Small Area Estimation in India - Crop Yield and Acreage Statistics

B.D. Tikkiwal
Institute of Developmental Research and Statistics, Jaipur, India

Dr. G.C. Tikkiwal
Jai Narain Vyas University, Jodhpur, India

ABSTRACT: This paper first presents the various landmarks in the development of crop yield and acreage statistics in India. It refers to the parental relation in these developments with the developments in other countries. Then, the paper presents findings of two studies, one by B.D. Tikkiwal and the other by G.C. Tikkiwal. These findings should be of help in the future in better organization of current surveys for crop yield and acreage estimation in India and elsewhere, particularly in other developing countries where similar methodology is used. The main findings themselves concern the improvement of estimation procedures for such surveys for big areas, the adoption of certain composite estimators for small area estimation, and the need for organization of such studies in series for some years for trying out B.D. Tikkiwal's SICURE model and developing other small area estimation methods for estimation of crop yield and acreage. In particular, it is empirically established that the estimation procedure used in GCES has provided in Rajasthan State of India underestimates of different crops' average yields, compared to those corresponding estimates provided by a more efficient and at the same time self-weighting procedure used in the Monitoring and Evaluation Unit of the Directorate of Agriculture, Government of Rajasthan.

1. Introduction and Summary

In the context of recent developments in Small Area Methodology, the first attempt to build up agricultural statistics in India for small areas seems to have been made by B.D. Tikkiwal [1991, 1993] in his attempt to try out his Simulation-Cum-Regression (SICURE) model in crop estimation for small areas. An earlier attempt by Panse, Rajgopalan and Pillai [1966] to build up crop yield estimates at the community development (CD) block level through classical Double Sampling Regression Methods using eye-appraisal of a large number of sample fields as auxiliary information did not succeed as the village-account (patwari) agency responsible for this work could not do this work satisfactorily along with his other responsibilities.

A working group on small area, set up by the Central Statistical Organization (CSO) of the Government of India, while considering the need for developing suitable methodologies for meeting the need of data requirements for key characteristics, observed in its report [1990] that from the limited work with regard to the estimation of crop yield for a small area, it appears that the SICURE Model is more promising than the various other methods. This is because, unlike other methods, the model first artificially enlarges, through simulation, the size of the sample at the small area level and then uses auxiliary information for building up multistage-cum-multiphase regression estimators of the type considered in the literature [B.D. Tikkiwal 1965a, Agarwal and Tikkiwal 1980]. The details of how to use the SICURE Model for the estimation of crop yield are given in B.D. Tikkiwal [1993]. According to these details, it is necessary in the beginning to conduct a large number of crop-cutting experiments in a given district to check on the efficacy of an ANOVA Model for simulation purposes. A project for trying out the SICURE Model in a given district was, in fact, submitted by him to a sponsoring agency here. Since, however, only limited funds were made available to him by this agency, he dropped this idea and carried out the project on a more modest scale.

The details of the project and its findings are presented in this paper. It turns out that it is possible to give reliable statistics at district, subdistrict and panchayat samiti / CD block levels with direct methods only. This is done through the data on wheat, obtained in the field study carried out by the M&E unit, Directorate of Agriculture, Government of Rajasthan, India in 1995-96 on our behalf. It is further shown empirically that the estimates of average yield through the GCES scheme, the only national scheme in India, for those crops dealt with by the M&E unit also, have been on the lower side of corresponding M&E estimates all these years because of the use of relatively less efficient estimation procedures in the GCES scheme. The M&E estimation procedure has the further advantage in its being self-weighting.

As regards providing estimates of average crop yield at a level lower than a subdistrict, say at Assistant Agricultural Officer (AAO) Circle level, we have to resort to Small Area Estimation methods including the SICURE model. They can be of the kind found useful in another study by G.C. Tikkiwal [1997], reported briefly in section 4.3 here, for providing reliable acreage statistics at district and lower levels up to the Inspector Land Revenue Circles (ILRCs) in a Tehsil, slightly bigger on an average than a CD block. In his study he uses estimators based on both direct and indirect estimators, referred to in the literature as composite estimators. This is done after reviewing the relevant work on composite estimators [Royall 1977, Schaible 1978, Singh et al. 1993] and then building up a generalized class of composite estimators. The composite estimators considered earlier lie in this generalized class.

Finally, this paper proposes that a series of such studies over some years should be organized not necessarily in Rajasthan alone, but in other parts of India and some other developing countries as well for further development of small area estimation methods for estimating crop yield and acreage. The findings presented in this paper from the said two studies should be of help, in the future, in a better organization of current surveys, for crop yield and acreage statistics, in India and elsewhere, particularly in other developing countries where similar methodology is used. This is clear when these findings are seen in the light of various land marks in the development of crop yield and acreage statistics in India and in the light that we now need such statistics for small areas as well.

2. Landmarks in the Development of Crop Yield Statistics

First we present a brief historical note on the development of methodology for crop estimation surveys for getting necessary yield statistics, not only in India, but also elsewhere. Then we present some details of getting yield statistics in India and their limitations.

2.1 Historical Note

Before the development of survey methodology for estimation of crop yield in India and elsewhere, there were only subjective methods for its estimation resulting in rather unreliable yield statistics. Now crop estimation surveys are being conducted in India and other developing countries through survey methodology developed mostly in the 1940s [Mahalanobis 1946, Sukhatme and Aggarwal 1946-47, 1947-48]. Earlier, there was some pioneering work done by Hubback [1927] in this area, to which a reference was made by Prof. R.A. Fisher in his memorandum dated 2nd March 1945 addressed to the then Imperial Council of Agricultural Research, Government of India. It stated [from Mahalanobis 1946, p. 269]:

> *"The use of the method of random sampling is theoretically sound. I may mention that its practicability, convenience and economy was demonstrated by an extensive series of crop-cutting experiments on paddy carried out by Hubback [later Sir John Hubback, Governor of Orissa] more than twenty years ago over a greater part of the rice tracts in Bihar and Orissa. So far as I know these were the earliest crop-cutting experiments based on the principle of random sampling anywhere in the world. They influenced greatly the development of my methods at Rothamsted."*

2.2 Crop Yield Statistics in India and their Limitations

The overall responsibility for satisfactory conduct of crop estimation surveys in different states and union territories of India was initially entrusted to the Crop Survey Wing of the Statistical Branch of the Indian Council of Agricultural Research. Later, in pursuance of a cabinet resolution in June 1952, this role was entrusted to the Directorate of National Sample Surveys (NSSO), one of the biggest sample survey organizations in the world. This arrangement, over the years, continues. The functions of this central agency, as in the case of former one, are:

- technical direction and overall coordination of crop estimation in different states,
- collaboration with the states in training of primary staff for their adherence to procedural statistical requirements, and
- central check on field work to bring lasting improvements in the agricultural statistics.

In addition to performing the above functions, the NSSO started the practice of conducting land utilization surveys and crop estimation surveys as part of its normal rounds to build up estimates at the all India level of its own from the tenth round onward in 1955-56. Subsequently, several steps were taken to effect improvements in the system of crop estimation in the states. In view of the improvements so effected, it was decided to drop this practice in 1970-71 and adopt one official estimate of crop production in the country based on the estimates of different states as approved and authorized by the respective State Governments. The entire scheme of getting the production estimates in different states and then in the country as a whole is referred to as the General Crop Estimation Surveys (GCES) scheme.

A scheme for improvement of crop statistics (ICS), introduced in 1973-74, provides for supervisory checks through subsamples on the data collected on both area and yield of various crops. In spite of this scheme being in operation for so long, some serious problems of non-sampling errors remain [Rao 1993, Srivastava and Mehta 1996]. They will be all the more serious in building up crop yield and acreage statistics for small areas.

As regards the demand at present for increases in sample sizes to meet small area needs, Rao in his paper [p. 33] makes the following pertinent observation:

> *"I think a ceiling should be imposed on the size of the GCES sample and no further increase should be allowed. In fact even the current scale of experiments (350,000 per annum) is excessive having regard to the lack of control over the field operations. I know there is growing demand for large sample sizes to furnish estimates for small areas especially in connection with the crop insurance scheme, but such needs should be met by separate arrangements by the agencies concerned and they should not form part of the GCES."*

The solution to the various problems arising in building up reliable crop yield statistics presumably lies in the following:

- to reduce the work load on the field agency in temporarily settled states in order to make field operations for GCES more effective. For this, adopt the practice of carrying out a census of all fields (girdawari) in all villages only once in five years, instead of every year, in temporarily settled states, and take other administrative measures to improve this agency. For replacing the present yearly census of all fields by a five yearly census, we have to modify the sample design of TRS to bring in the theory of repeat surveys suitably. This step will also reduce subjectivity in the working of this agency.

- to promote the use of the SICURE Model (described in section 1) after necessary trial. This model can provide a solution to the non-response problem arising out of the loss of planned crop-cutting experiments in GCES. Without this all these years, we have based yield statistics only on the remaining crop-cutting experiments, thus making our estimates biased and less efficient.
- to strengthen TRS and EARAS schemes (described in section 3) in order to meet small area needs and the need of deeper stratification for measuring the impact of irrigation, type of seed, etc.

It is to be noted here that while examining the functioning of state Monitoring and Evaluation (M&E) Units in different State Agricultural Departments of the country, Raheja and Jai Krishna [1991] make pertinent recommendations to improve and broaden the activities of these units. As regards the second phase of the M&E surveys for measuring yield rates of important crops for diagnostic purposes, they recommend suitable integration of these surveys with the GCES. Such an integration should be welcome to effect necessary economy and efficiency. For this, deeper stratification has to take care of the needs of M&E surveys as well.

3. Landmarks in the Development of Crop Acreage Statistics

In order to obtain yield and production statistics for a particular crop, we need reliable estimates of its acreage statistics. The responsibility for collection of acreage statistics for different crops lies with the Departments of Land Revenue and Land Records. The Indian territory for this purpose is categorized into (i) temporarily settled states and Union territories, (ii) permanently settled states, and (iii) states and Union territories of the North East Region. These three categories cover 88, 11 and 1 percent, respectively, of the total area in the country. In the first category, there is a permanent agency of village accountants (patwaris) which is responsible for collection of crop acreage statistics on a complete enumeration basis. In the second category, such data were collected through grid sampling until 1980-81; after this, data are being collected through a scheme known as "Establishment of an Agency for Reporting Agricultural Statistics" (EARAS). This scheme makes it possible to conduct area sample surveys on the lines of the Timely Reporting Scheme (TRS), a scheme introduced in the 1970s in category (i) for reducing time lag in the availability of crop area statistics. In the third category, there is still the old traditional method of giving crop estimates based on personal assessments of village chowkidars in the absence of any regular agency.

4. Meeting Small Area Needs for Crop Yield and Acreage Statistics

For decentralized planning and other purposes like crop insurance, we need reliable agricultural statistics at least at the levels of district and CD block. But the data provided by NSSO, TRS and EARAS are generally reliable at the state level and not at district level. We demonstrate how to meet these small area needs by taking the state of Rajasthan, one of the less developed states in India, for a case study.

4.1 Reliability of Crop Yield Statistics in Rajasthan for Meeting Small Area Needs

The reliability of crop yield statistics is measured in terms of relative standard error or coefficient of variation in percentage of the estimated average yield per hectare. The theoretical consideration under normality model suggests that this percentage should be anywhere from 1.2 to 2.6 in order that the estimated average yield does not differ from the true average yield by more than 5 percent with 95 percent confidence. If this percentage is higher, then the error in the estimated average yield proportionally increases and thus the reliability of the estimated average yield decreases. It is felt that for a good many practical situations, we may tolerate this error up to 5 percent, and up to 10 percent in rare cases.

The reliability of crop yield statistics in Rajasthan was examined in this light by B.D. Tikkiwal [1993, section 6.6] for three crops — maize, paddy and wheat — for five consecutive years (1981-82 to 1985-86). It was noted that the yield estimates for different crops at the state level were satisfactory as expected, but the situation was bad at the district level. This position prevails even now.

To meet the small area needs, we have to increase the number of crop-cutting experiments considerably for yield estimation alone, which is costly and otherwise not desirable in view of Rao's observation in section 2 of his paper. The degree of lack of control over the field operations in different states of India during the three year period 1986-89 is given in the Tabular Statement 5 of his paper. According to this statement, the percentage of crop-cutting experiments correctly conducted during this period in Rajasthan was only between 33 and 45 for both Kharif and Rabi crops. This reflected a highly unsatisfactory state of affairs in its field operations at that time. The situation is reported to have improved considerably since then. Still, far more efficient estimates than required can be obtained through the state agency of M&E Unit in the Directorate of Agriculture, which conducts a much larger number of crop-cutting experiments of its own for its M&E surveys every year. This we illustrate from the study conducted in Rajasthan under our project on small areas.

4.2 Rajasthan Study for Crop Yield Statistics

Wheat, a Rabi crop sown in October-November of the year and then harvested around March-April of the next year, is a major crop in Rajasthan. The said M&E Unit carried out the field study, on our behalf, for the wheat crop for the year 1995-96 in the Sikar district of the state, a district which is compact and has many on-going progressive schemes. The sample design adopted for the study was slightly different from the one followed in the state for these M&E surveys since the beginning of these surveys from Kharif of 1980-81. But this change in the sample design makes all the difference in putting the analysis of such surveys in the state on sound footing and further helps in scientific allocation of the total sample of clusters (chaks) in a year to different strata, keeping in view administrative convenience in distributing the work load evenly amongst the field staff concerned.

It may be noted that Rajasthan is the only state (and recently one more state, Karnatak) which independently conducts crop-cutting experiments, and as a result, the situation in the state is much better than other states in the country where M&E surveys are part of GCES. That the situation is bad in other states as regards the precision of various estimates is clear from the summary of the Report [Raheja and Jai Krishna 1991, section 2.2.3]. The various estimates in these states do not reflect the situation properly even at the district level whereas we need such estimates at least up to the CD block / panchayat samiti level. Thus, our selection of Rajasthan for our case study has added significance. This study should provide a good precedence for other states of the country, and of other developing countries as well, where similar methodology is followed in meeting their small area needs.

4.2.1 Sample Design in the Study

Before we present the sample design used in the study, we give relevant details for the M&E surveys in Rajasthan. For agricultural purposes, a district is divided into a certain number of subdistricts. In the case of the Sikar district, it is divided into two subdistricts: Sikar and Srimadhopur. A subdistrict consists of a certain number of AAO circles, each AAO circle then consists of a certain number of

VEW circles, each VEW circle consists of a certain number of chaks, and finally a chak consists of a certain number of farmers representing their families. In our case, Sikar subdistrict consists of 9 AAO circles and Srimadhopur of 11 AAO circles. Each AAO circle in Sikar subdistrict consists of 7 VEWs, and Srimadhopur of 6 VEWs, except one circle which consists of 5 VEWs. A mandal in each VEW circle of both districts consists of 160 farmers. Each of the chaks forming a VEW circle has necessarily some mandal farmers and some non-mandal farmers.

At the time of launching this study, the sample design in use in Rajasthan's M&E surveys was stratified multistage, the strata being the subdistrict in the concerned district. The first stage units were the chaks in the subdistrict, the second-stage units in the chaks were the farmers in the respective chaks, and then ultimate units of sampling were plots of specified size in one of the fields of each selected farmer. A simple random sampling scheme was adopted at all stages of sampling, though some varying probabilities sampling scheme with replacement at the first-stage could have been tried profitably [Sukhatme and Narain 1952, B.D. Tikkiwal 1965b]. For each crop, 50 chaks were taken in the sample; in turn, from each selected chak only one mandal farmer and one non-mandal farmer; and finally one plot for harvest from each type of farmer was selected. The size of this plot and the mode of its selection was the same as in the GCES scheme.

The above sample design has been prevalent in Rajasthan from Kharif of 1980-81 itself and is in conformity with the guidelines of the World Bank Manual by Slade and Feder [1981], strictly for monitoring surveys but easily extendable to M&E surveys as well. However, in order to have some even distribution of work load at the AAO level, 50 chaks were selected from a given subdistrict in a manner that a certain minimum number of chaks fell in each of the AAO circles of the subdistrict and that no two chaks fell in the same VEW circle. But this control over random selection invalidated the formula used for estimating the precision of the estimates of yield rate in these surveys. Whether this invalidation meant overestimation or underestimation of the precision was not immediately clear, as it meant technically arbitrary allocation of chaks to different AAO circles in a given subdistrict.

In order to overcome this difficulty in our study, AAO circles in each of the two subdistricts are taken as strata in place of subdistricts, now with VEW circles and chaks in the respective VEW circles as first-stage and second-stage units, the remaining stages of sampling remaining the same. However, this creates another difficulty, namely that it is no more possible to measure the exact precision of the various estimators of the average yield considered in the following subsection 4.2.2, as it is not possible to measure the variability between chaks within VEWs and between farmers within chaks with this sample design in which we select only one chak within a selected VEW and then only one farmer within a given selected chak. The respective formulae for various estimators used in the study give overestimates of their precision. However, this difficulty is overcome in some other way.

4.2.2 Main Findings of the Study

Let $\overline{X}_i$ denote the sample mean yield of a particular crop in the ith AAO circle, for different i, of a particular subdistrict. Let $\overline{X}_{st} = \sum_{i=1}^{S} W_i \overline{X}_i$ denote the weighted mean yield of the crop in the subdistrict, where S denotes the number of AAO circles in the subdistrict and W_i is the weight associated with the mean yield for the crop in the ith AAO circle. Let M_i and m_i denote respectively the total number of VEWs and the selected number of VEWs in the ith AAO circle. Let A_i denote the area sown for the crop in the ith AAO circle and A the area sown for the crop in the subdistrict. Thus $A = \sum_{i=1}^{S} A_i$.

By taking W_i equal to m_i/n, A_i/A, M_i/N and $1/S$, where $N = \sum_{i=1}^{S} M_i$ and $n = \sum_{i=1}^{S} m_i$, we get four different estimators of the mean yield of the crop in the subdistrict. The first estimator here corresponds to the estimator used in M&E surveys in Rajasthan. The second estimator corresponds to the one used in GCES [NSSO 1993-94, pp. 153-155]. It may be noted that if the total number n of VEWs selected from the subdistrict is distributed over different AAO circles of the subdistrict in proportion to the areas sown for the crop in different AAO circles, then the first estimator reduces to the second one. Table 1 gives various estimators, their estimates, their estimated variances $\hat{V}(\overline{X}_{st})$, their estimated gains in efficiency due to stratification, and their estimated coefficients of variation (CVs) for both subdistricts in Sikar district. In this table, $\hat{V}(\overline{X})$ denotes the estimate of the variance of the estimator when there is no stratification at the subdistrict level.

Table 1: Results Regarding Wheat Crop Yield in the Year 1995-96 for Mandal and Non-mandal Farmers in Sikar and Srimadhopur Subdistricts, Rajasthan

	Results for Mandal Farmers				Results for Non-mandal Farmers			
Estimator of the mean yield $\overline{X}_{st}$	Value of $\overline{X}_{st}$	$\hat{V}(\overline{X}_{st})$	Estimated gain in efficiency due to stratification (%)	Estimated coefficient of variation (%)	Value of $\overline{X}_{st}$	$\hat{V}(\overline{X}_{st})$	Estimated gain in efficiency due to stratification (%)	Estimated coefficient of variation (%)
	Sikar Subdistrict							
$\overline{X}_{st} = \frac{1}{n}\sum_{i=1}^{S} m_i \overline{X}_i$	7.817	0.0121	60.33	1.407	7.058	0.0115	97.39	1.519
$\overline{X}_{st} = \frac{1}{A}\sum_{i=1}^{S} A_i \overline{X}_i$	7.799	0.0157	23.57	1.607	6.765	0.0149	52.35	1.804
$\overline{X}_{st} = \frac{1}{N}\sum_{i=1}^{S} M_i \overline{X}_i$	7.869	0.0129	50.39	1.443	7.123	0.0120	89.17	1.538
$\overline{X}_{st} = \frac{1}{S}\sum_{i=1}^{S} \overline{X}_i$	7.869	0.0129	50.39	1.443	7.123	0.0120	89.17	1.538
	$\hat{V}(\overline{X})$ = 0.0194				$\hat{V}(\overline{X})$ = 0.0227			
	Srimadhopur Subdistrict							
$\overline{X}_{st} = \frac{1}{n}\sum_{i=1}^{S} m_i \overline{X}_i$	9.078	0.0264	62.50	1.790	8.638	0.0180	50.00	1.553
$\overline{X}_{st} = \frac{1}{A}\sum_{i=1}^{S} A_i \overline{X}_i$	9.035	0.0308	39.29	1.942	8.555	0.0194	39.18	1.628
$\overline{X}_{st} = \frac{1}{N}\sum_{i=1}^{S} M_i \overline{X}_i$	9.640	0.0334	28.44	1.896	9.458	0.0216	25.00	1.554
$\overline{X}_{st} = \frac{1}{S}\sum_{i=1}^{S} \overline{X}_i$	9.650	0.0323	32.82	1.862	8.596	0.0209	29.19	1.682
	$\hat{V}(\overline{X})$ = 0.0429				$\hat{V}(\overline{X})$ = 0.0270			

We note from the table that the estimator used in M&E surveys in Rajasthan has maximum precision amongst all the four estimators considered here. But the estimator used in GCES behaves differently.

In the case of Srimadhopur subdistrict, it is second best for both mandal and non-mandal farmers provided we ignore its CVs. But in the case of Sikar district, it is the poorest of the four estimators for both mandal and non-mandal farmers. Further, it underestimates the average yield of the crop in both subdistricts for mandal farmers as well as non-mandal farmers. Though this underestimation lies only between 4.2 and 0.23 percent of the corresponding M&E estimate, there is additional advantage in using the M&E estimator. It makes the modified sample design used in the study self-weighting, which in turn makes tabulation and subsequent analysis of data easy.

It has been observed all these years that M&E estimators of average yield for different crops are higher than those obtained through GCES in Rajasthan. The above discussion explains why it happens this way.

It may be noted here that each of the four estimators gives estimates of subdistrict yield with much higher precision than required. We can easily build up estimates also for panchayat samitis in Sikar district and also of the district itself. The precision of these estimates should also be much more than required, particularly for the district. However, from Table 2 it is clear that this is not invariably so at the AAO level. Considering that CVs normally should not exceed the 5 percent limit, the precision is as required in 19 cases out of 40 in all. To bring down the CVs in the rest, we need to resort to small area estimation methods of the kind presented in the following subsection. For this, we need to organize such field studies in series over some years.

Table 2: Coefficients of Variation (%) AAO-circlewise for Mandal and Non-mandal Farmers in Sikar and Srimadhopur Subdistricts, Rajasthan

Srimadhopur Subdistrict			Sikar Subdistrict		
AAO	CV for Mandal Farmers	CV for Non-mandal Farmers	AAO	CV for Mandal Farmers	CV for Non-mandal Farmers
1	0	0	1	0	0
2	5.284	4.978	2	5.796	6.276
3	10.570	6.384	3	5.221	5.387
4	3.611	5.176	4	1.083	4.297
5	3.346	3.561	5	3.493	2.123
6	5.453	6.994	6	6.483	3.750
7	6.270	5.480	7	6.346	6.306
8	17.505	14.580	8	2.885	4.364
9	3.860	4.522	9	4.213	5.091
10	5.904	6.773			
11	5.422	4.541			

In AAO circle 8 of Srimadhopur, the CVs for both types of farmers, based on three crop-cutting experiments for each type, appear to be outliers. Therefore, the data from this circle have not been included in the analysis of data presented in Table 1. But, even when we include this data in the analysis, the results regarding the relative performance of the two estimators, one used in GCES and the other in the Rajasthan M&E unit remain unaffected.

We have observed in subsection 4.2.1 that, with the design used in this study, it is not possible to measure the exact precision of the various estimators of the average yield for wheat, as it is not possible to measure the variability between chaks within VEWs and between farmers within chaks. In our future field studies for research suggested above, we have to modify our sample design further to take care of this limitation as well. However, it may be noted that the above observations hold good in spite of this limitation, particularly for the first two estimators, one used in M&E surveys and the other in GCES, where there is so much difference between their estimated gains in efficiency due to stratification. As regards the exact precision of these estimators, their CVs should not exceed more than 5 percent for any subdistrict in view of what follows.

From Table 1, we note that there is a considerable gain in efficiency due to stratification for all the estimators in the two subdistricts. This is going to remain so even when we take into account the variability between chaks within VEWs and the variability between farmers within chaks, as the contribution to the overall variance of an estimator due to further stages of sampling after the first is generally much less. Thus, the assumption that the sample of chaks is randomly drawn directly from the subdistrict itself, as was done in M&E surveys all along, will give an upper limit to the CV of an estimator. Therefore, through the formula used in M&E surveys, we get the upper limits of respective CVs as shown in Table 3.

Table 3. Upper Limits of the Coefficients of Variation (%) for Wheat Crop in the Year 1995-96 for Mandal and Non-mandal Farmers in Sikar and Srimadhopur Subdistricts, Rajasthan

Subdistrict	Type of Farmer	
	Mandal	Non-mandal
Sikar	3.34	1.12
Srimadhopur	4.31	4.18

4.3 Rajasthan Study for Crop Acreage Statistics

Some empirical studies such as by Wakesberg [1973] and Schaible et al. [1977] arrive at the conclusion that when sample sizes for small domains are relatively small, the synthetic or indirect estimator outperforms the direct estimator. However, for relatively large sample sizes, the reverse happens. In the absence of a precise idea about the nature of sample size for a given situation, it becomes necessary to use composite estimators, which are weighted sums of the two types of estimators.

Below we first define a generalized class of composite estimators, due to G.C. Tikkiwal [1997], having earlier composite estimators as special cases. Then, we give seven estimators, direct and indirect, as special cases of the generalized class and consider their relative performance in our Crop Acreage Study carried out in a Tehsil, only slightly bigger on an average than a CD block or panchayat samiti.

Let n_a and N_a denote the sample units and population units in the small domain a. Let $\overline{Y}_a$, $\overline{X}_a$ and $\overline{Y}$, $\overline{X}$ denote, respectively, the sample means based on n_a and $n = \sum_a n_a$ observations selected from N_a and $N = \sum_a N_a$ observations for characteristics y and x. Correspondingly, let $\overline{y}_a$, $\overline{x}_a$ and $\overline{y}$, $\overline{x}$ denote, respectively, the population means based on N_a and N observations. Now we define the generalized class of composite estimators for estimating the population mean $\overline{y}_a$ of the domain a as follows:

$$\overline{Y}_{c,a} = w_a \overline{Y}_a (\overline{X}_a / \overline{x}_a)^{\beta_1} + (1-w_a)\overline{Y}(\overline{X}/\overline{x}_a)^{\beta_2} \ , \qquad (1)$$

where β_1, β_2 and weights w_a are suitably chosen constants. The first component of the composite estimator is a direct estimator ($\overline{Y}_{d,a}$) due to Srivastava [1967] and the second component is a generalized synthetic estimator ($\overline{Y}_{syn,a}$) due to Ghiya [1997]. The best way to choose the values of β_1, β_2 and w_a is in which MSE($\overline{Y}_{c,a}$) is minimum. This gives approximately

$$\beta_1 = -\frac{C_{x_a y_a}}{C_{x_a}^2} \tag{2}$$

and

$$\beta_2 = \frac{\bar{y}(C_x^2 - 4C_{xy}) - 2\bar{y}_a(\frac{C_x^2}{2} - C_{xy})}{4\bar{y}C_x^2 - 2\bar{y}_a C_x^2} \tag{3}$$

in usual notations. The weight w_a is given by

$$w_a = \frac{MSE(\overline{Y}_{syn,a})}{MSE(\overline{Y}_{d,a}) + MSE(\overline{Y}_{syn,a})} \ . \tag{4}$$

We assess the relative performance empirically through a simulation study of this generalized composite estimator $\overline{Y}_{c,a}$ with some other estimators, direct and indirect, belonging to the general class. We give below the list of all these estimators of population total T_a of small domain a, as this total has a meaning for crop acreage.

Direct Estimators:

Direct ratio estimator	$t_{1,a} = N_a(\overline{Y}_a/\overline{X}_a)\bar{x}_a$
Direct general estimator	$t_{2,a} = N_a\overline{Y}_a(\overline{X}_a/\bar{x}_a)^{\beta_1}$

Indirect Estimators:

Ratio synthetic estimator	$t_{3,a} = N_a(\overline{Y}/\overline{X})\bar{x}_a$
Generalized synthetic estimator	$t_{4,a} = N_a\overline{Y}(\overline{X}/\bar{x}_a)^{\beta_2}$
Generalized composite estimator	$t_{5,a} = N_a\overline{Y}_{c,a}$
Composite estimator	$t_{6,a} = N_a[w_a\overline{Y}_a + (1-w_a)(\overline{Y}/\overline{X})\bar{x}_a]$
Composite estimator	$t_{7,a} = w_a t_{1,a} + (1-w_a)t_{3,a}$

In our 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. 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. Before simulation, we first examine the synthetic assumptions of ratio synthetic estimator $t_{3,a}$ and generalized synthetic estimator $t_{4,a}$ with respect to the above small domains. These assumptions are as follows:

$$\bar{y}_a(\bar{x}_a)^{\beta_2} \approx \bar{y}(\bar{x})^{\beta_2} \tag{5}$$

for $\beta_2 = -1$ for estimator $t_{3,a}$ and for the optimum value of β_2 given in (3) for estimator $t_{4,a}$. From this examination, we noted in this study that both assumptions closely meet in ILRCs 3, 5 and 7, deviate moderately in ILRCs 4 and 6, but deviate considerably in ILRCs 1 and 2.

Now taking villages as sampling units for simulation purposes and otherwise, 500 independent simple random samples for each size of 25, 50, 63, 76 and 88 are selected from the population of 252 villages of Jodhpur Tehsil. Then, to assess the relative performance of the estimators under consideration, their Absolute Relative Bias (ARB) and Simulated relative standard error (Srse) or simply coefficient of variation are calculated for each ILRC as follows:

$$ARB(t_{k,a}) = \frac{\left| \frac{1}{500} \sum_{s=1}^{500} t_{k,a}^{s} - T_a \right|}{T_a} \times 100 \tag{6}$$

and

$$Srse(t_{k,a}) = \frac{\sqrt{ASE(t_{k,a})}}{E(t_{k,a})} \times 100 \tag{7}$$

where

$$ASE(t_{k,a}) = \frac{1}{500} \sum_{s=1}^{500} (t_{k,a}^{s} - T_a)^2$$

for k=1..7 and a=1..7.

We present the results of ARB and Srse in Table 4 only for $n = 50$ (a sample of 20 percent villages, as presently done in TRS) as the findings from other tables are similar.

Table 4. Simulated Relative Standard Errors and Absolute Relative Biases, in percentage, of Various Estimators for n = 50 in Different ILRCs

Estimator	ILRC						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$t_{1,a}$	37.37 (0.21)	17.46 (2.28)	8.51 (0.76)	16.29 (0.13)	12.73 (2.41)	12.28 (0.32)	15.29 (2.78)
$t_{2,a}$	18.55 (0.96)	18.32 (1.50)	6.56 (0.12)	15.43 (0.18)	11.27 (1.12)	13.68 (0.54)	11.34 (0.61)
$t_{3,a}$	19.11 (17.90)	20.67 (19.50)	5.71 (0.72)	10.11 (8.66)	5.71 (0.05)	12.14 (11.03)	5.85 (1.02)
$t_{4,a}$	40.54 (39.67)	21.00 (19.84)	5.96 (0.11)	10.17 (8.68)	5.95 (0.18)	8.43 (4.06)	7.16 (3.97)
$t_{5,a}$	16.87 (5.70)	13.80 (9.29)	4.41 (0.10)	8.49 (6.21)	6.17 (0.88)	6.29 (2.82)	6.05 (3.09)
$t_{6,a}$	17.02 (0.91)	17.14 (1.20)	5.07 (0.11)	9.95 (0.17)	8.03 (1.09)	11.42 (0.51)	5.61 (0.59)
$t_{7,a}$	16.48 (8.40)	13.48 (10.20)	4.78 (0.50)	10.15 (6.30)	5.01 (0.38)	8.14 (4.60)	5.45 (1.20)

Note: The figures shown in parentheses are the absolute relative biases in percentage.

For assessing relative performance of the various estimators, we have to adopt some rule of thumb. Here, we adopt the rule that at the ILRC level, an estimator should not have Srse more than 10 percent and bias more than 5 percent. We note from the above table that none of the estimators satisfy the rule in ILRCs 1 and 2. This is happening because, in these circles, there is considerable deviation from the synthetic assumptions, as observed earlier. In ILRC 4, where the synthetic assumptions deviate moderately, $t_{6,a}$ alone satisfies the rule. But in ILRC 6, where deviations are moderate too, $t_{5,a}$ is best. $t_{5,a}$ is best in ILRC 4, too, provided we restrict ourselves only to Srse. In ILRC 3, where the synthetic assumptions closely meet, $t_{5,a}$ is best here also. However, in ILRCs 5 and 7, where the synthetic assumptions closely meet as well, $t_{7,a}$ alone is competitive with others.

From the above analysis and otherwise, we recommend the use of composite estimators $t_{5,a}$ and $t_{7,a}$ at the ILRC level, and thus up to the level of district in TRS, when there are no considerable deviations from synthetic assumptions. When this condition is not satisfied, we should look for other types of estimators such as those through the SICURE Model [B.D. Tikkiwal 1993] and assess their relative performance through studies of the kind, in series, over some years for crop acreage estimation. We have already seen, in section 4.2, a similar need for organization of further studies in series to develop small area methods for crop yield estimation, too.

References

Agarwal, C.L. and Tikkiwal, B.D. (1980), "Two-stages Sampling on Successive Occasions," Sankhya, C., pp. 31-44.

CSO (1990), *Report of the Working Group on Small Area Development Programme Statistics*, Department of Statistics, Ministry of Planning, Government of India, New Delhi.

Ghiya, A. (1997), "On Direct and Indirect Estimation in Survey Sampling," Ph.D. thesis submitted to the J.N.V. University, Jodhpur, India under the supervision of Dr. G.C. Tikkiwal.

Mahalanobis, P.C. (1946), "Sample Surveys of Crop Yields in India," Sankhya, pp. 269-280.

NSSO (1993-94), *Consolidated Results of Crop Estimation Survey on Principal Crops*, National Sample Survey Organization, Department of Statistics, Ministry of Planning and Programme Implementation, Government of India.

Panse, V.G., Rajgopalan, M. and Pillai, S.S. (1966), "Estimation of Crop Yield for Small Domains," *Biometrics*, pp. 374-384.

Raheja, S.K. and Jaya Krishna (1991), "A Report on Special Study on Functioning of State Monitoring and Evaluation Units," sponsored by the Directorate of Extension, Ministry of Agriculture, Government of India.

Rao, V.R. (1993), "Technical Address : Statistics for Rural Development - an Agenda for Nineties," Journal of the Indian Society for Agricultural Statistics, Vol. XLV No. 1, pp. 1-47.

Royall, R.A. (1977), "Statistical Theory of Small Area Estimates - Use of Prediction Models," unpublished report prepared under contract from the National Center for Health Statistics.

Schaible, W.L. (1978), "Choosing Weights for Composite Estimators for Small Area Statistics," *Proceedings of the Survey Research Methods Section*, American Statistical Association, Washington, D.C., pp. 741-746.

Schaible, W.L., Brock, D.B., Casady, R.J. and Schnack, G.A. (1977), "An Empirical Comparison of the Simple Inflation, Synthetic and Composite Estimators for Small Area Statistics," *Proceedings of the American Statistical Association, Social Statistics Section*, pp. 1017-1021.

Singh, D. (1993), "Some Studies in Small Area Statistics," Ph.D. thesis, Indian Agricultural Statistics Research Institute library.

Slade, R.H. and Feder G. (1981), *The Monitoring and Evaluation of Training and Visit Extension in India*, manual of Instructions, World Bank.

Srivastava, S.K. (1967), "An Estimator Using Auxiliary Information in Sample Surveys," *Calcutta Statistical Association Bulletin*, pp. 121-132.

Srivastava, S.S. and Mehta, R. (1996), "Qualitative Issues in Agricultural Statistics," Souvenir, 50th Annual Conference of Indian Society for Agricultural Statistics, pp. 49-55.

Sukhatme, P.V. and Aggarwal, O.P. (1946-47, 1947-48), "Report on the Crop-cutting Survey on Wheat by the Random Sampling Method in Delhi," Indian Council of Agricultural Research, New Delhi.

Sukhatme, P.V. and Narain, R.D. (1952), "Sampling with Replacement," *Journal of the Indian Society for Agricultural Statistics*, Vol. 4, pp. 42-49.

Tikkiwal, B.D. (1965a), "The Theory of Two-stage Sampling on Successive Occasions," Journal of the Indian Statistical Association, pp. 125-136.

Tikkiwal, B.D. (1965b), *Some Aspects of Sampling with Varying Probabilities from a Finite Population*, University of Rajasthan Studies (Science Series), pp. 31-52.

Tikkiwal, B.D. (1991), "Modeling Through Survey Data for Small Domains," keynote address at the Symposium on Modeling held at Kurukshetra University, March 7-9, 1991.

Tikkiwal, B.D. (1993), "Modeling Through Survey Data for Small Domains," *Proceedings International Scientific Conference on Small Area Statistics and Survey Design* held in September, 1992 at Warsaw, Poland.

Tikkiwal, G.C. (1997). "On Estimation of Crop Area Statistics for Small Domains," in press *Proceedings of the 2nd Annual Conference of Information Theory and Applications and National Symposium in Statistics with Focus on Applications in Agriculture and Allied Sciences*, held at CCS Haryana Agricultural University, Hissar February 7-9, 1997.