STATISTICS IN TRANSITION-new series, July 2010
Vol. 11, No. 1, pp. 75—89

GENERALIZED CLASS OF SYNTHETIC ESTIMATORS FOR SMALL AREAS UNDER SYSTEMATIC SAMPLING SCHEME

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

ABSTRACT

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

Key words: Synthetic Estimation; Small Domain; Inspector Land Revenue Circles (ILRCs); Timely Reporting Scheme (TRS); Absolute Relative Bias (ARB); Simulated relative standard error (Srse).

1. Introduction

The common feature of small area estimation problem is that when large-scale sample surveys 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.

The indirect estimators are based on methods which increase the effective sample size either by (i) simulating enough data through appropriate analysis of available data under appropriate modelling or (ii) by using data from other

1 Assistant Professor, CMES, University of Petroleum & Energy Studies, Dehradun- 248007, India.
2 Department of Mathematics & Statistics, J.N.V. University, Jodhpur, Rajasthan -342011, India.

domains and /or time periods through models that assume similarities across domain and /or time periods. The only known method so far belonging to category (i) is SICURE- modelling [TIKKIWAL.(1993)].The other methods of estimation like Synthetic, Composite, and Generalized Regression belong to category (ii). Among these the synthetic estimators are used for small area estimation, mainly because of its simplicity, applicability to general sampling design and potential to increase accuracy in estimation. However, if the implicit model assumption of similarities across domain and /or time period fails, the synthetic estimator may be badly design biased. GONZALEZ (1973), GONZALEZ and WAKESBERG (1973), GHANGURDE AND SINGH (1977, 78), Tikkiwal & Pandey (2007) among others study the synthetic estimator based on auxiliary variables, viz. the ratio synthetic estimator. These studies show that synthetic estimators provide reliable estimates to some extent.

Tikkiwal and Ghiya (2000) define and discuss a generalized class of synthetic estimators for small domains, using auxiliary information, under simple random sampling and stratified random sampling schemes. The generalized class, among others, includes simple, ratio and product synthetic estimators. The two authors compare empirically the relative performance of various direct and synthetic estimators for estimating crop acreage for small domains.

This paper discusses the generalized class of synthetic 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 sampling units like villages / households / agricultural fields etc. are selected by systematic sampling scheme. Systematic sampling scheme, apart from operationally more convenient, provides more efficient estimators under certain conditions [Cf. Cochran (1977), Sukhatme et al. (1984), Madow (1946) & Osborne, J.G. (1942)].

2. 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 be 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,

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

The various population and sample means for characteristic X &Y can be denoted by:

$\bar{X}$ & $\bar{Y}$ = Means of the population based on N observations.

$\bar{X}_a$ & $\bar{Y}_a$ = Population means of domain 'a' based on N_a observations.

$\bar{x}_{ai}$ & $\bar{y}_{ai}$ = Sample means 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 consists of 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 x_{aij} & y_{aij} denote the values of the auxiliary variate and characteristic under study respectively for the j[th] unit of the i[th] sample bearing serial number i +(j-1) k_a , i =1,......., k_a ; j =1,......., n_a . Therefore,

$$\bar{X}_a = \frac{1}{N_a}\sum_i \sum_j x_{aij} \ , \ \bar{x}_{ai.} = \frac{1}{n_a}\sum_{j=1}^{n_a} x_{aij} \ , \ \bar{Y}_a = \frac{1}{N_a}\sum_i \sum_j y_{aij} \ \text{ and}$$

$$\bar{y}_{ai.} = \frac{1}{n_a}\sum_{j=1}^{n_a} y_{aij}$$

Various mean squares and coefficient of variations of subpopulation 'U_a' for auxiliary variate x & characteristic under study, y is denoted by

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

$$C_{y_a} = \frac{S_{y_a}}{\bar{Y}_a}$$

The coefficient of covariance between X and Y is denoted by

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

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 design. 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 until 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

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

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

In this case the x_{aij} & y_{aij} denote the values of the auxiliary variate and characteristic under study 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 $j = 1, 2, \ldots\ldots, n_a$. The various mean squares and coefficient of variations of sub population 'U_a' for auxiliary variate x & characteristic under study, y in this case will be as follows:

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

$$C_{1y_a}^2 = \frac{S_{1y_a}^2}{\overline{Y}_a^2}$$

The coefficient of covariance between X and Y is denoted by

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

3. Generalized Class of Synthetic Estimators

Following Srivastava (1967), we define in this section a generalized class of synthetic estimators of population mean $\overline{Y}_a$ based on the auxiliary variable 'x' under Systematic Sampling Scheme as follows.

$$\bar{y}_{s,a} = \bar{y}_w \left(\frac{\bar{x}_w}{\bar{X}_a} \right)^{\beta} \tag{3.1}$$

Where β is a suitably chosen constant, and

$$\left. \begin{aligned} \bar{y}_w &= \sum{}' p_a \bar{y}_{ai.} + \sum{}'' p_a \bar{y}_{ai.} \\ \bar{x}_w &= \sum{}' p_a \bar{x}_{ai.} + \sum{}'' p_a \bar{x}_{ai.} \end{aligned} \right\} \tag{3.2}$$

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 $N_a \neq n_a k_a$ and $p_a = \frac{N_a}{N}$. Here, clearly

$$E(\bar{y}_w) = \bar{Y} \text{ and } \quad E(\bar{x}_w) = \bar{X} \tag{3.3}$$

The above estimator $\bar{y}_{s,a}$ perform well under the following condition

$$\bar{Y}_a \left(\bar{X}_a \right)^{\beta} \cong \bar{Y} (\bar{X})^{\beta} \tag{3.4}$$

It is noted that the synthetic estimator $\bar{y}_{s,a}$ is consistent if the condition given in (3.4) is satisfied.

Remark 3.1.

If β = 0, -1, 1, the estimator $\bar{y}_{s,a}$ in (3.1) reduces to $\bar{y}_{s,s,a} = \bar{y}_w$, $\bar{y}_{s,r,a} = \left(\frac{\bar{y}_w}{\bar{x}_w} \right) \bar{X}_a$, and $\bar{y}_{s,p,a} = \bar{y}_w \left(\frac{\bar{x}_w}{\bar{X}_a} \right)$ respectively with synthetic condition $\bar{Y}_a \cong \bar{Y}$, $\frac{\bar{Y}_a}{\bar{X}_a} \cong \frac{\bar{Y}}{\bar{X}}$, and $\bar{Y}_a (\bar{X}_a) \cong \bar{Y} \bar{X}$.

4. Design Bias and Mean Square Error of Generalized Synthetic Estimator

Design Bias and Mean Square Error of generalized synthetic estimator, under the synthetic condition given in (3.4), is as follows

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

$$+\frac{\beta(\beta-1)}{2}\left\{\sum_a' p_a^2 \frac{(k_a-1)}{k_a}\frac{S_{x_a}^2}{\bar{X}^2}+\sum_a'' p_a^2 \frac{(N_a-1)}{N_a}\frac{S_{1x_a}^2}{\bar{X}^2}\right\}\Bigg]\Bigg] \tag{4.1}$$

for a = 1, …., A.

$$MSE(\bar{y}_{s,a}) = E(\bar{y}_{s,a}-\bar{Y}_a)^2 = \bar{Y}_a^2\Bigg[\Bigg[\left\{\sum_a' p_a^2 \frac{(k_a-1)}{k_a}\frac{S_{y_a}^2}{\bar{Y}^2}+\sum_a'' p_a^2 \frac{(N_a-1)}{N_a}\frac{S_{1y_a}^2}{\bar{Y}^2}\right\}$$

$$+(2\beta-1)\beta\left\{\sum_a' p_a^2 \frac{(k_a-1)}{k_a}\frac{S_{x_a}^2}{\bar{X}^2}+\sum_a'' p_a^2 \frac{(N_a-1)}{N_a}\frac{S_{1x_a}^2}{\bar{X}^2}\right\}$$

$$+4\beta\left\{\sum_a' p_a^2 \frac{(k_a-1)}{k_a}\frac{S_{y_a x_a}}{\bar{X}\bar{Y}}+\sum_a'' p_a^2 \frac{(N_a-1)}{N_a}\frac{S_{1x_a y_a}}{\bar{X}\bar{Y}}\right\}\Bigg]\Bigg]$$

$$-2\bar{Y}_a^2\Bigg[\beta\left\{\sum_a' p_a^2 \frac{(k_a-1)}{k_a}\frac{S_{y_a x_a}}{\bar{X}\bar{Y}}+\sum_a'' p_a^2 \frac{(N_a-1)}{N_a}\frac{S_{1y_a x_a}}{\bar{X}\bar{Y}}\right\}$$

$$+\frac{\beta(\beta-1)}{2}\left\{\sum_a' p_a^2 \frac{(k_a-1)}{k_a}\frac{S_{x_a}^2}{\bar{X}^2}+\sum_a'' p_a^2 \frac{(N_a-1)}{N_a}\frac{S_{1x_a}^2}{\bar{X}^2}\right\}\Bigg]\Bigg] \tag{4.2}$$

The suitable value of β is the one for which $MSE(\bar{y}_{s,a})$ is minimum. So minimizing the $MSE(\bar{y}_{s,a})$ with respect to β under synthetic condition, gives simplified expression for β, if $\bar{X} \cong \bar{X}_a$ as follows

$$\beta = \frac{-\left\{\sum_a' p_a^2 \frac{(k_a-1)}{k_a}\frac{S_{y_a x_a}}{\bar{X}\bar{Y}}+\sum_a'' p_a^2 \frac{(N_a-1)}{N_a}\frac{S_{1x_a y_a}}{\bar{X}\bar{Y}}\right\}}{\left\{\sum_a' p_a^2 \frac{(k_a-1)}{k_a}\frac{S_{x_a}^2}{\bar{X}^2}+\sum_a'' p_a^2 \frac{(N_a-1)}{N_a}\frac{S_{1x_a}^2}{\bar{X}^2}\right\}} \tag{4.3}$$

It is noted that the expression of MSE for direct estimator under linear & circular systematic sampling design, is minimum if $\alpha = -\frac{C_{x_a y_a}}{C_{x_a}^2}$ [Cf. Srivastava (1967)].

5. Estimation of Mean square errors

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 $\overline{y}_{ai.}$ or $\overline{x}_{ai.}$. 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, 1985) investigate 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 $V(\overline{y}_{ai.})$, 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:

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

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

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

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

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

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 $\overline{y}_{ai.}$ and $\overline{x}_{ai.}$, given by Swain (1964), is

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

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

5.1. Estimation of mean square error of direct estimator

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

Its mean square under case (i) is

$$MSE(\overline{y}_{d,a}^{G}) = \overline{Y}_a^2\left[\frac{V(\overline{y}_{ai.})}{\overline{Y}_a^2} + \frac{V(\overline{x}_{ai.})}{\overline{X}_a^2} + \frac{2\alpha Cov(\overline{y}_{ai.}, \overline{x}_{ai.})}{\overline{X}_a\overline{Y}_a}\right]$$

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

where $R_a = \frac{\overline{Y}_a}{\overline{X}_a}$, thus a consistent estimator of $MSE(\overline{y}_{d,a}^{G})$ is given by

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

Where $r_a = \frac{\overline{y}_a}{\overline{x}_a}$ 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

$$MSE\left(\overline{y}_{d,a}^{G}\right)_c = \overline{Y}_a^2\left[\frac{V(\overline{y}_{ai.})_c}{\overline{Y}_a^2} + \alpha^2\frac{V(\overline{x}_{ai.})_c}{\overline{X}_a^2} + 2\alpha\frac{Cov(\overline{y}_{ai.}, \overline{x}_{ai.})_c}{\overline{X}_a\overline{Y}_a}\right]$$

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

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

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

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 calculated similarly as of v_{2x} and v_{2y}.

5.2. Estimation of mean square error of synthetic estimator

The expression for the Mean Square Error given in (4.2), can be approximated under the synthetic condition given in (3.4) as follows:

$$MSE(\bar{y}_{s,a}) = \sum\nolimits'_a p_a^2 V(\bar{y}_{ai.}) + \sum\nolimits''_a p_a^2 V(\bar{y}_{ai.})_c$$

$$+ \beta^2 R_a^2 \left\{ \sum\nolimits'_a p_a^2 V(\bar{x}_{ai.}) + \sum\nolimits''_a p_a^2 V(\bar{x}_{ai.})_c \right\}$$

$$+ 2\beta R_a \left\{ \sum\nolimits'_a p_a^2 Cov(\bar{y}_{ai.}, \bar{x}_{ai.}) + \sum\nolimits''_a p_a^2 Cov(\bar{y}_{ai.}, \bar{x}_{ai.})_c \right\} \tag{5.10}$$

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

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

$$+ 2\beta r_a \left\{ \sum\nolimits'_a p_a^2 r \sqrt{v_{2y} v_{2x}} + \sum\nolimits''_a p_a^2 r \sqrt{v'_{2y} v'_{2x}} \right\} \tag{5.11}$$

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

6. Crop Acreage Estimation for Small Domain – A Simulation Study

This section demonstrates the use of the generalized synthetic and ratio synthetic estimators to obtain crop acreage estimates for small domain and also compare their relative performance with the corresponding direct estimators empirically, 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)].

6.1. 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 thereby 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.

6.2. 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 is provided in the Table 6.2.1.

Table 6.2.1. Total Area (Irrigated and Unirrigated) under Bajara Crop in 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 is the list of all those estimators, whose relative performance is to be assessed for estimating population total T_a of small domain for ‘a’ = 1, 2 …7.

Direct estimators

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

Direct general estimator $\hat{T}_{2,a} = N_a\, \bar{y}^G_{d,a} = N_a \bar{y}_{ai.} \left(\frac{\bar{x}_{ai.}}{\bar{X}_a} \right)^{\alpha}$

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

Indirect estimators

Ratio synthetic estimator $\hat{T}_{3,a} = N_a\, \bar{y}_{s,r,a} = N_a \left(\frac{\bar{y}_w}{\bar{x}_w} \right) \bar{X}_a$

Generalized synthetic estimator $\hat{T}_{4,a} = N_a\, \bar{y}_{s,a} = N_a\, \bar{y}_w \left(\frac{\bar{x}_w}{\bar{X}_a} \right)^{\beta}$

where $\bar{y}_w = \sum' p_a \bar{y}_{ai.} + \sum'' p_a \bar{y}_{ai.}$; and $\bar{x}_w = \sum' p_a \bar{x}_{ai.} + \sum'' p_a \bar{x}_{ai.}$

Before simulation, we examine the condition of generalized synthetic and synthetic ratio estimators as given in Eq. (3.4) and in remark (3.1). These results are presented in following tables 6.2.2 & 6.2.3 respectively. We note that both the above conditions met for ILRCs (3), (5), (7) deviate moderately 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	$\bar{Y}_a / \bar{X}_a$	$\bar{Y} / \bar{X}$	$[\lvert(\bar{Y}_a / \bar{X}_a) - (\bar{Y} / \bar{X})\rvert \div (\bar{Y}_a / \bar{X}_a)] \times 100$
(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	$\bar{Y}_a(\bar{X}_a)^{\beta}$	$\bar{Y}(\bar{X})^{\beta}$	$\left[\left\lvert\{\bar{Y}_a(\bar{X}_a)^{\beta} - \bar{Y}(\bar{X})^{\beta}\}\right\rvert \div \bar{Y}_a(\bar{X}_a)^{\beta}\right] \times 100$
(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, 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. The simulation length was estimated with the help of the concept discussed by Whitt, W. (1989) & Murphy, K.E. Carter, C.M. & Wolfe, L. H. (2001) based on the steady state condition, that is selecting approximately 10 percent, 20 percent, 25 percent, 30 percent and 35 percent villages independently 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.

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

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

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.

6.3. Results

We present the results of ARB and Srse in Table (6.3.1) only for $n = 50$, (a sample of 20 present 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%.

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

Estimator	ILRCs						
	(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.*

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 a considerable deviation from the synthetic condition, as observed earlier. 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 $\hat{T}_{3,a}$.

From the above analysis it is clear that if the synthetic estimators do not deviate considerably from their corresponding synthetic condition then, performance of the synthetic estimators $\hat{T}_{3,a}$ and $\hat{T}_{4,a}$, based on a sample of 20 present villages (as presently being taken under TRS), is satisfactory at the level of ILRCs. Therefore, these estimators are also likely to perform better both at Tehsil and district levels. When the synthetic estimators deviate considerably from their corresponding synthetic condition, then we should look for other types of estimators such as those obtained through the ***SICURE MODEL*** [TIKKIWAL, B.D. (1993)] and assess their relative performance through studies of the kind, in series, over some years for crop acreage estimation.

Acknowledgement

The author is very thankful to the referee for his valuable suggestions.

REFERENCES

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.

COCHRAN, W.G. (1977).“*Sampling Techniques*”, Wiley & Sons.

GONZALEZ, M.E. (1973). “*Use and evaluation of synthetic estimates”,* Proceedings of the social statistical section of American Statistical Association, 33–36.

GONZALEZ, M. E. and WAKSBERG, J. (1973).“*Estimation of the error of synthetic estimates”,* paper presented at first meeting of the International association of survey statisticians, Vienna, Austria, 18–25, august 1973.

LAHIRI, D.B. (1954).“*On the question of bias of systematic sampling*”, Proceedings of the world Population Conference, 6, 349–62.

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.

MADOW, L.H. (1946).“*Systematic sampling and its relation to other sampling designs”,* Journal of American Statistical Association 41:204–217.

Osborne, J.G. (1942).“*Sampling errors of systematic & random survey of covertype areas*”, Journal of American Statistical Association 37:256–264.

SRIVASTAVA, S.K. (1967). “*An estimator using auxiliary information in sample surveys”,* Calcutta Statist. Assoc. Bull. , 121–132.

SWAIN, A.K.P.C. (1964).“*The use of systematic sampling in ratio estimates*”, JISA, 2, 160–164.

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.

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.

TIKKIWAL, B.D. (1993). "*Modelling through survey data for small domains",* Proceedings of International Scientific conference on small area statistics and survey Design (held in September, 1992 at Warsaw, Poland).

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.

WOLTER, K.M. (1984, 1985)."*An investigation of some estimators of variances for systematic sampling*", J. Amer. Stat. Assoc., 79, 781–90.

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.