB$(\bar{y}_{3a})$ varies between 0.21% to 1.99% whereas the Srse $(\bar{y}_{3a})$ is around 5% for the ILRCs 3,4 and 6.

2. We, further, note that percentage of ARBs and Srses of the three composite estimators, when weights are estimated through averaging over similar ILRCs 3, 4 and 6, is smaller than the case when weights are not averaged.

3. When we compare the Srse $(\bar{y}_{1\,a})$ and Srse $(\bar{y}_{2\,a})$ than Srse $(\bar{y}_{1\,a})$ is smaller. This is some what true for ARB $(\bar{y}_{1\,a})$, in most of the cases.

   Finally, we suggest the use of the composite estimator $\bar{y}_{3a}$ when the synthetic assumption meets. But when the synthetic assumption is not valid one should look for other types of estimators. Such as those obtained through the SICURE Model (Tikkiwal, 1993) or presented by Ghosh and Rao (1994).

Table 1: Absolute Differences (Relative) under Synthetic Assumption of Synthetic Ratio Estimator for Various ILRCs

| ILRC | $\bar{Y}_a/\bar{X}_a$ | $\left[\left\lvert \bar{Y}_a/\bar{X}_a - \bar{Y}/\bar{X} \right\rvert / \left(\bar{Y}_a/\bar{X}_a\right)\right] \times 100$ |
|---|---|---|
| 1 | 0.7304 | 18.1681 |
| 2 | 0.9416 | 8.3368 |
| 3 | 0.8596 | 0.4071 |
| 4 | 0.8663 | 0.3693 |
| 5 | 0.9667 | 10.7168 |
| 6 | 0.8816 | 2.0984 |
| 7 | 0.8917 | 3.2073 |

Note. $\bar{Y}/\bar{X} = 0.8631$

# References

[1] Ghosh, M. and Rao, J.N.K. (1994). Small Area Estimation: An Appraisal. *Statistical Science*, **91**, 55–93.

[2] Gonzalez, N.E. and Waksberg, J. (1973). Estimation of the error of synthetic estimates. Paper presented at first meeting of international association of survey statisticians, Vienna, Austria.

[3] Lahiri, D.B. (1951). A method of sample selection providing unbiased ratio estimates. *Bull. Int. Stat. Inst.* **3**, 133–40.

[4] Midzuno, H. (1952). On the sampling system with probability proportional to sum of sizes. *Ann. Inst. Stat. Math.*, **3**, 99–107.

[5] Rao, J.N.K. (2003). *Small area estimation*. Wiley Interscience.

[6] Sarndal, C.E., Swensson, B. and Wretman, J. (1992). *Model assisted survey sampling*, Springer-Verlag.

[7] Schaible, W.L. (1978). Choosing weights for composite estimators for small area statistics. Proceedings of the survey research methods section, Amer. Statist. Assoc., Washington, D.C., 741-746.

[8] Schaible, W.L., Brock, D.B., Casady, R.J. and Schnack, G.A. (1977). An empirical comparison of the simple synthetic and composite estimators for small area statistics. Proceedings of Amer. Statist. Assoc., Social Statistics Section 1017-1021.

[9] Tikkiwal, B.D. (1993). Modeling through survey data for small domains. Proceedings of International Scientific Conference on Small Area Statistics and Survey Design, Warsaw, Poland.

[10] Tikkiwal, B.D. and Tikkiwal, G. C. (1998). Crop yield and acreage statistics for small areas. *Jour. Ind. Soc. Agri. Statist.*, **51** (2&3), 265–282.

[11] Tikkiwal, G.C. and Ghiya, A. (2000). A generalized class of synthetic estimators with application of crop acreage estimation for small domains. *Biometrical Journal*, **42**, 865–876.

[12] Tikkiwal, G.C. and Ghiya, A. (2004). A generalized class of composite estimators with application to crop acreage estimation for small domains. *Statistics in Transitions*, **6**, 697–711.

[13] Tikkiwal, G.C. and Pandey, K.K. (2007). On Synthetic and Composite estimators for small area estimation under Lahiri- Midzuno Sampling Scheme. *Statistics in Transition*, **8**, 111–123.

Table 2: Percentage of Absolute Relative Bias (ARB) and Simulated Relative Standard Errors (Srse) of various composite estimators when weights are averaged over similar small domains.

| | n = 25 | | | n = 50 | | | n = 63 | | | n = 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ILRC | $\bar{y}_{1a}$ | $\bar{y}_{2a}$ | $\bar{y}_{3a}$ | $\bar{y}_{1a}$ | $\bar{y}_{2a}$ | $\bar{y}_{3a}$ | $\bar{y}_{1a}$ | $\bar{y}_{2a}$ | $\bar{y}_{3a}$ | $\bar{y}_{1a}$ | $\bar{y}_{2a}$ | $\bar{y}_{3a}$ |
| 1 | 7.02 (26.16) | 5.82 (32.32) | 18.52 (20.69) | 8.70 (23.56) | 8.26 (26.77) | 18.44 (19.50) | 10.27 (10.91) | 9.65 (23.00) | 18.39 (19.02) | 11.43 (19.41) | 10.68 (20.96) | 18.37 (18.79) |
| 2 | 13.12 (20.44) | 13.27 (27.26) | 8.15 (10.82) | 12.08 (18.56) | 11.53 (22.22) | 8.19 (9.54) | 10.63 (15.88) | 9.89 (18.37) | 8.21 (9.03) | 9.62 (13.53) | 8.95 (15.16) | 8.21 (8.75) |
| 3 | 4.90 (17.19) | 7.08 (25.64) | 0.57 (7.80) | 4.18 (14.59) | 5.55 (20.08) | 0.54 (5.38) | 3.30 (12.48) | 4.27 (16.66) | 0.52 (4,13) | 2.53 (10.63) | 3.22 (13.83) | 0.52 (3.35) |
| 4 | 2.63 (15.23) | 2.69 (22.57) | 0.22 (7.71) | 2.60 (13.07) | 2.79 (17.85) | 0.24 (5.31) | 1.98 (10.76) | 2.13 (14.33) | 0.26 (4.07) | 1.69 (9.23) | 1.84 (11.99) | 0.25 (3.29) |
| 5 | 16.44 (23.26) | 16.45 (26.96) | 10.59 (12.64) | 13.76 (18.56) | 13.08 (20.27) | 10.61 (11.62) | 12.75 (16.30) | 12.31 (17.50) | 10.62 (11.23) | 11.75 (14.57) | 11.27 (15.48) | 10.62 (11.01) |
| 6 | 5.10 (15.55) | 5.98 (22.59) | 1.85 (7.85) | 3.07 (11.81) | 2.93 (15.95) | 1.91 (5.58) | 2.57 (9.67) | 2.41 (12.83) | 1.94 (4.46) | 2.32 (8.05) | 2.18 (10.48) | 1.95 (3.78) |
| 7 | 5.68 (26.66) | 3.25 (30.80) | 20.49 (22.51) | 10.62 (22.30) | 9.57 (24.35) | 20.46 (21.44) | 11.73 (20.12) | 10.61 (21.29) | 20.43 (21.01) | 12.36 (18.85) | 11.21 (19.46) | 20.43 (20.81) |

Note : The percentage of Srse is given in the brackets.

Table 3: Percentage of Absolute Relative Bias (ARB) and Simulated Relative Standard Errors (Srse) of various composite estimators when weights are not averaged over similar small domains.

| | n = 25 | | | n = 50 | | | n = 63 | | | n = 100 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ILRC | $\bar{y}_{1a}$ | $\bar{y}_{2a}$ | $\bar{y}_{3a}$ | $\bar{y}_{1a}$ | $\bar{y}_{2a}$ | $\bar{y}_{3a}$ | $\bar{y}_{1a}$ | $\bar{y}_{2a}$ | $\bar{y}_{3a}$ | $\bar{y}_{1a}$ | $\bar{y}_{2a}$ | $\bar{y}_{3a}$ |
| 1 | 7.02 (26.16) | 5.82 (32.32) | 18.52 (20.69) | 8.70 (23.56) | 8.26 (26.77) | 18.44 (19.50) | 10.27 (20.91) | 9.65 (23.00) | 18.39 (19.02) | 11.43 (19.41) | 10.68 (20.96) | 18.37 (18.79) |
| 2 | 13.12 (20.44) | 13.27 (27.26) | 8.15 (10.82) | 12.08 (18.56) | 11.53 (22.22) | 8.19 (9.54) | 10.63 (15.88) | 9.89 (18.37) | 8.21 (9.03) | 9.62 (13.53) | 8.95 (15.16) | 8.21 (8.75) |
| 3 | 8.20 (21.59) | 9.38 (27.95) | 0.59 (7.81) | 6.07 (17.65) | 6.11 (20.64) | 0.56 (5.39) | 4.40 (14.00) | 4.26 (16.11) | 0.54 (4.14) | 3.18 (11.32) | 3.16 (13.03) | 0.53 (3.35) |
| 4 | 4.56 (15.68) | 3.33 (23.87) | 0.17 (7.74) | 3.91 (13.96) | 2.81 (18.87) | 0.21 (5.33) | 2.70 (11.26) | 2.04 (15.01) | 0.23 (4.08) | 2.19 (9.52) | 1.80 (12.51) | 0.24 (3.29) |
| 5 | 16.44 (23.26) | 16.45 (26.96) | 10.59 (12.64) | 13.76 (18.56) | 13.08 (20.27) | 10.61 (11.62) | 12.75 (16.30) | 12.31 (17.50) | 10.62 (11.23) | 11.75 (14.57) | 11.27 (15.48) | 10.62 (11.01) |
| 6 | 7.42 (18.33) | 6.81 (24.17) | 1.98 (7.82) | 4.18 (12.94) | 2.91 (16.88) | 1.99 (5.57) | 3.36 (10.55) | 2.35 (13.70) | 2.00 (4.46) | 2.73 (8.75) | 2.10 (11.6) | 1.99 (3.78) |
| 7 | 5.68 (26.66) | 3.25 (30.80) | 20.49 (22.51) | 10.62 (22.30) | 9.57 (24.35) | 20.46 (21.44) | 11.73 (20.12) | 10.61 (21.29) | 20.43 (20.01) | 12.36 (18.85) | 11.21 (19.46) | 20.43 (20.81) |

Note : The percentage of Srse is given in the brackets.