International Journal of Statistics and Systems
ISSN 0973 - 2675 Volume 5, Number 3 (2010), pp. 319–333

http://www.ripublication.com/ijss.htm

The Generalized Class of Composite Method of Estimation for Crop Acreage in Small Domain

Krishan K. Pandey[1] and G.C. Tikkiwal[2]

[1] *College of Management & Economic Studies, University of Petroleum & Energy Studies, Dehradun- 248007, India*
[2]*Department of Mathematics & Statistics, J.N.V. University, Jodhpur, Rajasthan -342011, India*

Abstract

This paper defines and discusses a generalized class of composite estimators for small domains, using auxiliary information, under systematic sampling scheme. The generalized class of composite estimators, among others, includes a number of direct, synthetic and composite estimators. Further, it demonstrates the use of the estimators belonging to the generalized class for estimating crop acreage for small domains and also compares their relative performance with the corresponding direct and synthetic estimators, through a simulation study.

Key words: Inspector Land Revenue Circles (ILRCs), Timely Reporting Scheme (TRS), Absolute Relative Bias (ARB), Simulated relative standard error (Srse), Simulation-Cum-Regression (SICURE) model.

Introduction

According to Purcell and Kish (1979), an area is said to be small area, if it accounts for $1/10^4$ to $1/10^{th}$ of a population. A small area may be a geographical area as well as a socioeconomic classification of the population. The common feature of small area estimation problem is that when large-scale sample survey are designed to produce reliable estimates at the national or state level; generally they do not provide estimates of adequate precision at lower levels like District, Tehsil / County, and Inspector land Revenue Circle. This is because the sample sizes at the lower level are generally insufficient to provide reliable estimates using traditional estimators. Therefore, the need was felt to develop alternative estimators to provide small area statistics using the data already collected through large-scale surveys. The traditional design based and alternative estimators are also termed, in the literature of small area estimation, respectively as direct and indirect estimators.

It is evident that at some point, as the sample size in a small area increases, a direct estimator becomes more desirable than a synthetic one. This is true whether or not the sample was designed to produce estimates for small areas.

Gonzalez (1973), Ghangurde and Singh (1977, 78), Tikkiwal, G.C. and Pandey, K.K. (2006, 07), among others have studed the synthetic estimator based on auxiliary variables viz. the ratio synthetic estimator. These studies show that synthetic estimators provide reliable estimates to some extent.

Gonzalez, Waksberg (1973) and Schaible, Brock, Casady, Schnack (1977) studied errors of synthetic and direct estimates for standard metropolitan statistical areas and Counties of United State of America. The authors of both the papers conclude that when in small domains sample sizes are relatively small the synthetic estimator outperforms 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 choose one over the other. In general, a composite estimator may be defined as follows:

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

Where $\bar{y}_{d,a}$ is a direct estimator and $\bar{y}_{syn,a}$ is a synthetic estimator of $\bar{Y}_a$, the population mean of small area 'a' and $w_a (a = 1, ..., A)$ are suitably chosen weights. Here $\bar{y}_{c,a}$ is a model dependent estimator as it is a combination of design-based estimator $\bar{y}_{d,a}$ and model dependent design based estimator $\bar{y}_{syn,a}$ [Cf. Sarndal (1984)]. The optimal 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

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

It can be seen in many practical situation $E\left(\bar{y}_{d,a} - \bar{Y}_a\right)\left(\bar{y}_{syn,a} - \bar{Y}_a\right)$ is small relative to $MSE\left(\bar{y}_{syn,a}\right)$, $w_a^{'}$ becomes more manageable. In this case $w_a^{'}$ may be approximated by

$$w_a^{'} = \frac{MSE\left(\bar{y}_{syn,a}\right)}{MSE\left(\bar{y}_{d,a}\right) + MSE\left(\bar{y}_{syn,a}\right)} = \frac{1}{1 + F_a}, \text{ Where } F_a = \frac{MSE\left(\bar{y}_{d,a}\right)}{MSE\left(\bar{y}_{syn,a}\right)}$$

The weights w_a^{*} can be estimated by replacing the mean square errors of $\bar{y}_{d,a}$ and $\bar{y}_{syn,a}$ by their usual estimates, i.e. $\hat{w}_a^{*} = \dfrac{\overline{mse}(\bar{y}_{syn,r,a})}{v(\bar{y}_{d,a}) + \overline{mse}(\bar{y}_{syn,r,a})}$

Further Schaible (1978) proposed a modified "average" weighting scheme based on several weighting variables under their different models.

This paper discusses the generalized class of composite estimators, using auxiliary information under systematic sampling scheme. The systematic sampling scheme, being operationally more convenient in practice, is often used in large – scale field surveys under multistage design. In such survey, like crop acreage surveys in India, ultimate stage of smapling units like villages / households / agricultural fields etc. are selected by systematic smapling scheme. Systematic sampling scheme, apart from operationally more convinient, provides more efficient estimators under certain conditions [Cf. Cochran (1977), Sukhatme et al (1984), Madow (1946) & Osborne, J.G. (1942)].

Formulation of the problem & Notations

Let us suppose that we have a finite population U = (1,…, i ,….N) which is divided into 'A' non-overlapping small areas U_a of size N_a (a = 1,…., A) for which estimates are required. Let the characteristic under study is denoted by 'y' and also assume that the auxiliary information is available which is denoted by 'x'. Suppose the population units in small area 'a' are numbered 1 to N_a i.e. $U_a = (1,......., N_a)$ and n_a units are to be selected by systematic sampling scheme. A systematic sample of size n_a is selected from each small area 'a', (a = 1,….., A) either (i) by linear systematic sampling scheme, (when $N_a = n_a k_a$, k_a being an integer) or (ii) by circular systematic sampling scheme, (when $N_a \neq n_a k_a$). Consequently,

$$\sum_{a=1}^{A} N_a = N \text{ and } \sum_{a=1} n_a = n \ ,$$

Further, the various population and sample means for characteristic Z = X, Y can be defined as:

$\bar{Z}$ = Mean of the population based on N observations.

$\bar{Z}_a$ = Population mean of domain 'a' based on N_a observations.

$\bar{z}$ = Mean of the sample 's' based on n observations.

$\bar{z}_a$ = Sample mean of domain 'a' based on n_a observations.

Case (i): For the case $N_a = n_a k_a$ i.e. for linear systematic sampling scheme, arrange the population units into $n_a k_a$ arrays and select a random number, say, i between 1 and k_a then every k_a^{th} unit thereafter. So the sample consist n_a units from $N_a (= n_a k_a)$ units, and the sample is $\{i, i + k_a, ..., i + (n_a - 1)k_a\}$. The number i, is called random start and k_a is the sampling interval. Further, let z_{aij} denotes the value of the auxiliary variate and characteristic under study of $z = x, y$ respectively for the jth unit of the ith sample bearing serial number i +(j-1) k_a , i =1,......., k_a ; j =1,......., n_a . Therefore,

$\bar{Z}_a = \frac{1}{N_a} \sum_i \sum_j z_{aij}$, $\bar{z}_{ai.} = \frac{1}{n_a} \sum_{j=1}^{n_a} z_{aij}$ = mean of the ith systematic sample of size n_a, selected from small area 'a' of characteristic $z = x, y$.

Various mean squares and coefficient of variations of subpopulation, 'U_a' for characteristic $z = x, y$ is denoted by

$$S_{z_a}^2 = \frac{1}{k_a - 1}\sum_{i=1}^{k_a}(\bar{z}_{ai.} - \bar{Z}_a)^2 \ , \ C_{z_a} = \frac{S_{z_a}}{\bar{Z}_a}$$

The coefficient of covariance between X and Y is denoted by

$$C_{x_a y_a} = \frac{S_{x_a y_a}}{\bar{X}_a \bar{Y}_a} \text{ , where } S_{x_a y_a} = \frac{1}{k_a - 1}\sum_{i=1}^{k_a}(\bar{y}_{ai.} - \bar{Y}_a)(\bar{x}_{ai.} - \bar{X}_a)$$

Case (ii): For the case $N_a \neq n_a k_a$ i.e. for those small areas where N_a / n_a is not an integer but k_a is the integer nearest to N_a / n_a, Lahiri (1954) suggested to use circular systematic sampling scheme. Here in this case a random number is chosen from 1 to N_a and the units corresponding to this random number are chosen as the random start. There after every k_a^{th} unit is chosen in a cyclic manner till a sample of n_a units is selected. Thus if i is a number selected at random from 1 to N_a, the sample consists of units corresponding to these numbers are

$\{i + (j-1)k_a\}$ if $i + (j-1)k_a \leq N_a$

$\{i + (j-1)k_a - N_a\}$ if $i + (j-1)k_a > N_a$, j =1,2,…….., n_a

Further, suppose that z_{aij} denotes, the value of the auxiliary variate and characteristic under study of $z = x, y$ respectively for the jth unit of the ith sample bearing the number $\{i + (j-1)k_a\}$ or $\{i + (j-1)k_a - N_a\}$ as the case may be for $j = 1, 2, ..., n_a$. The various mean squares and coefficient of variations of sub population 'U_a' for characteristics $z = x, y$ is given as below:

$$S_{1z_a}^2 = \frac{1}{N_a - 1}\sum_{i=1}^{N_a}(\bar{z}_{ai.} - \bar{Z}_a)^2 \ , \ C_{1z_a}^2 = \frac{S_{1z_a}^2}{\bar{Z}_a^2}$$

The coefficient of covariance between X and Y is denoted by

$$C_{1\,x_a y_a} = \frac{S_{1x_a y_a}}{\bar{X}_a \bar{Y}_a}\text{, where } S_{1x_a y_a} = \frac{1}{N_a - 1}\sum_{i=1}^{N_a}(\bar{y}_{ai.} - \bar{Y}_a)^2(\bar{x}_{ai.} - \bar{X}_a)^2$$

A Generalized class of composite estimators

Following Tikkiwal and Pandey (2007), we, in this section, define a generalized class of composite estimators of population mean $\bar{Y}_a$ based on auxiliary variable 'x' under systematic sampling scheme, as follows;

$$\bar{y}_{c,g} = w_a \bar{y}_{ai.}\left(\frac{\bar{x}_{ai.}}{\bar{X}_a}\right)^{\alpha} + (1 - w_a)\,\bar{y}_w\left(\frac{\bar{x}_w}{\bar{X}_a}\right)^{\beta} \qquad \text{… (3.1)}$$

Where α and β are suitably chosen constants. The estimator $\bar{y}_{c,g}$ is a weighted sum of generalized direct estimator $\bar{y}_{d,a}^{G}$, an estimator which is expected to perform well for fairly big range of values of α [Srivastava (1967)], and the generalized synthetic estimators $\bar{y}_{s,a}$ given by Tikkiwal & Ghiya (2004) defined as below:

$$\bar{y}_{d,a}^{G} = \bar{y}_{ai.}\left(\frac{\bar{x}_{ai.}}{\bar{X}_{a}}\right)^{\alpha} \quad \ldots (3.2)$$

and

$$\bar{y}_{s,a} = \bar{y}_{w}\left(\frac{\bar{x}_{w}}{\bar{X}_{a}}\right)^{\beta} \quad \ldots (3.3)$$

The proposed generalized class of composite estimators includes a number of direct, synthetic and composite estimators as special cases. The Table (3.1) shows a list of such estimators with corresponding choice of values of the different constants.

Table: 3.1: Various Direct & Indirect Estimators as special case of the generalized class of composite estimators.

No.	**Estimator**	w_a	$(1-w_a)$	α	β
1.	Simple Direct ($\bar{y}_{ai\bullet}$)	1	0	0	-
2.	Simple Synthetic ($\bar{y}_{w}$)	0	1	-	0
3.	Simple Ratio ($(\bar{y}_{ai\bullet}/\bar{x}_{ai\bullet})\bar{X}_{a}$)	1	0	-1	-
4.	Ratio Synthetic ($(\bar{y}_{w}/\bar{x}_{w})\bar{X}_{a}$)	0	1	-	-1
5.	Simple Product ($(\bar{x}_{ai\bullet}/\bar{X}_{a})\bar{y}_{ai\bullet}$)	1	0	1	-
6.	Product Synthetic ($(\bar{y}_{w}/\bar{X}_{a})\bar{x}_{w}$)	0	1	-	1
7.	Composite: combining simple direct with simple synthetic ($w_a\bar{y}_{ai\bullet}+(1-w_a)\bar{y}_{w}$)	w_a	$(1-w_a)$	0	0
8.	Composite: combining simple direct with ratio synthetic ($w_a\bar{y}_{ai\bullet}+(1-w_a)\left(\frac{\bar{y}_{w}}{\bar{x}_{w}}\right)\bar{X}_{a}$)	w_a	$(1-w_a)$	0	-1
9.	Composite: combining simple ratio with ratio synthetic ($w_a\left(\frac{\bar{y}_{ai\bullet}}{\bar{x}_{ai\bullet}}\right)\bar{X}_{a}+(1-w_a)\left(\frac{\bar{y}_{w}}{\bar{x}_{w}}\right)\bar{X}_{a}$)	w_a	$(1-w_a)$	-1	-1

The estimator given in (3.1) perform well under the following condition,

$$\bar{Y}_{a}\left(\bar{X}_{a}\right)^{\beta} \cong \bar{Y}(\bar{X})^{\beta} \quad \ldots (3.4)$$

It is noted that the composite estimator $\bar{y}_{c,g}$ is consistent; if the condition given in (3.4) is satisfied.

Design Bias and Mean Square Error

The design bias of the composite estimator $\bar{y}_{c,g}$, is given by

$$B\left(\bar{y}_{c,g}\right)=E\left(\bar{y}_{c,g}\right)-\bar{Y}_a=w_a\,B\left(\bar{y}^G_{d,a}\right)+(1-w_a)\,B\left(\bar{y}_{s,a}\right) \quad \ldots (4.1)$$

Where $B\left(\bar{y}^G_{d,a}\right)$ for the case of linear systematic sampling schemen, is given as follows;

$$B\left(\bar{y}^G_{d,a}\right)=\bar{Y}_a\left[\frac{(k_a-1)}{k_a}\left\{\alpha C_{x_a y_a}+\frac{\alpha(\alpha-1)}{2}C^2_{x_a}\right\}\right] \quad \ldots (4.2)$$

and for the case of circular systematic sampling scheme the expression is as follows;

$$B\left(\bar{y}^G_{d,a}\right)=\bar{Y}_a\left[\frac{(N_a-1)}{N_a}\left\{\alpha C_{1x_a y_a}+\frac{\alpha(\alpha-1)}{2}C^2_{1x_a}\right\}\right] \quad \ldots (4.3)$$

The bias of the estimator $\bar{y}_{s,a}$ is given by

$$B\left(\bar{y}_{s,a}\right)=\bar{Y}\left(\frac{\bar{x}}{\bar{X}_a}\right)^{\beta}\left[1+\beta\left\{\sum_a{}' p_a^2\frac{(k_a-1)}{k_a}\frac{S_{y_a x_a}}{\bar{X}_a\bar{Y}_a}+\sum{}'' p_a^2\frac{(N_a-1)}{N_a}\frac{S_{1y_a x_a}}{\bar{X}_a\bar{Y}_a}\right\}\right.$$
$$\left.+\frac{\beta(\beta-1)}{2}\left\{\sum_a{}' p_a^2\frac{(k_a-1)}{k_a}\frac{S^2_{x_a}}{\bar{X}^2_a}+\sum{}'' p_a^2\frac{(N_a-1)}{N_a}\frac{S^2_{1x_a}}{\bar{X}^2_a}\right\}\right]-\bar{Y}_a \quad \ldots (4.4)$$

Where $\sum{}'$ denotes the summation over those small areas, where $N_a=n_a k_a$, and $\sum{}''$ denotes summation over those small areas where $N_a\neq n_a k_a$. Further mean square error of $\bar{y}_{c,g}$ is given by

$$MSE\left(\bar{y}_{c,g}\right)=w_a^2 MSE\left(\bar{y}^G_{d,a}\right)+(1-w_a)^2 MSE\left(\bar{y}_{s,a}\right)+2w_a(1-w_a)E\left(\bar{y}^G_{d,a}-\bar{Y}_a\right)\left(\bar{y}_{s,a}-\bar{Y}_a\right)$$

Now under the condition that the covariance term $E(\bar{y}^G_{d,a}-\bar{Y}_a)(\bar{y}_{s,a}-\bar{Y}_a)$ is small relative to $MSE(\bar{y}_{s,a})$, as discussed in Section 1, the above expression of $MSE\left(\bar{y}_{c,g}\right)$, can be written as $MSE\left(\bar{y}_{c,g}\right)=w_a^{*2} MSE\left(\bar{y}^G_{d,a}\right)+\left(1-w_a^*\right)^2 MSE\left(\bar{y}_{s,a}\right)$ . . . (4.5)

Where expression of $MSE(\bar{y}^G_{d,a})$ is given under linear & circular systematic sampling scheme, respectively as follows;

$$MSE(\bar{y}^G_{d,a})=\bar{Y}_a^2\,\frac{(k_a-1)}{k_a}\left[C^2_{y_a}+\alpha^2 C^2_{x_a}+2\alpha C_{x_a y_a}\right] \quad . . . (4.6)$$

and $MSE(\bar{y}^G_{d,a})=\bar{Y}_a^2\,\frac{(N_a-1)}{N_a}\left[C^2_{1y_a}+\alpha^2 C^2_{1x_a}+2\alpha C_{1x_a y_a}\right]$ . . . (4.7)

Which is minimum if $\alpha = -\frac{C_{x_a y_a}}{C_{x_a}^2}$ [**C**f. Srivastava (1967)]. Further, design bias and mean square error of generalized synthetic estimator $\bar{y}_{s,a}$, is given as follows;

$$B(\bar{y}_{s,a}) = \bar{Y}_a \left[\beta \left\{ \sum_a^{'} p_a^2 \frac{(k_a - 1)}{k_a} \frac{S_{y_a x_a}}{\bar{X}\bar{Y}} + \sum_a^{''} p_a^2 \frac{(N_a - 1)}{N_a} \frac{S_{1x_a y_a}}{\bar{X}\bar{Y}} \right\} + \frac{\beta(\beta - 1)}{2} \left\{ \sum_a^{'} p_a^2 \frac{(k_a - 1)}{k_a} \frac{S_{x_a}^2}{\bar{X}^2} + \sum_a^{''} p_a^2 \frac{(N_a - 1)}{N_a} \frac{S_{1x_a}^2}{\bar{X}^2} \right\} \right] \quad \text{... (4.8)}$$

And

$$MSE(\bar{y}_{s,a}) = E(\bar{y}_{s,a} - \bar{Y}_a)^2 = \bar{Y}^2 \left(\frac{\bar{X}}{\bar{X}_a} \right)^{2\beta} \left[1 + \sum_a^{'} p_a^2 \frac{(k_a - 1)}{k_a} \frac{S_{y_a}^2}{\bar{Y}^2} + \sum_a^{''} p_a^2 \frac{(N_a - 1)}{N_a} \frac{S_{1y_a}^2}{\bar{Y}^2} \right.$$
$$+ (2\beta - 1)\beta \left\{ \sum_a^{'} p_a^2 \frac{(k_a - 1)}{k_a} \frac{S_{x_a}^2}{\bar{X}^2} + \sum_a^{''} p_a^2 \frac{(N_a - 1)}{N_a} \frac{S_{1x_a}^2}{\bar{X}^2} \right\}$$
$$\left. + 4\beta \left\{ \sum_a^{'} p_a^2 \frac{(k_a - 1)}{k_a} \frac{S_{y_a x_a}}{\bar{X}\bar{Y}} + \sum_a^{''} p_a^2 \frac{(N_a - 1)}{N_a} \frac{S_{1x_a y_a}}{\bar{X}\bar{Y}} \right\} \right]$$
$$+ \bar{Y}_a^2 - 2\bar{Y}_a \bar{Y} \left(\frac{\bar{X}}{\bar{X}_a} \right)^{\beta} \left[1 + \beta \left\{ \sum_a^{'} p_a^2 \frac{(k_a - 1)}{k_a} \frac{S_{y_a x_a}}{\bar{X}\bar{Y}} + \sum_a^{''} p_a^2 \frac{(N_a - 1)}{N_a} \frac{S_{1y_a x_a}}{\bar{X}\bar{Y}} \right\} \right.$$
$$\left. + \frac{\beta(\beta - 1)}{2} \left\{ \sum_a^{'} p_a^2 \frac{(k_a - 1)}{k_a} \frac{S_{x_a}^2}{\bar{X}^2} + \sum_a^{''} p_a^2 \frac{(N_a - 1)}{N_a} \frac{S_{1x_a}^2}{\bar{X}^2} \right\} \right] \quad \text{… (4.9)}$$

The suitable value of β is the one for $MSE(\bar{y}_{s,a})$ is minimum. So minimizing the $MSE(\bar{y}_{s,a})$ with respect to β, gives simplified expression of β, which is given as follows;

$$\beta = \frac{-\left\{ \sum_a^{'} p_a^2 \frac{(k_a - 1)}{k_a} \frac{S_{y_a x_a}}{\bar{X}\bar{Y}} + \sum_a^{''} p_a^2 \frac{(N_a - 1)}{N_a} \frac{S_{1x_a y_a}}{\bar{X}\bar{Y}} \right\}}{\left\{ \sum_a^{'} p_a^2 \frac{(k_a - 1)}{k_a} \frac{S_{x_a}^2}{\bar{X}^2} + \sum_a^{''} p_a^2 \frac{(N_a - 1)}{N_a} \frac{S_{1x_a}^2}{\bar{X}^2} \right\}} \quad \text{… (4.10)}$$

Estimation of MSE of composite estimator under systematic sampling scheme

Since a systematic sample can be regarded as a random selection of one cluster, it is not possible to give an unbiased or even consistent estimator of the design variances of $\bar{y}_{ai.}$ or $\bar{x}_{ai.}$. A common practice in applied survey work is to regard the sample as random and for lack of knowing what else to do, estimate the variance using simple random sample formulae. Unfortunately, if followed indiscriminately this practice can lead to badly biased estimators and incorrect inferences concerning the population parameters of interest. Wolter (1984) investigated several biased estimators of variances with a goal of providing some guidance about when a given estimator may be more appropriate than other estimators. The criterion to judge the various estimators on the basis of their bias, their mean square error, and proportion of confidence interval formed using the variance estimators which contain the true population parameter of interest. This study suggests the use of biased but simple estimator v_{2y} for $V(\bar{y}_{ai}.)$, when sample size is very small for both the situations viz., when $N_a = n_a k_a$ and $N_a \neq n_a k_a$. The expression of v_{2y} is given as follows;

$$v_{2y} = (1-f)\left(\frac{1}{n_a}\right)\sum_{j=2}^{n_a}\frac{a_{ij}^2}{2(n_a-1)} \qquad \dots (5.1)$$

$$\left.\begin{aligned} &\text{where} \quad a_{ij} = \Delta y_{ij} = y_{ij} - y_{i,\,j-1} \\ &\text{and} \quad f = \frac{n_a}{N_a} \end{aligned}\right\} \qquad \dots (5.2)$$

Similarly estimate of $V(\bar{x}_{ai.})$ is given by v_{2x}, where

$$v_{2x} = (1-f)\left(\frac{1}{n_a}\right)\sum_{j=2}^{n_a}\frac{b_{ij}^2}{2(n_a-1)} \qquad \dots (5.3)$$

$$\left.\begin{aligned} &\text{where} \quad b_{ij} = \Delta x_{ij} = x_{ij} - x_{i,\,j-1} \\ &\text{and} \quad f = \frac{n_a}{N_a} \end{aligned}\right\} \qquad \dots (5.4)$$

We note that above estimators v_{2y} and v_{2x} are based on overlapping differences of Δy_{ij} & Δx_{ij} respectively. Further, the estimate of covariance term between $\bar{y}_{ai.}$ and $\bar{x}_{ai.}$, given by Swain (1964), is

$$C\hat{o}v(\bar{y}_{ai}., \bar{x}_{ai}.) = r\sqrt{v_{2y}v_{2x}} \qquad \dots (5.5)$$

Where r is correlation coefficient between x and y observations based on the sample of size n_a.

Estimation of mean square error of direct estimator

Following Srivastava (1967), the generalized class of direct estimators of $\bar{Y}_a$ under systematic sampling scheme is $\bar{y}_{d,a}^G = \bar{y}_{ai.}\left(\frac{\bar{x}_{ai.}}{\bar{X}_a}\right)^{\alpha}$. Its mean square under case (i) is

$$MSE(\bar{y}_{d,a}^G) = V(\bar{y}_{ai}.) + \alpha^2 R_a^2 V(\bar{x}_{ai}.) + 2\alpha R_a Cov(\bar{y}_{ai}., \bar{x}_{ai}.) \qquad \dots (5.6)$$

Where $R_a = \frac{\bar{Y}_a}{\bar{X}_a}$. Thus a consistent estimator of $MSE(\bar{y}_{d,a}^{G})$ is given by

$$mse(\bar{y}_{d,a}^{g}) = v_{2y} + \alpha^2 r_a^2 v_{2x} + 2\alpha r_a r\sqrt{v_{2y} v_{2x}} \quad \ldots (5.7)$$

Where $r_a = \frac{\bar{y}_a}{\bar{x}_a}$ is the ratio of sample means. It is also observed that the mean square error for direct estimator in case of circular systematic sampling is given by

$$MSE\left(\bar{y}_{d,a}^{G}\right)_c = V(\bar{y}_{ai}.)_c + \alpha^2 R_a^2 V(\bar{x}_{ai}.)_c + 2\alpha R_a Cov(\bar{y}_{ai}., \bar{x}_{ai}.)_c \quad \ldots (5.8)$$

Thus consistent estimator of $MSE\left(\bar{y}_{d,a}^{G}\right)_c$ is given by

$$mse(\bar{y}_{d,a}^{g})_c = v_{2y}^{'} + \alpha^2 r_a^2 v_{2x}^{'} + 2\alpha r_a r\sqrt{v_{2y}^{'} v_{2x}^{'}} \quad \ldots (5.9)$$

Where $v_{2y}^{'}$ and $v_{2x}^{'}$ are the estimates of variances of $V(\bar{y}_{ai}.)_c$ and $V(\bar{x}_{ai}.)_c$ respectively in case of circular systematic sampling design. To be calculate similarly as of v_{2x} and v_{2y}.

Estimation of mean square error of synthetic estimator

The expression for the Mean Square Error given in (4.9), can be approximated under the synthetic condition given in (3.4) as follows;

$$MSE(\bar{y}_{s,a}) = \sum_a^{'} p_a^2 V(\bar{y}_{ai}.) + \sum_a^{''} p_a^2 V(\bar{y}_{ai}.)_c + \beta^2 R_a^2 \left\{\sum_a^{'} p_a^2 V(\bar{x}_{ai}.) + \sum_a^{''} p_a^2 V(\bar{x}_{ai}.)_c\right\}$$

$$+ 2\beta R_a \left\{\sum_a^{'} p_a^2 Cov(\bar{y}_{ai}., \bar{x}_{ai}.) + \sum_a^{''} p_a^2 Cov(\bar{y}_{ai}., \bar{x}_{ai}.)_c\right\} \quad \ldots (5.10)$$

Thus a consistent estimator of $MSE(\bar{y}_{s,a})$ is given by

$$mse(\bar{y}_{s,a}) = \left\{\sum_a^{'} p_a^2 v_{2y} + \sum_a^{''} p_a^2 v_{2y}^{'}\right\} + \beta^2 r_a^2 \left\{\sum_a^{'} p_a^2 v_{2x} + \sum_a^{''} p_a^2 v_{2x}^{'}\right\}$$

$$+ 2\beta r_a \left\{\sum_a^{'} p_a^2 r\sqrt{v_{2y} v_{2x}} + \sum_a^{''} p_a^2 r\sqrt{v_{2y}^{'} v_{2x}^{'}}\right\} \quad \ldots (5.11)$$

Where $r_a = \frac{\bar{y}_a}{\bar{x}_a}$ is the ratio of sample means.

Crop Acreage Estimation for Small Domains - A Simulation Study

This section demonstrates the use of the generalized class of composite estimator $\bar{y}_{c,g}$, along with the various direct and indirect estimators to obtain crop acreage estimates for small domains and also compare their relative performance through a simulation study. This is done by taking up the state of Rajasthan, one of the states in India, for case study [Cf. Tikkiwal & Ghiya (2000)].

Existing methodology for estimation

In order to improve timelines and quality of crop acreage statistics, Timely Reporting Scheme (TRS) is used by most of the States of India. The TRS has the objective of providing quick and reliable estimates of crop acreage statistics and there-by production of the principle crops (i.e. Jowar, Bajra, Maize etc.) 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, selected by stratified linear systematic sampling design taking Tehsil (a sub-division of the District) as a stratum. These statistics are further used to provide state level estimates using direct estimators viz. unbiased (based on sample mean) and ratio estimators.

Details of the Simulation Study

For collection of revenue and 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. The various information regarding the ILRCs of Jodhpur Tehsil are provided in the Table 6.2.1.

Table 6.2.1: Total Area (Irrigated and Unirrigated) under Bajara Crop in Inspector Land Revenue Circles (ILRCs) of Jodhpur Tehsil for Agricultural seasons 1992-93 and 1993-94.

S.No	ILRCs of Jodhpur Tehsil	No. of villages in ILRC	Total area(Irr.+U.Irr.) under the crop Bajra in 1992-93	Total area(Irr.+U.Irr.) under the crop Bajra in 1993-94
1	Jodhpur (1)	29	7799.5899	5696.5000
2	Keru (2)	44	21209.5880	15699.6656
3	Dhundhada (3)	32	19019.0288	16476.4863
4	Bisalpur (4)	30	15153.9248	14269.0000
5	Luni (5)	33	19570.1323	16821.4508
6	Dhava (6)	40	25940.0979	25075.5000
7	Jajawal Kalan (7)	44	18007.4120	15875.0000
	Total	**252**	**126699.7737**	**109913.6027**

Below the list of all those estimators, whose relative performance is to be assessed for estimating population total T_a of small domain a for 'a' = 1, 2, …7.

(1) Direct ratio estimator $\hat{T}_{1,a} = N_a\, \bar{y}_{d,r,a} = N_a \left(\dfrac{\bar{y}_{ai.}}{\bar{x}_{ai.}} \right) \bar{X}_a$

Where $\bar{y}_{ai.} = \dfrac{1}{n_a} \sum_{j=1}^{n_a} y_{aij}$; and $\bar{x}_{ai.} = \dfrac{1}{n_a} \sum_{j=1}^{n_a} x_{aij}$

(2) Ratio synthetic estimator $\hat{T}_{2,a} = N_a\, \bar{y}_{s,r,a} = N_a \left(\dfrac{\bar{y}_w}{\bar{x}_w} \right) \bar{X}_a$

Where, $\bar{y}_w = \sum' p_a \bar{y}_{ai.} + \sum'' p_a \bar{y}_{ai.}$

And, $\bar{x}_w = \sum' p_a \bar{x}_{ai.} + \sum'' p_a \bar{x}_{ai.}$

(3) Composite estimator (a weighted average of simple direct and synthetic ratio estimators)

$$\hat{T}_{3,a} = N_a \left[w_a \bar{y}_{d,s,a} + (1-w_a) \bar{y}_{s,r,a} \right] = N_a \left[w_a \bar{y}_{ai.} + (1-w_a) \bar{y}_w \left(\frac{\bar{X}_a}{\bar{x}_w} \right) \right]$$

(4) Composite estimator (a weighted average of direct ratio and synthetic ratio estimators)

$$\hat{T}_{4,a} = N_a \left[w_a \bar{y}_{d,r,a} + (1-w_a) \bar{y}_{s,r,a} \right] = N_a \left[w_a \left(\frac{\bar{y}_{ai.}}{\bar{x}_{ai.}} \right) \bar{X}_a + (1-w_a) \left(\frac{\bar{y}_w}{\bar{x}_w} \right) \bar{X}_a \right]$$

Prior to simulation, we examine the condition of generalized synthetic and synthetic ratio estimators as given in Eq. (3.4). These results are presented in following tables 6.2.2 & 6.2.3 respectively. We note that both the above conditions meet for ILRCs (3), (5), (7) deviate moderality for ILRCs (4) & (6) and deviate considerably for ILRC (7).

Table 6.2.2: Absolute Differences (Relative) under Synthetic Assumption of Synthetic Ratio Estimator for Various ILRCs.

ILRC	$\bar{Y}_a / \bar{X}_a$	$\bar{Y} / \bar{X}$	$\left[\left| (\bar{Y}_a / \bar{X}_a) - (\bar{Y} / \bar{X}) \right| \div (\bar{Y}_a / \bar{X}_a) \right] \times 100$
(1)	**0.73036**	**0.86751**	**18.17**
(2)	**0.7402**	**0.86751**	**17.19**
(3)	**0.8663**	**0.86751**	**0.13**
(4)	**0.9416**	**0.86751**	**7.86**
(5)	**0.8595**	**0.86751**	**0.91**
(6)	**0.9666**	**0.86751**	**10.25**
(7)	**0.8815**	**0.86751**	**1.58**

Table 6.2.3: Absolute Differences under Synthetic Assumption of Generalized Synthetic Estimator for Various ILRCs.

ILRC	$\bar{Y}_a(\bar{X}_a)^\beta$	$\bar{Y}(\bar{X})^\beta$	$\left[\left\|\{\bar{Y}_a(\bar{X}_a)^\beta - \bar{Y}(\bar{X})^\beta\}\right\| \div \bar{Y}_a(\bar{X}_a)^\beta\right] \times 100$
(1)	**3.31157**	**4.6578**	**40.65232**
(2)	**2.11349**	**2.4947**	**18.03699**
(3)	**0.77584**	**0.7791**	**0.42019**
(4)	**1.23143**	**1.1343**	**7.887578**
(5)	**0.8136**	**0.82231**	**1.070551**
(6)	**0.14251**	**0.13789**	**3.241878**
(7)	**2.40008**	**2.44412**	**1.834939**

Now for simulation study, [Length of simulation is estimated with the help of concept discussed by Whitt, W. (1989) & Murphy, K.E. Carter, C.M. & Wolfe, L. H. (2001), based on the steady state condition] taking villages as sampling units, 500 independent systematic samples each of size 25, 50, 63, 76 and 88 are selected by the procedure described in section 2 from the population of 252 villages of Jodhpur Tehsil. That is selecting approximately 10 percent, 20 percent, 25 percent, 30 percent and 35 percent villages independly form each ILRC. For each small area estimator under consideration and for each sample size we compute Absolute Relative Bias (ARB) and Average Square Error (ASE), as defined below.

$$ARB(\hat{T}_{k,a}) = \frac{\left|\frac{1}{500}\sum_{s=1}^{500}\hat{T}_{k,a}^{s} - T_a\right|}{T_a} \times 100 \qquad \text{... (6.1)}$$

and $$Srse(\hat{T}_{k,a}) = \frac{\sqrt{ASE(\hat{T}_{k,a})}}{E(\hat{T}_{k,a})} \times 100 \qquad \text{... (6.2)}$$

Where $ASE(\hat{T}_{k,a}) = \frac{1}{500}\sum_{s=1}^{500}\left(\hat{T}_{k,a}^{s} - T_a\right)^2$ and $E(\hat{T}_{k,a}) = \frac{1}{500}\sum_{s=1}^{500}\hat{T}_{k,a}^{s}$

For k = 1,……,4 and a =1,……., 7.

Where, subscript ‘k’ is used for a particular small area estimator and subscript ‘a’ is for a particular ILRC.

Results & Conclusion

We present the results of ARB and Srse in Table (6.3.1) only for $n = 50$, (a sample of 20 percent villages, as presently adopted in TRS) as the findings from other tables are similar.

For assessing the relative performance of the various estimators, we have to adopt some rule of thumb. Here we adopt the rule that at the ILRCs level, an estimator should not have Srse more than 10 % and bias more than 5%.

We note from the table that none of the estimators satisfy the rule in ILRCs 1 and 2. This may be because, in these circles, there is considerable deviation from the synthetic condition, as observed earlier.

Table 6.3.1: Simulated relative standard error (in %) and Absolute Relative Bias (in %) for various ILRCs under SRSWOR scheme, for n = 50.

ILRCs

Estimator	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
$\hat{T}_{1,a}$	**37.83 (20.00)**	**24.91 (20.83)**	**8.63 (0.81)**	**16.63 (9.87)**	**13.01 (0.193)**	**17.87 (12.00)**	**15.41 (1.181)**
$\hat{T}_{2,a}$	**19.67 (19.12)**	**21.31 (19.60)**	**8.21 (0.75)**	**14.44 (9.66)**	**9.03 (0.085)**	**17.56 (11.53)**	**10.47 (1.071)**
$\hat{T}_{3,a}$	**18.46 (9.80)**	**17.62 (10.18)**	**6.18 (0.98)**	**12.02 (7.32)**	**8.13 (0.523)**	**11.86 (6.61)**	**6.51 (1.68)**
$\hat{T}_{4,a}$	**17.02 (9.00)**	**13.99 (10.09)**	**4.82 (0.8)**	**11.12 (7.10)**	**7.06 (0.47)**	**8.99 (5.20)**	**5.53 (1.50)**

(***Note:*** The figures shown in parentheses are the Absolute Relative Biases in percentage.)

In ILRCs 4 and 6, where the condition deviate moderately, $\hat{T}_{4,a}$ alone satisfies the rule to some extent. In ILRCs 3, 5 and 7, where the synthetic condition closely meet, both $\hat{T}_{3,a}$ and $\hat{T}_{4,a}$ satisfy the rule but $\hat{T}_{4,a}$'s performance is slightly better than $\hat{T}_{3,a}$.

Finally, in view of the above discussion, it is recommended that the use of composite estimator $\hat{T}_{4,a}$ (which is a weighted sum of direct ratio and ratio synthetic estimators), for crop acreage estimation for small domains like ILRCs, tehsils and districts under the TRS scheme, for the cases where the synthetic assumption is satisfied. For other cases we suggest to investigate SICURE model or Bayesian approach to small area estimation problems.

Reference

[1] Cochran, W.G. (1977). "*Sampling Techniques*", Wiley & Sons.

[2] Ghangurde, P.D. and Singh, M.P. (1978). "*Evaluation of efficiency of synthetic estimates*", Proceeding of the Social Statistical Section of the American Statistical Association, 53-61.

[3] Ghangurde, P.D. & Singh, M.P. (1977). "*Synthetic estimates in periodic household surveys*" Survey Methodology, 3, 151-181.

[4] Gonzalez, M.E. (1973). *"Use and evaluation of synthetic estimates"*, Proceedings of the social statistical section of American Statistical Association, 33-36.

[5] 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, 18-25.

[6] Lahiri, D.B. (1954)."*On the question of bias of systematic sampling*", Proceedings of the world Population Conference, 6, 349-62.

[7] Madow, L.H. (1946)."*Systematic sampling and its relation to other sampling designs"*, Journal of American Statistical Association 41:204-217.

[8] Murphy, K.E. Carter, C.M. & Wolfe, L.H. (2001). "*How long should I simulate, and for how many trials?A practical guide to reliability simulations"*,Reliability and Maintainability Symposium, 2001, Proceedings, Philadelphia, p.p. 207-212.

[9] Osborne, J.G. (1942)."*Sampling errors of systematic & random survey of covertype areas*", Journal of American Statistical Association 37:256-264.

[10] Purcell, N.J. and Kish, L. (1979). "*Estimation for small domain"*. Biometrics, 35, 365-384.

[11] Sarndal, C. E.,(1984). "*Design- consistent versus model-dependent estimation for small domains"*, Jour. Amer. Stat. Assoc.,**79**, 624-31.

[12] 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.

[13] 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.

[14] Srivastava, S.K. (1967). "*An estimator using auxiliary information in sample surveys"*, Calcutta Statist. Assoc. Bull. , 121-132.

[15] Sukhatme, P. V. Sukhatme, B. V., Sukhatme, S. and Asok, C. (1984). "*Sampling Theory of Surveys with Applications (3rd Edition)"*, Indian Society of Agricultural Statistics, New Delhi.

[16] Swain, A.K.P.C. (1964)."*The use of systematic sampling in ratio estimates"*, JISA, 2, 160-164.

[17] Tikkiwal, G.C. and Ghiya A. (2000). "*A generalized class of synthetic estimators with application of crop acreage estimation for small domains*", Biom. J., 42, 7, 865-876.

[18] 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), 5, 697-711.

[19] 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 New Series, vol. 8. No. 1, 111-123.

[20] Tikkiwal, G.C. and Pandey, K.K. (2006). "*Generalized Class of Synthetic Estimators for Small Area under Systematic Sampling Scheme"*, A paper

presented in National Conference at Saurastra University, Gujrat form November 11-13, 2006.

[21] Whitt,W. (1989).“*Simulation run length planning”*, in the proceedings of the 1989 winter simulation conference, AT & T Bell laboratories U.S.A., pp- 106-112.

[22] Wolter, K.M. (1984). “*An investigation of some estimators of variances for systematic sampling*”, J. Amer. Stat. Assoc., 79, 781-90.