"AN EMPIRICAL STUDY FOR ESTIMATION OF CROP PRODUCTION AT SMALLER GEOGRAPHICAL AREA BASED ON VARIOUS FACTORS INCLUDING WEATHER VARIABLES"

THESIS

SUBMITTED TO THE
NARENDRA DEVA UNIVERSITY OF AGRICULTURE & TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE (AGRICULTURE)

IN

AGRICULTURAL STATISTICS

By

Ashish Shukla

I.D. No. A-8263/14

DEPARTMENT OF AGRICULTURAL STATISTICS
NARENDRA DEVA UNIVERSITY OF AGRICULTURE & TECHNOLOGY,
NARENDRA NAGAR (KUMARGANJ), FAIZABAD-224 229 (U.P.)
INDIA

2016

DEDICATED
To My
Beloved Parents

Mr. Janardan Shukla

Mrs. Savitri Shukla

Ashish Shukla . . . .

Dr. B.V.S. Sisodia
Professor& Head

Deptt. Of Agricultural Statistics
N.D. University of Agriculture &Technology,
Narendra Nagar (Kumarganj),
Faizabad- 224 229, (U.P.) INDIA

CERTIFICATE-I

This is to certify that the thesis entitled ***"An empirical study for estimation of crop production at smaller geographical area based on various factors including weather variables"*** submitted for the degree of '**M.Sc. (Ag.)**' in subject of '**Agricultural Statistics**' of the Narendra Deva University of Agriculture & Technology, Narendra Nagar (Kumarganj), Faizabad (U.P.) is a bonafide research work carried out by **Mr. *Ashish Shukla***, I.D No. **A-8263/14**, under my supervision and that no part of this thesis has been submitted for any other degree.

The assistance and help received during the course of investigation have been duly acknowledged.

Narendra Nagar **(B. V. S. Sisodia)**
8 July, 2016 Major Advisor & Chairman

ACKNOWLEDGEMENT

Acknowledgment for a few might be just a trifle thing written on a piece of paper. But in its true sense it gives us an opportunity to remember and express our feelings for those whom we love and revere. Here I get a chance to express my token of thanks to people who have touched me in one way or the other by their small things. Words are not enough to express my feelings for them yet these lines are not exaggeration but feelings, which come straight from heart.

*All and every praise be for **"GOD"**, the most magnificent, the most beneficent. I pray to him for his guidance and protection throughout my life. It is with his guidance and help that I have been able to complete this huge task.*

*Let me first of all, I feel great pleasure in taking the opportunity to express my heart-felt thanks, sense of gratitude and reverence to my **Major Advisor and Chairman of Advisory Committee, Dr. B. V. S. Sisodia**, Prof. and Head, Department of Agril. Statistics for his incredible support, meticulous guidance, indelible inspiration, persistent encouragement, ingenious suggestions, mellifluous nature, continued patience, constructive criticism, indefatigable attitude during the course of investigation. Without his stimulating guidance, meticulous counsel and tremendous encouragement, this study would not have been possible.*

*I am highly indebted to my advisory committee **Dr. R. A. Singh,** Professor, Department of Agril. Economics, **Dr. Ghanshyam Singh,** Professor, Department of Agronomy, **Dr. Prakash Singh,** Associate Professor, Department of Ext. Education, and **Dr. V. N. Rai,** Associate Professor, Department of Agril. Statistics who have paid special attention to the constructive criticism of this manuscript and they have also taken pain in giving proper suggestions, decent advice and liberal help rendered throughout the period of study.*

*I also offer gratefulness to all teaching staff of my department especially to **Shri Sunil Kumar,** Associate Professor for providing necessary facilities and* *fruitful suggestions during preparation of this manuscript and entire courses of my study.*

I feel paucity of words in expressing my outmost gratefulness and reverence to ***Dr. Akhtar Haseeb****, Hon'ble Vice*-Chancellor, ***Dr. P. K. Singh****, Registrar and* ***Dr. Bhagwan Singh****,* Dean, College of Agriculture of this university for providing all sorts of facilities to carry out the present investigation.

I feel golden opportunity and proud privilege to work under the talented and inspiring guidance of ***Dr. B. V. S. Sisodia****, Prof. and Head, Department of Agril. Statistics, for his imperative and facile suggestion, conspicuous ability, constructive criticism and above all the inculcating in me a spirit of independent thinking with confidence which made this endeavor possible. I am extremely indebted to him for being meticulous through.*

And thanks for the non-teaching staff members ***Sri D. K. Singh, Sri Om Prakash, Sri Ram Lal Tiwari, Sri Saligram Yadav, Sri Ramroop, Sri Santosh, Sri Ram Manorath Mishra, Sri Jagram, Sri Ram Murti and Sri J. P. Shukla*** *of Department of Agricultural Statistics for providing necessary facilities and cordial help whenever needed.*

I found no rhetorical gems from the ocean of words to express my hearted sense of veneration and gratitude to my father ***Mr. Janardan Shukla****, my mother* ***Mrs. Savitri Shukla,*** *my elder sisters* ***Mrs. Anupama Pathak, Mrs. Nirupama Dubey, Mrs. Swarupama Pandey and Mrs. Binu Pandey*** *Jijus* ***Mr. Mukesh Dubey, Mr. Arun Pandey, Mr. Amit Pandey*** *lovely sisters* *Antima, Ankita, Priyanka, Shreya, Shaumya* *younger brothers* ***Shashank, Ashutosh, Abhishek,*** *and my elder brother* ***Nitesh****, who have always provided support for my higher education and chi-tiring activities for whole of the day that remain a source of inspiration and engorgement to recharge me to work hard and exhilaration has always animated me to rise against problems and face them smiling.*

I am highly thankful to my UG coordinators ***Dr. H. N. Tripathi****, Professor, Department of Agricultural Engineering (DDU) and* ***Dr. Rajesh Kumar Mishra****,* *Associate Professor, Department of Philosophy (MGKV) for their regular guidance, motivation and incredible support.*

I would like to express my affection to my loving friends ***Ram Ashish, Happy, Chandra Sekhar, Awadhesh, Ravi, Nitin, Avinash, Alekh, Sanjeev, Awadh Bihari, Shailesh, Ritesh, Ravindar, Vibhansu, Praveen, Sriom, Rajkumar, Amaresh, Ranjan, Pradeep, Karudendar.***

I also wish to express my indebtedness to ***My Seniors Dr. Dhyanesh Shukla, Dr. Annu Bhaskar, Dr. K. K. Maurya, Dr. Amar Singh, Sri Hargovind Singh , Dr. Monika, Dr. Azfar, Ramesh, Nitin, Anshul, Snehdeep, Sweta, Sandhya, Manoj, Dayaram, Sanjay, Ajay, Sandeep, Dheerendra, Ekta, Sarvesh, Anupam, and Uday Bhan, and My colleagues Elango. S, Amit Mishra, Amit Singh and Alok Mani Tripati and My lovely Juniors Pallavi, Naveen, Ravi, Sarvesh*** *because without their patience, inspiration, love and co-operation, this task would have been a more difficult challenge to me.*

Besides these, many more people helped me. If I don't list them all it is not for lack of gratitude but for lack of space which compels me to hold my pen down and therefore, I give them all my thanks.

Acknowledgment are generally written as the preface of a thesis in order to thank those who have been helpful in either ways to complete it but I am going to keep those all people here so that wherever I go through it, I should be very honest to accept that it was **not only me** *who did it.*

Again I thank, ***'God'*** *and my family for his blessing bestowed upon me.*

Narendra Nagar, Faizabad
8July, 2016 ***(Ashish Shukla)***

INTRODUCTION

Sample surveys have long been used as cost-effective means for data collection. Such data is used to provide suitable statistics not only for population targeted by the survey but also for a variety of a sub – populations, often called domains. In recent years, there has been increasing demand in most countries for reliable data not only at the national level, but also for sub-national levels or domains.

The small area statistics have been derived from large scale surveys where survey statistician is not only interested in estimating the population parameters such as mean or total but also the parameters of sub-populations, called domain. A special feature of domain studies is that survey is not planned specifically for estimating domain parameters due to high resources requirement but these are developed with the help of information obtained in the large scale surveys. When the domain sizes are very small, sometimes the usual direct sampling approach may have inadequate representation of such domains in the sample and therefore, estimates of parameters for these domains may not be reliable. Such types of small domains are also called small area. According to Brackston (1987) “any small domain may be regarded as a small area for which direct design based estimates cannot be reliably produced from the current sample survey programme”.

In view of the government emphasis on lower level administrative units for all policy planning, role of techniques based on small area estimation has increased considerably. These small areas may be in the form of geographical or administrative units such as tehsils, Assembly constituencies, Community Development Blocks or still smaller units like Municipalities, Mandals or Village Panchayats in India. Sometimes, group of individuals belonging to different economic strata may also be considered as small area. Welfare of such groups may be important for the government and other welfare agencies.

This has become more relevant in the context of India due to enactment of Panchayati Raj system in the country where the distribution of central funds are often made on a local or regional basis and it may depend on variables such as number of unemployed, condition of the housing, use of fertilizers, principal occupation of the households, infrastructural needs, land and water endowments *etc.* The number of techniques for estimation of small area statistics has been evolved. These techniques have been widely used in developed countries like UK, USA etc. in the area of health, education, employment, industries, labour forces, *etc.* particularly for estimation of small area statistics.

In India, the organization like National Sample Survey Organization (NSSO), Central Statistical Organization (CSO), Indian Institute of Population Studies (IIPS) *etc.* are mainly responsible to provide statistics on various aspects such as socio-economic indicators, unemployment, education, industrial parameters, health statistics, poverty statistics, labour forces *etc.* through large scale surveys. However, these organizations have not been able to integrate the techniques of small area estimation in their large scale surveys to achieve the goal of small area statistics for the decentralized planning process in the country. Srivastava *et al.* (2007) worked on small area estimation with National Sample Survey Data. They used already available small area estimation techniques to derive district level estimates of amount of loan outstanding per household using data from the 2002-03 Debt-Investment Survey of National Sample Survey Organization (NSSO) for the rural areas of Uttar Pradesh.

Agriculture in India is the means of livelihood of almost two third of the work force in the country. It is the major source of income for about three-fourths of India's population who live in villages. Agriculture is not only an important occupation of the people, but also way of life, culture and custom. Further, as the Indian economy is mainly based on agriculture, its proper planning is very important. The planning in agriculture is mainly looked after by the Planning Commission of India (now Niti Ayog) which operates and executes under the aegis of the Government of India. The sole objective of the Planning Commission in terms of Agriculture Planning in India is to enhance the total output of agriculture and boost the economic growth of the country. However, availability of reliable statistics is a key for success of any planning process and their monitoring. India has a well-established National Agricultural Statistics System. The system is very comprehensive providing data on variety of parameters on interest of agricultural production system. As a result, reliable estimates of various parameters of interest are available at the macro level. In view of the decentralized system of planning in the country, reliable estimates of various parameters are required at the micro level.

The productivity and production of crops are two important parameters of interest for policy formulation including crop insurance which are available at district level through scientifically designed Crop Cutting Experiments (CCEs) under the scheme of General Crop Estimation Surveys (GCES). For micro level planning, estimates of crop productivity/Production are required at small geographical area, for example, Blocks or Gram Panchayat levels (GP). CCEs are not economically viable proposition at block or GP levels in view of their large numbers. Accumulations of large non-sampling errors are also likely to creep in. An attempt was made long back by Panse *et al* (1966) and Singh (1968) to estimate block level crop production by CCEs using double sampling approach. But this could not succeed due to physical constraints within frame of sampling design.

Srivastava *et al* (1999) and Sud *et al* (2001) proposed some methodologies for estimation of crop production at Block/GP level but their methods lacked practical implication. Sisodia and Singh (2001), Sisodia and Chandra (2012), Sisodia and Singh (2012) and Singh *et al* (2012) have developed some methodologies based on regression model between crop productivity/production and related covariates using time series data on these items which are available at district levels. Sharma *et al* (2015) and Sharma and Sisodia (2016) have made an attempt to apply technique of principal component and discriminant function analysis for estimation of crop production at block level. The fitted model at district level was used to develop block/GP level estimate of crop production provided information on the same covariates are also available at Block/GP levels.

The empirical study conducted by the preceding authors on rice and wheat crops in Barabanki and Faizabad district of Uttar Pradesh, India, has revealed that the block estimates of crop production obtained through regression model were quite close to the actual block estimates based on CCEs for almost all the block except for few blocks with reasonable standard error. It may be mentioned here that the block estimates of crop-production based on CCEs were available for some specific years through pilot studies conducted by U.P. State Government, which were used by the preceding authors for validation of the results obtained by their proposed methodologies.

It is also obvious that all the blocks in the district may not be uniform in terms of land-type, soil types, agro-ecosystem *etc*. Thus, the block effects are random and their effects need to be taken into account to develop more precise estimate of crop-production at block-level. Block specific variability typically remains even after accounting for the auxiliary variables used in the regression model at the district level. This limitation can be handled by an alternative estimation technique based on an explicit linking model, which provides a better approach to small area estimation (SAE) by incorporating random block-specific effects that account for the between block variation beyond that is explained by the auxiliary variables included in the model, referred to as the mixed model (Fay and Herriot, 1979). The random block effects in the mixed model capture the dissimilarity between the blocks. Sisodia *et al* (2015) have used area level model to develop more estimate of crop production at Block level by accounting Block variation.

The methodologies suggested by aforesaid research workers provided consistent estimates of block level estimate of crop production. The covariates they considered were mostly input factors. Climate plays an important role in crop production besides input factors. Climate change is now the crucial factor in agricultural production these days. Therefore, the information on weather variables could be the important factors in developing predictive models for estimation of crop production besides other input factors at smaller geographical area like Block or GP level.

In view of the above facts, the present investigation has been undertaken to develop predictive model for estimation of crop production at small area using information on weather variable and other input factors with the following specific objectives.

1) To develop empirical predictive model for estimation of crop production at Block level using weather variables and other factors.
2) To carry out empirical study using time series data on crop yield and other factors including weather variables for validation of the proposed model in (1).

REVIEW OF LITERATURE

2.1. Introduction

Sample survey is an economically viable statistical technique to draw inference about population parameters in finite population sampling. Most of the sample surveys are designed to obtain the estimate of population parameters like population mean, total etc. at aggregate level, i.e. at national or state or district level. When the estimation of parameters for the part of the population is required, domain studies are quite common in sample surveys where such parts of the population are referred to as domains of interest. A special feature of domain studies is that surveys are not planned specifically for estimating domain parameters but they are developed on the basis of available samples in the respective domain. Of course, the choice of suitable estimator depends on domain sizes, or availability of some other related domain parameters in some situations. The crux of this approach is that domains must be large enough to ensure sufficient sizes of the samples from the domains so that valid estimate could be developed. The requirements of still smaller areas are generally not met satisfactorily through usual estimation procedure in sampling. Such small areas require different approach. Purcell and Kish (1979) defined approximate ranges for various types of domains as major domains (1/10 of the population or more), minor domains (between 1/10 to 1/100 of the population), mini domains (1/100 to 1/10,000 of the population) and rare domains (< 1/10,000 of the population). Although, this classification was not so rigid, but it provides a fairly good idea about the smallness of "Small Area". For example, the smaller areas may be in the form of geographical or administrative units such as Community Development Blocks or still small units like Mandals or Gram Panchayats in India or may be groups of individuals belonging to different economic strata *viz*., the group of people belonging to socially deprived classes or categories of farmers. Local estimation of population related to health and vital statistics is another aspect of small area estimation problem. Welfare of such groups may be some times crucial for the government and other concerned agencies. A lot of works have been done by various research workers for the development of Small Area Estimation (SAE) techniques in the past. Various approaches have been resorted to develop estimation techniques for small area statistics, viz.

(i) Direct and indirect (synthetic) method of estimation based on sample surveys data available for larger area

(ii) Demographic methods

(iii) Model based estimation techniques based on sample surveys data available for larger area, and

(iv) Estimation/prediction approach using secondary data (time series data)or primary data available at larger area by applying suitable regression model.

A comprehensive review of the works in these above headings is presented in this chapter.

2.2. Direct and indirect (synthetic) method of estimation based on sample surveys data available for larger area.

2.2.1. Direct Method of Estimation.

The early work in the small area estimation can be dated back to 1966 when Panse *et al.* (1966) and Singh (1968) examined the feasibility of using double sampling for estimation of yield at the block level. The technique consisted in selecting a large sample of villages and fields from the block for eye estimation of yield prior to the harvest and combining it with results of crop cutting experiments conducted on a sub-sample for obtaining estimates of average yield at the block level. The technique envisaged presence of a strong positive correlation between eye estimate and the actual yield harvested. However, this approach could not succeed due to various physical constraints. It may be remarked that small area statistical techniques are not the substitute for sample survey approach if the survey approach is within the budget constraints.

If the sample survey is planned for larger area, then small area statistics can be developed using sample information available for larger area. The direct method of estimation of small area statistics is described below.

Let us consider the population U = {1, 2, …, N } be divided into D non-overlapping small area $\sum_{d=1}^{D} N_d = N$. Let the population also be divided into G non-overlapping groups which are generally considered to be larger than small area. It is also obvious that $\sum_{g=1}^{G} N_g = N$. Let Y be the characteristics of interest and the objective is to estimate $\overline{Y}_d$ or Y_d, the mean or total for small area d, d = 1, 2, …, D. Let us consider, a sample 's' of size 'n' is drawn from the population with probability P(s). π_i and π_{ij} are inclusion probability of i^{th} and paired unit i and j in the sample of a sampling design P. Let s_d denotes the intersection of s and U_d with cardinality n_d, which is a random variable subject to $\sum_{d=1}^{D} n_d = n$. Similarly, let s_g denotes the intersection of s and U_g with cardinality n_g s.t. $\sum_{g=1}^{G} n_g = n$. Also, s_{dg} denotes the intersection of s and U_{dg} with cardinality n_{dg} s.t. $\sum_{d=1}^{D}\sum_{g=1}^{G} n_{dg} = n$.

The most common direct estimator of Y_d is the Horvitz-Thompson estimator

$$\hat{Y}_{d(HT)} = \sum_{s_d} Y_k/\pi_k \qquad (2.2.1)$$

When simple random sampling without replacement (SRSWOR) is applied, then $\pi_k = \frac{n}{N}$ and (2.2.1.1) reduces to

$$\hat{Y}_d = \frac{N}{n} n_d \bar{y}_d \qquad (2.2.2)$$

where $\bar{y}_d$ is sample mean for d^{th} small area based on n_d units. A reasonable alternative estimator of Y_d is,

$$\hat{Y}'_d = N_d \bar{y}_d \qquad (2.2.3)$$

but it is not unbiased.

Another estimator for Y_d is,

$$\hat{Y}''_d = \sum_{g=1}^{G} N_{dg} \bar{y}_{dg} \qquad (2.2.4)$$

which is equivalent to an estimator in stratified random sampling. However, a post-stratified HT-type estimator based on G groups counts is

$$\begin{aligned}\hat{Y}_{d(HT/C)} &= \sum_g N_{dg} \Big(\sum_{s_{dg}} y_k/\pi_k\Big)\Big(\sum_{s_{dg}} 1/\pi_k\Big)^{-1} \\ &= \sum_g N_{dg} (\hat{Y}_{dg(HT)}/\hat{N}_{dg(HT)}) \\ &= \sum \hat{Y}_{dg}(G_{a/c}) \qquad (2.2.5)\end{aligned}$$

where $\hat{Y}_{dg}(G_{a/c})$ is Haj'ek's estimator of Y_{dg} $(= \sum_{P_{dg}} y_k)$ based on cell counts (Haj'ek, 1959).

If the values x_{dgk} of an auxiliary variable x closely related with y_{dgk} study variable Y (k= 1, 2, … n_{dg}) are available for all (d, g, k), a post-stratified HT- type estimator based on G groups-ratios is

$$\hat{Y}_{d(HT/r)} = \sum_g X_{dg} (\hat{Y}_{dg(HT)}/\hat{X}_{dg(HT)}) \qquad (2.2.6)$$

If all samples s_{dg} are non-empty, the post-stratified estimators (2.2.4) to (2.2.6) are approximately unbiased. If n_d is small, n_{dg} may possibly be zero in some cases, so the estimator in (2.2.4) to (2.2.6) may be useless estimators. However, the practitioners sometimes combine cells to avoid the problems with zero counts. Nevertheless, to overcome the problems with zero counts in some cells the methods of indirect or synthetic estimation are developed which are being reviewed in the following section.

2.2.2. Indirect or Synthetic Estimation Method

At first time, National Center for Health Statistics of USA (1968) used synthetic estimation to calculate state estimates of long and short-term physical disabilities from the National Health Interview Survey data.

Gonzalez (1973) described synthetic estimates as follows: “An unbiased estimate is obtained from a sample survey from a large area; when this estimate is used to derive estimates for sub-areas under assumption that the small areas have the same characteristics as the large area, and identify these estimates as synthetic estimates”. This method is traditionally used for small area estimation, mainly because of its simplicity, applicability to general sampling designs and potential of increased accuracy in estimation by borrowing information from similar small areas.

Using the notations of sub-section 2.2.1, the synthetic estimators of Y_d are given by

$$\hat{Y}_d^S = \sum_{g=1}^{G} N_{dg} \bar{y}_{.g} \qquad (2.2.1.1)$$

and $$\hat{Y}^{S}_{d/r} = \sum_{g=1}^{G} (X_{dg}/X_{.g})\hat{Y}'_{.g} \qquad (2.2.1.2)$$

where $\bar{y}_{.g} = \frac{1}{n_g}\sum_{d}\sum_{k}\hat{y}_{dgk}$ and $\hat{Y}'_{.g}$ is estimate of Y_g resulting into $\hat{Y}' = \sum \hat{Y}'_{.g}$ as an estimate of Y. The estimator (2.2.1.2) is due to Purcell and Linacre (1976) and Ghangurde and Singh (1977).

The estimator in (2.2.1.1) and (2.2.1.2) are not unbiased but more useful than unbiased direct estimator. However, the estimator (2.2.1.2) has the desirable consistency property that $\sum_{d}\hat{Y}^{s}_{d/r}$ equals the reliable direct estimator $\hat{Y}' = \sum_{g}\hat{Y}'_{.g}$ of the population total Y, unlike the original estimator proposed by the National Center for Health Statistics (NCHS) in 1968 which uses the ratio $X_{dg} / \sum_{g} X_{dg}$ instead of $X_{dg}/X_{.g}$. The synthetic estimator of NCHS will be unbiased if the following assumptions are satisfied.

(i) $\bar{Y}_{dg} = \bar{Y}_{.g}$

(ii) $\bar{X}_{dg} = \bar{X}_{.g}$

where $\bar{Y}_{dg}$ and $\bar{Y}_{.g}$ are, respectively, mean value of Y for the $(dg)^{th}$ cell and g^{th} group. The assumption (ii) states that cell mean for the auxiliary variable is equal to the small area mean. This means that the groups are formed in such way that variation between groups is small for the auxiliary character. This assumption is quite restrictive, and may not be achievable in practice.

The direct estimator $\hat{Y}'_{.g}$ used in (2.2.1.2) is typically a ratio estimator of the form

$$\hat{Y}_{.g} = \left[\sum_{k=s_{.g}} (\omega_k y_k) / \sum_{k=s_{.g}} (\omega_k x_k) \right] X_{.g} = (\hat{Y}_{.g} / \hat{X}_{.g}) X_{.g} \qquad (2.2.1.3)$$

where ω_k is the sampling weight attached to the kth element and $s_{.g}$ denotes the sample in the large domain g. For this choice, the synthetic estimator (2.2.1.2) reduces to

$$\hat{Y}^{s}_{d/r} = \sum_{g} X_{dg} (\hat{Y}_{.g} / \hat{X}_{.g}) \qquad (2.2.1.4)$$

If $\hat{Y}'_{.g}$ is approximately design-unbiased, the design bias of $\hat{Y}^{s}_{d/r}$ is given by

$$E\ (\hat{Y}^{s}_{d/r}) - Y_d = \sum_{g} X_{dg} \left(\frac{\overline{Y}_{.g}}{\overline{X}_{.g}} - \frac{Y_{dg}}{X_{dg}} \right) = \sum_{g} X_{dg} \left(\frac{\overline{Y}_{.g}}{\overline{X}_{.g}} - \frac{\overline{Y}_{dg}}{\overline{X}_{dg}} \right)$$

which is not zero unless $\frac{Y_{dg}}{X_{dg}} = \frac{\overline{Y}_{.g}}{\overline{X}_{.g}}$ for all g. In the special case where the auxiliary information X_{dg} equals to the population count N_{dg}, the latter condition is equivalent to assuming that the small area means $\overline{Y}_{dg}$ in each group g equal the overall groups mean, $\overline{Y}_{.g}$. Such an assumption is quite strong, and in fact synthetic estimators for some of the areas can be heavily biased in the design-based framework. However, if the groups are formed on the principal of stratification, obviously $\overline{X}_{dg} = \overline{X}_{.g}$ and in such situation aforesaid condition may be more realistic as against NCHS estimator.

It follows from (2.2.1.2) that the design-variance of $\hat{Y}^{s}_{d/r}$ will be small since it depends only on the variances and co-variances of the reliable estimators $\hat{Y}'_{.g}$. The variance of $\hat{Y}^{s}_{d/r}$ is readily estimated, but it is more difficult to estimate the MSE of $\hat{Y}^{s}_{d/r}$. Under assumption Cov($\hat{Y}_{d}$, $\hat{Y}^{s}_{d/r}$) = 0, where $\hat{Y}_d$ is a direct (an unbiased estimator of Y_d), an approximately unbiased estimator of MSE is given by

$$\text{MSE}\,(\hat{Y}^{s}_{d/r}) = (\hat{Y}^{s}_{d/r} - \hat{Y}_{d})^2 - V(\hat{Y}_{d}) \qquad (2.2.1.5)$$

Here V($\hat{Y}_{d}$) is a design-unbiased estimator of variance of $\hat{Y}_{d}$. The estimator (2.2.1.5), however, is very unstable. Consequently, it is customary to average these estimators over d to get a stable estimator of MSE (Gonzalez, 1973), but such a global measure of uncertainly can be misleading. Note that the assumption Cov ($\hat{Y}_{d}$, $\hat{Y}^{s}_{d/r}$) = 0 may be realistic in practice since $\hat{Y}^{s}_{d/r}$ is much less variable than $\hat{Y}_{d}$.

Singh *et al*. (2002) proposed a ratio synthetic estimator

$$\hat{Y}_{d} = \sum_{g} w_{dg} \frac{\hat{\overline{Y}}_{.g}}{\hat{\overline{X}}_{.g}} \overline{X}_{.g} \qquad (2.2.1.6)$$

where $w_{dg} = X_{dg} / \overline{X}_{.g}$ following the same condition in which the estimator due to Ghangurde and Singh (1977) is unbiased. This estimator (2.2.1.6) is biased. For bias to be zero, the assumption $\rho_{g} = c_{gx} / c_{gy}$ should be satisfied, where ρ_g, c_{gx} and c_{gy} stand for correlation coefficient between Y and X in the g^{th} group and the coefficient of variation of X and Y, respectively, in g^{th} group. They also showed that the ratio synthetic estimator (2.2.1.6) is more efficient than that of due to Ghangurde and Singh (1977) if $\rho_g > \frac{1}{2}\frac{c_{gx}}{c_{gy}}$ in each of the groups.

They also evaluated the post-stratified synthetic estimator due to Laake (1978) and the synthetic estimator due to Ghangurde and Singh (1977). They found that latter one is always efficient than the former one.

Nichol (1977) suggested adding the synthetic estimate $\hat{Y}^{s}_{d/r}$, as an additional independent variable in the sample-regression method. This method called the combined synthetic regression method, showed improvement, in empirical studies, over both the synthetic and sample-regression estimates.

Chambers and Finney (1977) and Purcell and Kish (1980) suggested structure-preserving estimation (SPREE) as a generalization of synthetic estimation in the sense it makes a full use of reliable direct estimates. SPREE uses the well-known method of interactive proportional fitting of margins in a multi-way table, where the margins are direct estimates.

Prabhu-Ajgaonkar (1988) suggested a method of small area estimation using the Woodruff's techniques. This method also suggests a kind of predictive approach for the primary stage units which are not included in the sample. The primary stage units (PSU's) were itself considered as a small area.

Wolter and Causey (1991) evaluated across-the-board ratio estimation and synthetic estimation, two techniques that used for improving population estimates for small areas.

Chaudhuri and Adhikary (1995) have also made an empirical study on generalized regression estimators of small areas. They concluded that synthetic generalised regression estimators performed well as compared to other estimators.

Chaudhuri and Maiti (1994, 1996, and 1997) developed the methodology for small area estimation through borrowing strength across time and similar domains. In fact, they made a case study to illustrate different methods of small area estimation such as Empirical Bayes, models with parametric function and Kalman Filter *etc.*

Srivastava *et al.* (1998) have given an overview of small area estimation technique in agriculture. In fact, in agriculture the crop yield at block level are not available and a method has been suggested by them to estimate the crop yield at block level through simulation process. However, this process requires yield at crop cutting experiments and other related information at field level in order to classify the population into homogeneous groups. Such information are, however, not readily available and even sometimes very difficult to get it from the sources responsible for it.

Datta *et al.* (1999) derived theory for multivariate small area estimation and Prasad and Rao (1999) robustified estimation by incorporating the design weights.

Consiglio *et al.* (2003) examined conditional and unconditional properties of some well-known small area estimator: expansion, post-stratified ratio, synthetic, composite, sample size dependent and the empirical best linear unbiased predictor.

Sharma *et al.* (2004) proposed two different types of estimator for small area estimation to estimate the crop yield at the gram panchayat level as an application to crop insurance was undertaken at IASRI information of crop yield on selected field by enquiring from the farmers, which was used judiciously for obtaining correlation factors and revealed that both the estimators perform satisfactorily in terms of the criterion of percentage root mean square error. The approach was akin to SAE in the since that block level estimates were scaled down to gram panchayat level.

Rao (2004) has discussed application of small area estimation like direct, indirect and composite method of estimation in agriculture.

Sinha and Rao (2009) investigated the robustness properties of the classical estimators and propose a resistant method for small are estimation, which is useful for down weighting any influential observations in the data when estimation the model parameters. To estimate the mean squared errors of the robust estimators of small area means, a parametric bootstrap method is adopted here, which is applicable to models with block diagonal covariance structures. Simulations are carried out to study the behaviour of the proposed robust estimators in the presence of outliers, and these estimators are also compared to the EBLUP estimators. Performance of the bootstrap mean squared error estimator is also investigated in the simulation study. The proposed robust method is also applied to some real data to estimate crop areas for counties in Iowa, using farm-interview data on crop areas and LANDSAT satellite data as auxiliary information.

2.2.3. Composite Estimation Method

Synthetic estimators are very sensitive to the assumption that parts of small areas within the same domain resemble each other. Even a small departure from this assumption can make the design-bias of the synthetic estimator very high. Probability samples have, therefore, suggested combining direct and synthetic estimators by using suitable weights. A combined estimator is a weighted average of a design-based direct estimator and a synthetic estimator.

$$\hat{Y}_d^C = \omega_d \hat{Y}_d + (1-\omega_d)\hat{Y}_d^S \qquad (2.2.2.1)$$

where ω_d ($0 \leq \omega_d \leq 1$) is a suitably chosen weight. The aim is to balance the potential bias of the estimator against the instability of design estimator. Many of the estimators reported in the literature, both design-based and model-based, have the form (2.2.2.1).

In the design - based approach, optimal weight is obtained by minimizing MSE ($\hat{Y}_d^C$) with respect to ω_d assuming Cov($\hat{Y}_d, \hat{Y}_d^S$) $\cong 0$:

$$\omega_d\,(opt) = MSE\,(\hat{Y}_d^S) \,/\, [\,MSE\,(\hat{Y}_d^S) + V\,(\hat{Y}_d)] \qquad (2.2.2.2)$$

The optimal weight ω_d (opt) can be estimated by using the MSE ($\hat{Y}_d^S$) given in (2.2.1.5) for the numerator and ($\hat{Y}_d^S - \hat{Y}_d$)^2 for the denominator. The resulting estimator can be very unstable. Schiable (1978) provided an 'average' weighting scheme based on several variables to overcome this difficulty, noting that the composite estimator is quite robust to deviations from $\hat{\omega}_d$ (opt).Purcell and Kish (1979) suggested another approach that uses a common weight, ω and then minimizes the average MSE *i.e.* $\sum$MSE ($\hat{Y}_d^C$)/m, with respect to ω. This leads to estimated weight of the form

$$\omega(\text{opt}) = 1 - \sum_d V\left(\hat{Y}_d\right) / \sum_d \left(\hat{Y}_d^S - \hat{Y}_d\right)^2 \qquad (2.2.2.3)$$

If the variances of $\hat{Y}_d$'s are approximately equal, then it may replace v($\hat{Y}_d$) by the average $\bar{v} = \sum_d v\left(\hat{Y}_d\right)/m$ in which case (2.2.2.3) reduces to the James-Stein type weight:

$$\hat{\omega}(\text{opt}) = 1 - m\bar{v} \;/\; \sum(\hat{Y}_d^S - \hat{Y}_d)^2 \qquad (2.2.2.4)$$

The choice of a common weight is not, however, very reasonable if the individual variances V($\hat{Y}_d$) vary considerably. Also, the James-Stein estimator cannot be less efficient than the direct estimator, $\hat{Y}_d$, for some individual areas if the small areas that are pooled are not similar (Rao and Shinozaki, 1978).

Simple weights, ω_d, that depend only on the realized sample values of covariate x have also been given in the literature. For example, **Drew, Singh and Chaudhary (1982)** suggested the sample-size-dependent estimator that uses the weight:

$$\omega_d(D) = \begin{cases} 1 & \text{if } \hat{N}_d \geq \delta N_d) \\ \hat{N}_d/(\delta(\ _d) & \text{, otherwise} \end{cases} \tag{2.2.2.5}$$

where $\hat{N}_d$ is the direct, unbiased estimator of the known domain population size N_d and δ is subjectively chosen to control the contribution of the synthetic estimator.

This estimator with δ=2/3 and a generalized regression synthetic estimator replacing the ratio synthetic estimator $\hat{Y}_d^S$ is currently being used in the Canadian Labour Force Survey to produce domain estimates. Sarndal and Hiniroglou (1989) suggested an alternative estimator that uses the weight:

$$\omega_d(S) = \begin{cases} 1 & \text{if } \hat{N}_d \geq N_d \\ \hat{N}_d/(N_d)^{l-1}, & \text{otherwise} \end{cases} \tag{2.2.2.6}$$

where l is chosen judgmentally to control the contribution of the bias of the synthetic component. They, however, suggested l = 2 as a general purpose value. Note that the weights (2.2.2.5) and (2.2.2.6) are identical if one chooses δ = 1 and l = 2.

Pandey and Kathuria (1995) developed a model based composite estimators using the models given by Holt *et al.* (1979) for small area estimation. The estimators so developed by them have been empirically tested with a data of sample survey conducted for estimation of milk yield of cows in different districts of Himachal Pradesh state.

2.3. Demographic Methods

For local estimation of population and other characteristics of interest, demographers have been using several methods in post-sensual years. **Purcell and Kish (1980)** grouped these methods under the following headings:

(i) Symptomatic Accounting Technique (SAT)

This procedure is the oldest among all the small area estimation methods. In this method, basic demographic, accounting educations relating births, deaths and migration or change in the population. Other related variables such as number of births and deaths, school enrolments, income tax returns, *etc.* are also accounted for, through relationship between the population growth and these symptomatic variables.

Bogue (1950) reported a generalization of SAT as Vital Rate Technique which uses the changes in birth and death rates rather than the raw values of changes. This method assumes that the ratio of birth (or death) rate for a given small area to the birth (or death) rate for the larger region has remained constant since the last census. The limitation of this method is that it has been used only for studying population counts and depends on good current registers of births, deaths, *etc*.

(ii) Structure Preservation Method of Estimation (SPREE)

Small area estimation methods discussed previously use census information at the larger area level for the purpose of scaling down the current survey estimate at the small area. The method should be adequate if it is assumed that the census information is true and the structure of the population remains same at the proposed inter-censual period. This assumption may not always be true. This small area is an important aspect to be considered for accuracy in the small area estimate at the inter-censual period. Purcell and Kish (1980) described a method initially for the count data and called it as structure preservation method of estimation. This method requires following information:

(a) **The association structure:** Recently census information, which establishes relationship between study variable and a set of associated variables.

(b) **The allocation structure**: The current information needed to update the relationship obtained in the association structure.

This method updates the association structure using the allocation structure with the help of procedure similar to raking ratio method of estimation.

The SPREE estimates are biased to the extent that the underlying association structure, which is derived from the past census results, does not correctly represent the true structure present in the current population.

(iii) Regression Symptomatic Method

The formulation of Regression-symptomatic procedure is based on Marker (1983). The multivariate symptomatic regression method uses the model

$$Y = X\beta + \varepsilon, \qquad \varepsilon \sim N(0, \sigma_1^2)$$

and Y_d , d = 1, 2, …, D are the ratios for the dependent variable of the most recent census to the preceding census, and X_{id} , i = 1, 2, …, P are the ratio for the P symptomatic variables of the most recent census to the preceding census. Using least square estimate $\hat{\beta} = (X'X)^{-1}X'Y$, the estimates $Y^* = X^* \hat{\beta}$ are found, where X* are matrix of ratios for P symptomatic variables of their present values to their values at the time of most recent census. The estimates of change Y_d^* is multiplied by the most recent value for the small area to give the symptomatic regression estimate of the present total for the small area d. The procedure has got two variants in the form of ratio-correlation formulation and difference-correlation formulation. In ratio-correlation formulation, Y_d's and X_d's are defined as ratio or the proportions of small area units to the larger ones over the two time periods (*i.e.* census) while in difference correlation formulation, the corresponding differences of Proportions are defined as Y_d's and X_d's.

(iv) Sample Regression Method

The sample regression method is based on the regression fitted for n sample observations for the ratios of the sample values to the previous census values as dependent variables and the ratios of the current symptomatic variables to their previous census values as independent variables. It is assumed that the same relationship also holds for all those units falling in the small area of interest, say d, and may be expressed as :

$$Y_d = X_d\beta + \varepsilon_d$$

where, Y_d = value of the character under study for small area d, and $\varepsilon_d \sim N(0, \sigma_1^2)$ and its estimator is given by

$$\hat{Y}_{d(reg.)} = X_d\, b$$

where,

$\hat{Y}_{d(reg.)}$ = the regression sample estimate of the character under study for the small area d,

X_{di} = the small area values of i = 1, 2,…, P symptomatic variables, and

b_i = the estimated regression coefficient.

This method is an improvement over regression symptomatic method in the sense that it relies only on the consistency of the relationship between independent and dependent variables over the recent census rather than assuming this type of consistency over the previous two censuses. It can be seen that this method effectively utilises the predictive role of regression analysis. Ericksen (1973) suggested an important empirical result on the selection of symptomatic variables and stratification of PSU's to improve the accuracy of the estimate. Ericksen (1974) evaluated the performance of the small area estimates based on:

(i) the ratio-correlation method

(ii) the sample regression method, and

(iii) the sample–regression method in which ratio correlation estimate is included as an independent variable.

Using 1960 and 1970 census data, Ericksen found that sample-regression method performed almost better than the ratio-correlation method but it was out-performed by the sample-regression method, when ratio-correlation estimate was used as independent variable.

(v) Base Unit Method

It is a common practice to treat the local area units as the smallest level for which the estimates are made. However, **Kalsbeek (1973)** suggested breaking up the local area units into contiguous geographical units called 'base units' such as townships, enumeration districts, *etc.* In Indian context, village can be considered as a suitable example of a base unit, where small area may be Gram Panchayat and C.D. blocks. The local area for which a variable of interest is to be estimated is referred to as the "target area" and further sub-divided into "target area base unit". Unlike other methods which use symptomatic information directly for the purpose of estimation, this procedure uses the information to group base units (sample base units) from the total population. The symptomatic information is also used to classify “target area base units” into the appropriate group.

Firstly, a random sample of n base units is selected from the total population of N base units. The sample base units (possibly including some “target area base units”) are required to possess both symptomatic and criterion information. The base units are divided into groups (post-strata) using either or both types of information available. All “target area base units” are classified to one of the h groups with respect to symptomatic information. An estimate for each of the “target area base units” with respect to the criterion variable is obtained from the sample base units in the group to which it belongs. These estimates are then pooled to get the final estimate for the target area. This method is considered to have more flexibility in terms of utilizing the symptomatic information.

2.4. Model Based techniques based on sample surveys data available for larger area

(i) Small Area Models

Historically, small area statistics having long been used in sample survey. Even such statistics existed in the eleventh century, England and seventeenth century Canada, based on either census or on administrate records (Brackstone, 1987). The SAE problem has received a lot of attention in recent years due to growing demand for reliable small area statistics. Traditional area-specific direct estimators do not provide adequate precision because sample size in small area seldom large enough.

This makes it necessary to employ indirect estimators that borrow strength from related areas, in particular, model-based indirect estimators.

Considering that area-specific auxiliary data $X_d = (x_{d1}, x_{d2}, \ldots, x_{dp})^T$ are available and the parameters of interest, θ_d, are assumed to be related to X_d. The model assumed is

$$\theta_d = X_d^T \beta + v_d z_d , \quad (d = 1,2,\ldots,D) \qquad (2.4.1)$$

where the z_d's are known positive constants, β is the vector of regression parameters and the v_d's are independently and identically distributed (i.i.d.) random variables with $E(v_d)=0$ & $V(v_d)=\sigma_v^2$.

In an element-specific auxiliary data $X_{dj} = (x_{dj1}, x_{dj2}, \ldots, x_{djp})^T$, for the population elements, and the variable of interest, y_{dj}, a nested error regression model

$$Y_{dj} = X_{dj}^T \beta + v_d + e_{dj} , \quad j = 1, 2, \ldots, N_d \qquad (2.4.2)$$

$$d = 1, 2, \ldots, D$$

Here $e_{dj} = \tilde{e}_{dj}\, l_{dj}$ and the $\tilde{e}_{dj}$'s are i.i.d. random variables, independent of the v_d's with $E(\tilde{e}_{dj}) = 0$ & $V(\tilde{e}_{dj}) = \sigma^2$.

The l_{dj} 's being known constants and N_d the number of elements in the d^{th} small area.

The parameters of inferential interest here are the small area totals Y_d or the means $\bar{Y}_d = Y_d / N_d$.

For making inferences about the θ_d's under model (2.4.1), we assume that direct estimators $\hat{\theta}_d$ are available and that

$$\hat{\theta}_d = \theta_d + e_d \qquad (2.4.3)$$

where e_d's are sampling errors, E (e_d/θ_d) = 0 and $V(e_d/\theta_d) = \psi_d$ *i.e.* the estimator $\hat{\theta}_d$ are design-unbiased. It is also customary to assume that the sampling variances, ψ_d, are known. For example, in the case of adjustment for census under enumeration, the estimates, θ_d, obtained from a post-enumeration survey (PES) could be seriously biased, as noted by Freedman and Navidi (1986).

Combining (2.4.3) and (2.4.1), the model becomes

$$\hat{\theta}_d = X_d^T \beta + v_d z_d + e_d \qquad (2.4.4)$$

which is a special case of the general mixed linear model. Note that (2.4.4) involves design-induced random variables, e.g., as well as model-based random variables v_d.

Fay and Herriot (1979) considered basic area level model for small area estimation. Later, this model is called Fay-Herriot model in the small area literature because they were first to introduce such model for small area estimation.

Fay (1987) advocated a multivariate approach to small area estimation to produce reliable estimates of median incomes for four-person families for the fifty states of U.S. and the districts of Columbia by including information from three- and fire person families as well. Fuller and Harter (1987) also considered this approach for estimating a finite population mean vector of multiple characteristics for each small area.

Prasad and Rao (1990) obtained a second order approximation to the mean square error of the small area estimator, based on a general linear mixed model with a block diagonal covariance structure.

Srivastava (2009) worked on estimating poverty at small area level, *i.e.* at district level. He estimated the indicators of poverty and economic inequality using Poverty Ratio, Poverty Gap Ratio, Squared Poverty Gap Ratio, Foster-Greer-Thorbecke (FGT) Index and also a measure of Income Inequality with the help of Gini Coefficient, poverty estimation. He emphasised on small area estimation approach with following procedure. He considered some explicit model-based methods which are essentially mixed models and are used in specific situations based on data availability on the response variables of interest. These are (i) area level models where information on response variable is available only at the small area level; and (ii) unit level models where information on the response variable is available at the unit level.

An area level mixed model is represented as

$$\hat{\theta}_d = z_d^T\beta + v_d + e_d, \quad d = 1..., D$$

where $\hat{\theta}_d$ is the direct survey estimate of the parameter θ_d, as obtained from the sample survey data, z_d^T is the vector of concomitant variates, the model errors $v_d$are assumed to be independent and identically distributed with mean zero, variance σ_d^2 and e_d are the sampling errors which are assumed to be independent across small areas with mean zero and known variances χ_d. Here e_d and χ_d are design-based and model-based random variables respectively. The model variance σ_d^2 is a measure of homogeneity of the areas after accounting for the covariates zd.

For the mixed model considered above, an Empirical Best Linear Unbiased Predictor (EBLUP) of θ_d is given by

$$\theta_d^* = \gamma_d \hat{\theta}_d + (1-\gamma_d) z_d^T \hat{\beta}$$

This estimator is a linear combination of direct estimator $\hat{\theta}_d$ and the regression synthetic estimator $z_d^T \hat{\beta}$. Further γ_d and $\hat{\beta}$ are defined as follows:

$$\gamma_d = \frac{\hat{\sigma}_v^2}{\Psi_d + \hat{\sigma}_v^2} \text{ and } \beta = \left[\sum_d^D \frac{z_d z_d^T}{\Psi_d + \hat{\sigma}_v^2} \right]^{-1} \left[\sum_d^D \frac{z_d \hat{\theta}_d}{\Psi_d + \hat{\sigma}_v^2} \right]$$

and variance

$$\sigma_v^{2(\alpha+1)} = \sigma_v^2 + \frac{1}{h^{'}(\sigma_v^{2(\alpha(})} \left[m - p - h(\sigma_v^{2(\alpha(}) \right]$$

Estimation of MSE (θ_d^*) for small areas which are in the sample

$$\text{mse}(\theta_d^*) = g_{1d}(\hat{\sigma}_v^2) - b_{\hat{\sigma}_v^2}(\hat{\sigma}_v^2)\Delta_{g_{1d}}(\hat{\sigma}_v^2) + g_{2d}(\hat{\sigma}_v^2) + 2g_{2d}(\hat{\sigma}_v^2)$$

where, $g_{1d}(\hat{\sigma}_v^2) = \gamma_d \Psi_d$, $g_{2d}(\hat{\sigma}_v^2) = (1-\gamma_d)^2 X_d^T \left[\sum_d^D \frac{z_d z_d^T}{\Psi_d + \hat{\sigma}_v^2} \right]^{-1} z_d$

and $g_{3d}(\hat{\sigma}_v^2) = \Psi_d^2 (\Psi_d + \hat{\sigma}_d^2)^{-3} \overline{V}(\hat{\sigma}_v^2)$

where $\overline{V}(\hat{\sigma}_v^2) = 2m \left[\sum_i \frac{1}{\Psi_d + \hat{\sigma}_v^2} \right]^{-2}$

$$b_{\hat{\sigma}_v^2}(\hat{\sigma}_v^2) = \frac{2\left[m\sum_d (\Psi_d + \hat{\sigma}_d^2)^{-2} - \left\{\sum_d (\Psi_d + \hat{\sigma}_d^2)^{-1}\right\}^2\right]}{\left[\sum_d (\Psi_d + \hat{\sigma}_d^2)^{-1}\right]^3}$$

and $\nabla_{g1d}\left(\hat{\sigma}_v^2\right) = \left(1 - \gamma_d\right)^2$

Estimation of MSE (A_i^*) for small areas which are not in the sample

$$mse(\theta_{d'}^*) = z_{d'}^T\left[\sum_i \frac{z_d z_d^T}{\Psi_d + \hat{\sigma}_v^2}\right]^{-1} z_{d'}\hat{\sigma}_v^2$$

(ii) Unit level models

$$y_{dj} = x_{dj}^T\beta + v_d + e_{dj'} j = 1,..., N_d; \; d = 1,..., D$$

The analysis carried out here indicates that it is feasible to estimate poverty indicators at district level by scaling down the State level poverty estimates utilizing small area estimation techniques. The choice of SAE model and corresponding variables is crucial for successful application of the SAE method. In the process of application of SAE method, it was realized that still there is enough scope for the choice of variables. Efforts for improving the estimates and to apply it for other States, is in process. If unit level data from census may be available, then other methods for poverty mapping may also be attempted. Estimates for poverty indicators are based on consumption expenditure survey data. It may be worthwhile to examine the poverty estimates based on income data, if reliable information on income may be obtained from household surveys. The distributions of expenditure and income are likely to differ and the differences should depend on the income expenditure levels of households.

The socio-economic and spatial factors may also contribute towards the variability in the income expenditure patterns. One of the sources for household level income data is the surveys conducted by NCAER. SAE methods have got an important role to play in disaggregated estimates at small area levels. Estimation of trends and changes are important in poverty studies. Comparability of results sometimes poses serious problems. One of the consumption expenditure surveys (55th round) of NSSO is an example, in which an attempt was made to try different reference periods for different items. The idea was to take care of recall lapse and improve the quality of data, but there are problems in comparability of results with other rounds.

(iii) Best Linear Unbiased Prediction Approach

Holt *et al.*, (1979) obtained a best linear unbiased prediction (BLUP) estimator of Y_d under the following model for the finite population:

$$y_{dgk} = \mu_g + e_{dgk} \qquad (2.4.5)$$

$$k = 1, 2, \ldots, N_{dg}; \quad g = 1, 2, \ldots, G; \quad d = 1, 2, \ldots, D.$$

where y_{dgk} is the value of the k^{th} unit in the cell (d, g), μ_g 's are fixed effects and the errors e_{dgk} are uncorrelated with zero means and variances σ_d^2. Let us consider n_{dg} be an elements in a sample of size n fall in cell (d, g) and let $\bar{y}_{dg}$ and $\bar{y}_{.g}$ denote the sample mean for (d, g) and g, respectively.

The best linear unbiased estimator μ_g under (5.4.5) is $\hat{\mu}_g = \bar{y}_{.g}$ which in turn leads to the BULP estimator of Y_d given by

$$\hat{Y}_d = \sum_g \hat{Y}^C_{dg} = \sum_{g=1}^{G} N_{dg}[(n_{dg}/N_{dg}).\bar{y}_{dg} + (1 - n_{dg}/N_{dg}).\bar{y}_{.g}] \qquad (2.4.6)$$

where $\hat{Y}_d$ is a called predictor for Y_d. It is also a model based synthetic predictor.

The model variance of the $\hat{Y}_d$ is

$$V_M(\hat{Y}_d) = \sigma^2 \sum_{g=1}^{G} N^2_{dg} / n_{.g}, \text{ for } N_{dg} > n_{dg} \qquad (2.4.7)$$

The estimator for σ^2 is

$$\hat{\sigma}^2 = \sum_d \sum_g \sum_k (y_{dgk} - y_{.g})^2/(n - G) \qquad (2.4.8)$$

Since the predictors heavily depend on the model assumed, their performances are seriously affected when assumed predictor model does not represent the characteristics of the population of interest. Cassel, Sarndal and Wretman (1976) suggested an alternative called method model-assisted approach. This approach combines sampling design and model both in getting optimal estimator.

Considering small area-specific model

$$Y_{dk} = \beta_d X_{dk} + e_{dk} \qquad (2.4.9)$$

$k = 1, 2, \ldots, N_d;\quad d = 1, 2, \ldots, D.$

where $E(e_{dk})=0$, $V(e_{dk})= \sigma^2_d$ and co-variances are assumed to be zero.

Sarndal (1981) proposed the following predictor for Y_d

$$\hat{Y}_d = \sum_{k=1}^{n_d}(y_{dk} / \pi_k + (X_d - \sum_{k=1}^{n_d} x_{dk}/ \pi_k)\hat{\beta}_d \tag{2.4.10}$$

This is called as generalised regression predictor ($G_{reg.}$).

$$\hat{\beta}_d = \sum_{k=1}^{n}(y_{dk}x_{dk}I_{dk}\theta_k \;/\; \sum_{k=1}^{n}x_{dk}^2\theta_k I_{dk} \tag{2.4.11}$$

θ_k are arbitrary assignable choice of constant and I_{dk} is random variable with value 1 if i^{th} unit belongs to the small area U_d and 0 otherwise. If, however, each small area has same slope, *i.e.*$\beta_d=\beta$, then by borrowing strength across the small area, d. Synthetic $g_{reg.}$ predictor is given by

$$\hat{Y}'_d = \sum_{k=1}^{n_d}(y_{dk} /\pi_k) + (X_d - \sum_{k=1}^{n_d}(x_{dk} /\pi_k)\, \hat{\beta} \tag{2.4.12}$$

where $\hat{\beta} = \sum_{d=1}^{D}\sum_{k=1}^{n_d}(y_{dk}\; x_{dk}\, \theta_k) / \sum_{d=1}^{D}\sum_{k=1}^{n_d}(x_{dk}^2\theta_k)$

A composite estimator by combining $\hat{Y}_d$ and $\hat{Y}'_d$ is given by

$$\hat{Y}_{\theta d} = \theta\hat{Y}_d + (1-\theta)\hat{Y}'_d \tag{2.4.13}$$

where θ should be in [0, 1] , so that $\hat{Y}_{\theta d}$ may be a convex combination of $\hat{Y}_d$ and $\hat{Y}'_d$. The optimal value of θ is given by

$$\theta_{opt.} = \frac{V_p(\hat{Y}_d) - C_p(\hat{Y}_d, \hat{Y}'_d)}{V_p(\hat{Y}_d) + V_p(\hat{Y}'_d) - 2C_p(\hat{Y}_d, \hat{Y}'_d)} \tag{2.4.14}$$

Generally, $\theta_{opt.}$ is unknown as it involves unknown parameters. The best choice of $\theta_{opt.}$ is studied by Schaible (1978) and Chaudhuri and Maiti (1996).

Battese *et al.* (1988) proposed and applied the nested error regression model for estimation of corn and soybean in Iowa State and showed that the mean is sum of a fixed component, involving unknown parameters to be estimated and a random component to be predicted and it is called Best Linear Unbiased Predictor. They also defined variance-component estimators in the nested-error model and obtained generalized least square estimators of the parameters of the linear model. Beside this, they constructed an estimator of the variance of the error in the predictor, including terms arising from the estimation of the parameters of the model.

A paper by **Ghosh and Rao (1994)** and **Rao (2003)** provide an excellent account of various model based on small area estimation techniques.

Chambers and Tzavidis (2006) described a new approach to small area estimation that is based on modelling quantile-like parameters of the conditional distribution of the target variable given the covariates. This avoids the problems associated with specification of random effects, allowing inter-area differences to be characterised by area-specific M-quantile coefficients. The proposed approach is easily made robust against outlying data values and can be adapted for estimation of a wide range of area-specific parameters, including quantiles of the distribution of the target variable in the different small areas. The difference between M-qunatile and random effects models were discussed and the alternative approaches to small area estimation are compared using both simulated and real data.

Jiang and Lahiri (2006) has given an excellent overview and appraisal of mixed model prediction in the context of small area estimation.

Chandra *et al.* (2007) investigated the SAE based on linear models with spatially correlated small area effects where the neighbourhood structure is described by a contiguity matrix. Such models allow efficient use of spatial auxiliary information in SAE. In particular, they use simulation studies to compare the performances of model-based direct estimation (MBDE) and empirical best linear unbiased prediction (EBLUP) under such models. These simulations are based on theoretically generated populations as well as data obtained from two real populations (the ISTAT farm structure survey in Tuscany and the US Environmental Monitoring and Assessment Program survey). The results of this study an initial exploration of the use of unit level models with spatially correlated area effects in small area estimation. In particular, they show how the EBLUP and MBD methods of estimation can be adapted for this situation. However, their empirical results based on both real data as well as on simulated data under the spatial model, indicate that the gains from inclusion of spatial structure in small area estimation do not appear to be large. This is especially true for model-based direct estimation based on this structure (SMBDE), where the extra spatial information seems to have very little impact on the distribution of the SEBLUP weights that characterise this method of estimation.

Salvati *et al*. (2007) investigated the use of GWR in small area estimation based on the M-quantile GWR model described in their study allows modeling of between area variability without the need to explicitly specify the area-specific random components of the model and this model does not explicitly depends on any particular small area geography, and so can be easily adapted to different geographies as the need arises.

Srivastava *et al.* (2007) worked on small area estimation with National Sample Survey Data. They used already available small area estimation techniques to derive district level estimates of amount of loan outstanding per household using data from the 2002-03 Debt-Investment Survey of National Sample Survey Organization (NSSO) for the rural areas of Uttar Pradesh. Fay and Herriot model (Fay and Herriot, 1979) has been used to obtain the model-based district level estimates. The diagnostic analysis shows that the model-based estimates are reasonably reliable and representative of the districts to which they belong. The model-based method has been found to be very effective for developing district level estimates of average amount of loan outstanding per household. For most of the district's the reduction in coefficient of variation is quite evident. However, the diagnostics results presented in previous section show only marginal gains in the model based estimates. This was expected since we used 1995-96 Agriculture Census data (the latest census data was not available) for collecting information on the covariates. Due to this we could not get very high correlation between the study variable and the covariates. We already indicated that the success of model-based SAE methods lie in the correct specification of the underlying model and availability of good covariates. This possibly explains the aberration in the diagnostics results.

Pratesi and Salvati (2008) proposed the small area indirect estimators under area level random effect models when only area level data are available and the random effects are correlated. The performance of the Spatial Empirical Best Linear Unbiased Predictor (SEBLUP) has explored with a Monte Carlo simulation study on lattice data and it was applied to the results of the sample survey on Life Conditions in Tuscany (Italy). The mean squared error (MSE) problem was discussed illustrating the MSE of estimator in comparison with the MSE of the empirical sampling distribution of SEBLUP estimator. A clear tendency in their empirical findings was that the introduction of spatially correlated random area effects reduces both the variance and the bias of the EBLUP estimator.

Chandra (2009) estimated the small area proportions under unit level spatial models. They investigate Small Area Estimation based on generalized linear mixed model (GLMM) with spatially correlated random area effects where the neighbourhood structure is described by a contiguity matrix. They used simulation studies to compare the performance of empirical best predictor for small area proportions under such models with and without spatially correlated area effects. The simulation studies are based on two real data sets. The results show only marginal gains when spatial dependence between small areas is incorporated into the SAE model.

Chandra and Chambers (2009) described how such multipurpose sample weights can be constructed when small area estimates of the survey variables are required. The approach is based on the model-based direct (MBD) method of small area estimation described in Chandra and Chambers (2005). Empirical results reported that MBD estimators for small areas based on multipurpose weights perform well across a range of variables that are often of interest in business surveys. Furthermore, these results show that the proposed approach is robust to model misspecification when applied to variables (e.g., those that contain a significant proportion of zeros) that are not suited to linear model-based small area estimation methods. They develop two loss functions that can be used to compute optimal multipurpose weights suitable for use in small area estimation using MBD estimators. They also investigated an alternative "weight smoothing" approach to constructing multipurpose weights. A final comment concerns the EBLUP is the most efficient linear estimator provided its model assumptions hold. However, this efficiency was not evident in simulations, in all probability because this estimator's underlying model assumptions fail when applied to the AAGIS data. Their concern were to demonstrate that combining a simple linear method

(MBD) of model-based small area estimation with a multipurpose approach to sample weighting leads to a robust estimation method, *i.e.*, one that works well in a variety of situations, even those where underlying model assumptions are approximate at best.

Johnson *et al.* (2010) studied the uses of small area estimation techniques to derive model-based district-level estimates of institutional births in Ghana by liking data from the 2003 GDHS and the 2000 Population and Housing Census. The models indicate considerable variability in the estimates, with institutional births ranging between 7% and 27% in the districts of the Northern region, compared to 78% and 85% in the districts of the Greater Accra Region. The diagnostic measures indicate that the model-based estimates are reliable and representative of the district to which they belong.

2.5. Estimation / prediction approach using time series or primary data available at larger area by applying suitable regression model

Stasny *et al.* (1991) have made use of regression model to predict wheat production at county level in USA. They, in fact, developed the multiple regression models to exhibit best possible relationship between farm production and some predictor variables related with farm production for predicting wheat production at county level However, their approach was based on cross-sectional data.

Crop production/yield are available through CCEs at district level. Conducting large number of CCEs to estimate crop yield at Block level is not viable proposition in view of lack of basic infrastructures, huge cost involved and accumulation of large non-sampling errors that are likely to creep in. Sisodia and Singh (2001) first time proposed a scale down approach using time series data on yield/production and related covariates by applying multiple regression model to obtain the block level estimate from the district level crop-production estimate. The methodology is briefly discussed below.

They assumed a multiple regression model between the crop production and auxiliary variables at district level, i.e.

$$Z_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_p X_{ip} + \varepsilon_i \qquad (2.5.1)$$

where $Z_i$is the crop production in the year i(i =1,.....,*n*) , $X_{ij}$is the value of auxiliary variable j(j =1,....., p) in the year i, β=(β_0, β_1, …., β_p)' is vector of unknown model parameters and $e_i$is error term assumed to follow independently normal distribution with mean 0 and variance σ^2. The fitted model with dates by ordinary least square is given by

$$\hat{Z}_i = \beta_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + ... + \hat{\beta}_p X_{ip} + \varepsilon_i \qquad (2.5.2)$$

where $\hat{\beta}$ is a least square estimate of β and $\hat{Z}_i$ is the estimated value of $Z_i$for corresponding values of X_{ij}'s in the year i. The sum of squares due to regression, *i.e.*, $SS_R(\beta_1, \beta_2, …., \beta_p|\beta_0)$ is decomposed to define a weight that determines the relative contribution of each predictor variable included in the model(see Montegomery & Peck,1982) as follows:

$$w_j = \frac{\text{SS due to } j^{th} \text{ predictor}}{SS_R\ (\beta_1, \beta_2, ..., \beta_p | \beta_0)} \qquad (2.5.3)$$

It is assumed that the information on X_{ij}. ij are also available at Block level using these weights, an estimator of crop production Z_q for the block q is constructed for a given year as :

$$\hat{Z}_q = \left[\sum_{j=1}^{P} w_j x_j\right] \hat{\bar{Z}};\ q = 1, \ldots Q \qquad (2.5.4)$$

where *Q* is total number of Block in a given District, x_j is the value of j-*th* predictor at Block level in a given year, $\hat{\bar{Z}} = \hat{Z}/A$ and $\hat{Z}$ is obtained through the fitted model (2.5.5) and *A* is the area under the crop in a given year. Note that the weight w_j depends on the set of data on Z and X_{ij} used to fit the model (2.5.1). The estimator $\hat{Z}_q$ is an unbiased estimator of Z_q if $\sum_{j=1}^{p} w_j x_j$ is considered to be a constant quantity for a given Block since under model (2.5.1) expected value of $\hat{\bar{Z}}$ is $\bar{Z}$. The variance of $\hat{Z}_q$ is given by

$$V(\hat{Z}_q) = (\delta_q / A)^2 V(\hat{Z}) \qquad (2.5.5)$$

where $\delta_q = \sum_{j=1}^{p} w_j x_j$. The variance of $\hat{Z}$ is easily available from the fitting model (2.5.2), which is equal to $\hat{\sigma}^2$, the estimated error variance. It is obvious that, in general, $\sum_{q=1}^{Q} \hat{Z}_q \neq Z$, where Z is the actual crop production reported at District level through the CCEs in a given year.

Thus a scaled estimator of Z_q was suggested by them as follows

$$\tilde{Z}_q = a_q \hat{Z}_q \tag{2.5.6}$$

such that $\sum_{q=1}^{Q} \tilde{Z}_q = \sum_{q=1}^{Q} a_q \hat{Z}_q = Z$, where a_q are constant. There can be three alternative choices of a_q. The third choice was due to Sisodia and Singh (2012) and Sisodia and Chandra (2012). These choices are in fact different adjustments which are described below.

Choice- I: Ratio Adjustment

The simplest choice of a_q is to take $a_q = a$, *i.e.* the constant for each block. It can easily be shown that

i.e. $$a \sum_{q=1}^{Q} \hat{Z}_q = Z \tag{2.5.7}$$

Thus, a new scaled estimator of Z_q is given by,

$$\tilde{Z}_q^{(1)} = \hat{Z}_q \left[Z / \sum_{q=1}^{Q} \hat{Z}_q \right] \tag{2.5.8}$$

Obviously, the estimator $\tilde{Z}_q^{(1)}$ is not an unbiased estimator of Z_q as $E\left[\tilde{Z}_q^{(1)}\right] = Z_q \left[Z / \sum_{q=1}^{Q} \hat{Z}_q \right]$, because $Z / \sum_{q=1}^{Q} \hat{Z}_q$ is constant for each block q, q=1,2,…,Q.

Choice- II: Least square adjustment

Another choice of a_q could be one that minimizes the sum of squared difference between $\tilde{Z}_q$ and $\hat{Z}_q$ subject to condition that $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$.

To obtain a_q which minimizes the sum of squared difference between $\tilde{Z}_q$

and $\hat{Z}_q$ subject to condition $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$, we must minimize $\sum_{q=1}^{Q} (\tilde{Z}_q - \hat{Z}_q)^2$ with respect to a_q using a Lagrange multiplier, λ, to impose the desired constraints. This minimisation leads to

$$a_q = \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{Q\hat{Z}_q} + 1 \qquad (2.5.9)$$

and resultant new scaled estimator of Z_q is given by,

$$\tilde{Z}_q^{(2)} = \hat{Z}_q + \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{Q} \qquad (2.5.10)$$

Infact, $\tilde{Z}_q^{(2)}$ is an unbiased estimator of Z_q.

Choice- III: Relative least square adjustment

The third choice of a_q could be one that minimizes the sum of square of relative differences $\left(\tilde{Z}_q - \hat{Z}_q\right)/\hat{Z}_q$ subject to condition that $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$. To obtain a_q which minimizes $\sum_{q=1}^{Q} \left[\left(\tilde{Z}_q - \hat{Z}_q\right)/\hat{Z}_q\right]^2$ subject to $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$, we minimize the following function with respect to a_q. This leads to

$$a_q = 1 + \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{\sum_{q=1}^{Q} \hat{Z}_q^2} . \hat{Z}_q \qquad (2.5.11)$$

and resultant new scaled estimator of Z_q is given by

$$\tilde{Z}_q^{(3)} = \hat{Z}_q + \frac{\hat{Z}_q^2}{\sum_{q=1}^{Q} \hat{Z}_q^2} \left(Z - \sum_{q=1}^{Q} \hat{Z}_q \right) \qquad (2.5.12)$$

It may also be noted that $\tilde{Z}_q^{(3)}$ is not an unbiased estimator of Z_q as

$$E\left[\tilde{Z}_q^{(3)}\right] = Z_q + E\left[\frac{\hat{Z}_q^2}{\sum_{q=1}^{Q}\hat{Z}_q^2}\left(Z - \sum_{q=1}^{Q}\hat{Z}_q\right)\right] \qquad (2.5.13)$$

Evidently, the second term in right hand side of (2.5.13) may not be zero. Thus, $\tilde{Z}_q^{(3)}$ is in general a biased estimator of Z_q. Up to first degree of approximation (Sukhatme and Sukhatme, 1970), the second term in (2.5.13) is, however, almost zero because the ratio $\frac{\hat{Z}_q^2}{\sum_{q=1}^{Q}\hat{Z}_q^2}$ is also expected to be very close to zero. Thus $\tilde{Z}_q^{(3)}$ is also an almost unbiased estimator.

The conditional bias and conditional MSE/ Variance

The bias in $\tilde{Z}_q^{(1)}$ is obtained as

$$\text{Bias}\left[\tilde{Z}_q^{(1)}\right] = E\left[\tilde{Z}_q^{(1)}\right] - Z_q$$

$$= \frac{\left(Z - \sum_{q=1}^{Q}\hat{Z}_q\right)Z_q}{\sum_{q=1}^{Q}\hat{Z}_q} \qquad (2.5.14)$$

MSE of $\tilde{Z}_q^{(1)}$ is given by

$$MSE\left[\tilde{Z}_q^{(1)}\right] = \left[\text{Bias}\left(\tilde{Z}_q^{(1)}\right)\right]^2 + V\left(\tilde{Z}_q^{(1)}\right)$$

$$MSE\left[\tilde{Z}_q^{(1)}\right] = \frac{\left(Z - \sum_{q=1}^{Q}\hat{Z}_q\right)^2 Z_q^2}{\left(\sum_{q=1}^{Q}\hat{Z}_q\right)^2} + \frac{Z^2}{\left(\sum_{q=1}^{Q}\hat{Z}_q\right)^2}V\left(\hat{Z}_q\right) \qquad (2.5.15)$$

Variance of $\tilde{Z}_q^{(2)}$ is given by

$$V\left(\tilde{Z}_q^{(2)}\right) = \frac{Q-2}{Q} V\left(\hat{Z}_q\right) + \frac{1}{Q^2} \sum_{q=1}^{Q} V\left(\hat{Z}_q\right) \qquad (2.5.16)$$

The estimate of variance of $\tilde{Z}_q^{(2)}$ can easily be obtained by computing $V\left(\hat{Z}_q\right)$ from the fitted model.

Bias & Mean Square Error (MSE) of $\tilde{Z}_q^{(3)}$

The bias of $\tilde{Z}_q^{(3)}$ can directly be written from the expression (2.5.13) as

$$\text{Bias}\left[\tilde{Z}_q^{(3)}\right] = E\left[\tilde{Z}_q^{(3)}\right] - Z_q$$

$$= E\left[\frac{\hat{Z}_q^2}{\sum_{q=1}^{Q} \hat{Z}_q^2} \left(Z - \sum_{q=1}^{Q} \hat{Z}_q \right)\right] \qquad (2.5.17)$$

Evidently, the bias will be close to zero if $\sum_{q=1}^{Q} \hat{Z}_q$ is close to Z. If $Z \cong \sum_{q=1}^{Q} \hat{Z}_q$, then approximate MSE of $\tilde{Z}_q^{(3)}$ is given by

$$MSE\left[\tilde{Z}_q^{(3)}\right] = \left[1 + \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{\sum_{q=1}^{Q} \hat{Z}^2{}_q} \hat{Z}_q \right]^2 V\left(\hat{Z}_q\right) \qquad (2.5.18)$$

Comparison for relative efficiency of $\tilde{Z}_q^{(1)}, \tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ over, $\hat{Z}_q$

Taking difference of $V(\hat{Z}_q)$ and $MSE(\tilde{Z}_q^{(1)})$, i.e. $V(\hat{Z}_q)$ - $MSE(\tilde{Z}_q^{(1)})$ and after simplification, we get the inequality

$$V(\hat{Z}_q) \geq \frac{\left(z - \sum_{q=1}^{Q} \hat{z}_q\right)^2 z_q^2}{\left[\left(\sum_{q=1}^{Q} \hat{z}_q\right)^2 - z^2\right]} \tag{2.5.19}$$

for $\tilde{Z}_q^{(1)}$ to be efficient than $\hat{Z}_q$. However, it seems that inequality (2.5.19) may not hold true in general.

Taking difference of $V(\hat{Z}_q)$ and $V(\tilde{Z}_q^{(2)})$, i.e. $V(\hat{Z}_q)$-$V(\tilde{Z}_q^{(2)})$, and after simplification, we get the following inequality

$$V(\hat{Z}_q) \geq \frac{1}{2Q} \sum_{q=1}^{Q} V(\hat{Z}_q) \tag{2.5.20}$$

for $\tilde{Z}_q^{(2)}$ to be efficient than $\hat{Z}_q$. It is obvious from the inequality (2.5.20) that the $V(\hat{Z}_q)$ must be greater than the half of the average of variances of $\hat{Z}_q$ (q=1, 2, … Q), which is quite possible in general.

Taking difference of $V(\hat{Z}_q)$ and $MSE(\tilde{Z}_q^{(3)})$, i.e. $V(\hat{Z}_q)$ - $MSE(\tilde{Z}_q^{(3)})$, we get the following inequality

$$1 - \left\{1 + \frac{z - \sum_{q=1}^{Q} \hat{z}_q}{\sum_{q=1}^{Q} \hat{z}_q^2} \hat{z}_q\right\}^2 \geq 0 \tag{2.5.21}$$

which is not possible as the value of LHS of the above inequality will always be negative. This shows that $\hat{Z}_q$ will always be efficient than $\tilde{Z}_q^{(3)}$ in general.

The above comparisons clearly indicate that the $\tilde{Z}_q^{(2)}$ is expected to perform better than $\hat{Z}_q, \tilde{Z}_q^{(1)}$ and $\tilde{Z}_q^{(3)}$.

The empirical study conducted for rice production in Faizabad district (U.P.) by **Sisodia and Singh (2001)** and for wheat production in Barabanki district (U.P.) by **Sisodia and Singh (2012)** and **Sisodia and Chandra (2012)** showed that the block estimate of crop production based on the estimator $\tilde{Z}_q^{(2)}$ was most precise as compared to other estimators. It can observed that the weight w_j used in $\hat{Z}_q$ is subjective one as it depends upon the set of Z and X_j's chosen for the fitting of the model (2.5.1). Sisodia and Chandra (2012) has, however, advocated for stable value of w_j obtained through iterative procedure of fitting of the model (2.5.1).

Singh *et al*. (2012) proposed a predictive approach of estimation of crop-production at block level using the model (2.1) fitted at district level. Their procedure is described briefly here.

A linear multiple regression models between the crop yield and independent auxiliary variables at district level can be specified as given below

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\ldots+\beta_pX_{ip}+\varepsilon_i \qquad (2.5.22)$$

where Y_i is the crop yield in the i^{th} year, (i= 1, 2, …n), X_{ij} is the value of j^{th} predictor (j=1, 2, … , P) in the i^{th} year, $\beta' = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$ is vector of unknown parameters of the model and ε_i is error term. It is also assumed that ε_i's follow independently normal distribution with mean 0 and variance σ^2. For example, X_{ij}'s may be percent irrigated area under a crop, relative area under the crop as percentage of the gross-cropped area (GCA), fertilizer consumption in kg per hectare for the crop, etc. The model (2.5.22) is fitted with the data at district level and let the fitted model be denoted as

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_p x_{ip} \qquad (2.5.23)$$

where $\hat{\beta}$ is the least square estimate of β and $\hat{Y}_i$ is the estimated value of Y_i for corresponding values of X_{ij}'s.

The fitted model (2.5.23) is then used directly at block level to predict the Z_q, the yield at block q (q=1, 2, …Q). It is well known from the theory of linear regression models that prediction is always considered to be more precise when the values of X_j's corresponding to which the prediction is to be made is within the ranges of X_j's used of fit the model (3.2.24). It is generally expected that percent irrigated area under the crop, fertilizer consumption (kg/ha), and relative area under the crop as percent to GCA at district and block level are almost within a similar range.

Let $\hat{Y}_q$ be the predicted value of the crop yield Y_q (Y_q= yield per unit of area) for the block q during a particular year of interest, obtained from the fitted model (2.5.23) as

$$\hat{Y}_q = \hat{\beta}_0 + \hat{\beta}_1 x_{q1} + \ldots + \hat{\beta}_p x_{qp} \qquad (2.5.24)$$

where X_{qj}'s; j= 1, 2, …. p, are values of predictors in the q^{th} block. Therefore, an unbiased estimator of Z_q, the crop production at q^{th} block is given by

$$\hat{Z}_q = \delta_q \hat{Y}_q \qquad (2.5.25)$$

where δ_q is the area under the crop at q^{th} block. It may be noted that the area under a given crop is being reported through complete enumeration by Revenue Department of State Government in India. Note that $\hat{Z}_q$ is an unbiased estimator of Z_q as $E(\hat{Z}_q) = Z_q$ because $E(\hat{Y}_q) = Y_q$ under the assumption of regression model (2.5.22).

Following **Montgomery and Peck (1982)** the variance of $\hat{Z}_q$ is given by

$$V(\hat{Z}_q) = \delta_q^2 \sigma^2 C'_q (X'X)^{-1} C_q \qquad (2.5.26)$$

where σ^2 is the residual variance corresponding to the regression model 2.5.22), X is matrix of X_{ij}'s at district level given by

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & . & . & . & X_{1p} \\ 1 & X_{21} & X_{22} & . & . & . & X_{2p} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & X_{n1} & X_{n2} & . & . & . & X_{np} \end{bmatrix}$$

and X' is the transpose of X matrix. C_q is column vector of X_{qi}'s at q^{th} block level given by

$$C'_q = [1 \;\; X_{q1} \;\; X_{q2} \;\; \dots \;\; X_{qp}]$$

It is obvious that in general $\sum_{q=1}^{Q}\hat{Z}_q \neq Z$, where Z is the actual crop production reported at district level through crop cutting experiment in a given year. Thus, a new scaled estimator Z_q is given by

$$\tilde{Z}_q = a_q \hat{Z}_q \qquad (2.5.27)$$

where a_q is constant such at $\sum_{q=1}^{Q}\tilde{Z}_q = Z$ or $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$

The choices of a_q and resultant seated estimators along with their conditional variance/MSE will be same as described earlier. The only difference would be in the values of $\hat{Z}_q$ and variance of $\hat{Z}_q$ as these are based on prediction approach.

Ahamad *et al* (2014) used cross section data of farmers and weather variables to obtained wheat production data at Taluka level in karnatka state following the procedure of Sisodia & Singh (2001). They in fact showed that how independent variables may make changes in wheat production at Taluka level.

Sisodia *et al* (2015) argued that the Blocks are not homogeneous in nature. Therefore, the variation between Blocks could effected the block estimates of crop production obtained by the methods proposed by Sisodia & Singh (2001, 2012), Sisodia and Chandra (2012) and Singh et al (2012). They considered an area level model to account for Block variation by including block level covariates as independent variables and block estimates developed by Sisodia and Chandra (2012) as dependent variable is the model. They showed that the precision of the block estimates of crop productions improved considerably.

Sharma *et al* (2015) used principal component analysis of the auxiliary variables at district and Block levels. They used first few principal components as independent variables in the regression model and followed the same procedure of Singh et al (2012). They showed that the precision of Block estimates of Crop-production improved considerably.

Sharma and Sisodia (2016) made use of discriminant function analysis of auxiliary variables at district and block levels. Discriminant scores were used as independent variables in the regression model at district level and followed the same procedure of Singh et al (2012) for estimating crop production at Block level. They found increase in the precision of the Block estimates of Crop – production.

RESEARCH METHODOLOGY AND MATERIALS

3.1. Introduction

The term small area is commonly referred to describe small geographical area such as Community Development Block (generally called as Block), Panchayat or a village. Besides, small sub – populations such as age - sex - race groups or health groups defined by various disorders/disease etc. may also constitute-small area. Crop production statistics for small area like Community Development Block or Panchayat level have become essential for micro level planning in India because regional planning and the distribution of central funds are often made on a local or regional basis, and it may depend on availability of statistics for small area on various items such as crop production/ yield, the number of unemployed person, condition of the housing, use of fertilizers, principal occupation of the households, infrastructural needs, land and water endowment etc. The estimates of crop production or yield through scientifically designed crop cutting experiment (CCEs) are being reported at district level and these estimates are aggregated at state and country level. The estimates of crop production are needed at Block or Panchayat level for formulating policies and programs in agriculture in the context of de-centralized planning process at micro level in India. The size of the block is generally 1/20th of the district. The first attempt, quite earlier, was made by Panse *et al.* (1966) and Singh (1968) to develop some methodologies for estimating the crop yield at block level using double sampling approach. However, the methods developed by them could not succeed due to certain physical constraints. Srivastava *et al*. (1998) made an attempt to develop the some synthetic estimators for crop production/ yield at block level under design-based approach of survey sampling. They also performed a simulation study to assess performance of the estimators. However, their approaches lack practical application because of cost sensitivity. Sharma *et al*. (2004) proposed two different estimators for small area estimation applying design-based approach of sample survey to estimate the crop yield at the gram panchayat level (smaller area than block) and revealed that both the estimators performed satisfactorily in terms of the criterion of percentage root mean square error. However, their procedures also suffer physical constraints. The methodology proposed in the present investigation is based on scaled down approach of Sisodia and Singh (2001) and others by applying suitable regression model between series data of crop production/yield and related auxiliary variables & weather variables at district level to estimate crop production at Block level/Panchayat level.

3.2. A brief review of work done using regression model for estimating crop- production at block level:

Stasny *et al.* (1991) have made use of regression model to predict wheat production at county level in USA. They, in fact, developed the multiple regression models to exhibit best possible relationship between farm production and some predictor variables related with farm production for predicting wheat production at county level However, their approach was based on cross-sectional data. Crop production/yield are available through CCEs at district level. Conducting large number of CCEs to estimate crop yield at Block level is not viable proposition in view of lack of basic infrastructures, huge cost involved and accumulation of large non-sampling errors that are likely to creep in. Sisodia and Singh (2001) first time proposed a scale down approach using time series data on yield/production and related covariates by applying multiple regression model to obtain the block level estimate from the district level crop-production estimate. The methodology is briefly discussed below.

They assumed a multiple regression model between the crop production and auxiliary variables at district level, i.e.

$$Z_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots\dots\dots\dots + \beta_p X_{ip} + \varepsilon_i \qquad (3.2.1)$$

where Z_i is the crop production in the year i(i =1,.....,n) , X_{ij} is the value of auxiliary variable j(j =1,....., p) in the year i, β'=(β_0, β_1, …., β_p) is vector of unknown model parameters and e_i is error term assumed to follow independently normal distribution with mean 0 and variance σ^2. The fitted model with data by ordinary least square is given by

$$\hat{Z}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_{i2} + \dots\dots\dots\dots + \hat{\beta}_p X_{ip} \qquad (3.2.2)$$

where $\hat{\beta}$ is a least square estimate of β and $\hat{Z}_i$ is the estimated value of Z_i for corresponding values of X_{ij}'s in the year *i*. The sum of squares due to regression, *i.e.*, $SS_R(\beta_1, \beta_2, \ldots, \beta_p|\beta_0)$ is decomposed to define a weight that determines the relative contribution of each predictor variable included in the model(see Montegomery & Peck,1982) as follows:

$$w_j = \frac{\text{SS due to j}^{\text{th}}\text{ predictor}}{SS_R(\beta_1, \beta_2, \ldots, \beta_p|\beta_0)} \qquad (3.2.3)$$

It is assumed that the information on $X_{ij}^{'s}$ are also available at Block level. Using these weights, an estimator of crop production Z_q for the block q is constructed for a given year as :

$$\hat{Z}_q = \left[\sum_{j=1}^{P} w_j x_j\right]\hat{\hat{Z}}; \text{ q= 1, … Q,} \qquad (3.2.4)$$

where Q is total number of Block in a given district, x_j is the value of j-*th* predictor at Block level in a given year, $\hat{\hat{Z}} = \hat{Z}/A$ and $\hat{Z}$ is obtained through the fitted model (3.2.2) and *A* is the area under the crop in a given year. Note that the weight w_j depends on the set of data on Z and X_{ij} used to fit the model (3.2.1). The estimator $\hat{Z}_q$ is an unbiased estimator of Z_q if $\sum_{j=1}^{p} w_j x_j$ is considered to be a constant quantity for a given Block since under model (3.2.1) expected value of $\hat{\hat{Z}}$ is $\overline{Z}$. The variance of $\hat{Z}_q$ is given by

$$V(\hat{Z}_q) = (\delta_q/A)^2 V(\hat{Z}) \qquad (3.2.5)$$

where $\delta_q = \sum_{j=1}^{p} w_j x_j$. The variance of $\hat{Z}$ is easily available from the fitting model (3.2.2), which is equal to $\hat{\sigma}^2$, the estimated error variance. It is obvious that, in general, $\sum_{q=1}^{Q} \hat{Z}_q \neq Z$, where Z is the actual crop production reported at District level through the CCEs in a given year. Thus, a scaled estimator of Z_q was suggested by them as follows

$$\tilde{Z}_q = a_q \hat{Z}_q \qquad (3.2.6)$$

such that $\sum_{q=1}^{Q} \tilde{Z}_q = \sum_{q=1}^{Q} a_q \hat{Z}_q = Z$, where a_q is constant. There can be three alternative choices of a_q. The third choice was due to Sisodia and Singh (2012) and Sisodia and Chandra (2012). These choices are in fact different adjustments which are described below.

<u>Choice- I</u>: Ratio Adjustment

The simplest choice of a_q is to take $a_q = a$, *i.e.* the constant for each block. It can easily be shown that

$$a \sum_{q=1}^{Q} \hat{Z}_q = Z$$

i.e.,

$$a = Z / \sum_{q}^{Q} \hat{Z}_q \qquad (3.2.7)$$

Thus, a new scaled estimator of Z_q is given by,

$$\tilde{Z}_q^{(1)} = \hat{Z}_q \left[Z / \sum_{q=1}^{Q} \hat{Z}_q \right] \qquad (3.2.8)$$

Obviously, the estimator $\tilde{Z}_q^{(1)}$ is not an unbiased estimator of Z_q as $E\left[\tilde{Z}_q^{(1)}\right] = Z_q\left[Z/\sum_{q=1}^{Q}\hat{Z}_q\right]$, because $Z/\sum_{q=1}^{Q}\hat{Z}_q$ is constant for each block q, q=1,2,…,Q.

Choice- II: Least square adjustment

Another choice of a_q could be one that minimizes the sum of squared difference between $\tilde{Z}_q$ and $\hat{Z}_q$ subject to condition that $\sum_{q=1}^{Q} a_q\hat{Z}_q = Z$. To obtain a_q which minimizes the sum of squared difference between $\tilde{Z}_q$ and $\hat{Z}_q$ subject to condition $\sum_{q=1}^{Q} a_q\hat{Z}_q = Z$, we must minimize $\sum_{q=1}^{Q}(\tilde{Z}_q - \hat{Z}_q)^2$ with respect to a_q using a Lagrange multiplier, λ, to impose the desired constraints.

i.e., we minimize the function,

$$\varphi = \sum_{q=1}^{Q}(\tilde{Z}_q - \hat{Z}_q)^2 - 2\lambda\left(\sum_{q=1}^{Q}\tilde{Z}_q - Z\right)$$

or

$$\varphi = \sum_{q=1}^{Q}(a_q\hat{Z}_q - \hat{Z}_q)^2 - 2\lambda\left(\sum_{q=1}^{Q}a_q\hat{Z}_q - Z\right)$$

Differentiating ϕ with respect to a_q, and equating it equal to zero and solving for a_q, we get

$$a_q = \frac{\lambda}{\hat{Z}_q} + 1 \quad (3.2.9)$$

Since $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$, then it implies that,

$$\sum_{q=1}^{Q}\left(\frac{\lambda}{\hat{Z}_q} + 1\right)\hat{Z}_q = Z$$

and we get

$$\lambda = \left(Z - \sum_{q=1}^{Q}\hat{Z}_q\right)/Q \quad (3.2.10)$$

From (3.2.9) and (3.2.10), we get

$$a_q = \frac{Z - \sum_{q=1}^{Q}\hat{Z}_q}{Q\hat{Z}_q} + 1 \quad (3.2.11)$$

Thus, another new scaled estimator of Z_q is given by,

$$\tilde{Z}_q^{(2)} = \left[\frac{Z - \sum_{q=1}^{Q}\hat{Z}_q}{Q\hat{Z}_q} + 1\right].\hat{Z}_q$$

or

$$\tilde{Z}_q^{(2)} = \hat{Z}_q + \frac{Z - \sum_{q=1}^{Q}\hat{Z}_q}{Q} \quad (3.2.12)$$

Note that the scaled estimator $\tilde{Z}_q^{(2)}$ is obtained by adjusting the original estimator $\hat{Z}_q$ by adding a factor which is ratio of the difference between the actual crop production at district level, Z, and the sum of the original block estimates, $\sum_{q=1}^{Q}\hat{Z}_q$, to number of block in the district. If, we take expectation of $\tilde{Z}_q^{(2)}$, we have

$$E\left[Z_q^{(2)}\right] = E\left(\hat{Z}_q\right) + \frac{Z - \sum_{q=1}^{Q} E\left(\hat{Z}_q\right)}{Q}$$

$$= Z_q + \frac{Z - \sum_{q=1}^{Q} Z_q}{Q}$$

$$= Z_q + \frac{Z - Z}{Q}$$

$$= Z_q$$

Therefore, $\tilde{Z}_q^{(2)}$ is an unbiased estimator of Z_q.

Choice- III: Relative least square adjustment

The third choice of a_q could be one that minimizes the sum of square of relative differences $\left(\tilde{Z}_q - \hat{Z}_q\right)/\hat{Z}_q$ subject to condition that $\sum_{q=1}^{Q} a_q\hat{Z}_q = Z$. To obtain a_q which minimizes $\sum_{q=1}^{Q}\left[\left(\tilde{Z}_q - \hat{Z}_q\right)/\hat{Z}_q\right]^2$ subject to $\sum_{q=1}^{Q} a_q\hat{Z}_q = Z$, we minimize the following function with respect to a_q.

$$\varphi = \sum_{q=1}^{Q}\left[\left(\tilde{Z}_q - \hat{Z}_q\right)/\hat{Z}_q\right]^2 - 2\lambda\left(\sum_{q=1}^{Q} a_q\hat{Z}_q - Z\right)$$

or $$\varphi = \sum_{q=1}^{Q}\left(a_q - 1\right)^2 - 2\lambda\sum_{q=1}^{Q}\left(a_q\hat{Z}_q - Z\right) \qquad (3.2.13)$$

Differentiating the above function w.r.t. a_q and equating it equal to zero and solving for a_q, we get

$$a_q = 1 + \lambda \hat{Z}_q \tag{3.2.14}$$

Substituting a_q from (3.2.14) in the constraint $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$ and solving for λ, we get

$$\lambda = \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{\sum_{q=1}^{Q} \hat{Z}_q^2} \tag{3.2.15}$$

From (3.2.14) and (3.2.15), we have

$$a_q = 1 + \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{\sum_{q=1}^{Q} \hat{Z}_q^2} . \hat{Z}_q \tag{3.2.16}$$

Therefore, an another new scaled estimator of Z_q is given by

$$\tilde{Z}_q^{(3)} = \left[1 + \frac{Z - \sum_{q=1}^{Q} \hat{Z}_q}{\sum_{q=1}^{Q} \hat{Z}_q^2} . \hat{Z}_q \right] . \hat{Z}_q$$

$$\tilde{Z}_q^{(3)} = \hat{Z}_q + \frac{\hat{Z}_q^2}{\sum_{q=1}^{Q} \hat{Z}_q^2} \left(Z - \sum_{q=1}^{Q} \hat{Z}_q \right) \tag{3.2.17}$$

It is obvious that scaled estimator $\tilde{Z}_q^{(3)}$ is obtained by adjusting the original estimator $\hat{Z}_q$ by adding a factor which is a proportion of the difference between district total Y and the sum of the original estimates of blocks. The Proportion here is based on the squared values of the original estimates $\hat{Z}_q$.

It may also be noted that $\tilde{Z}_q^{(3)}$ is not an unbiased estimator of Z_q as

$$E\left[\tilde{Z}_q^{(3)}\right] = Z_q + E\left[\frac{\hat{Z}_q^2}{\sum_{q=1}^{Q}\hat{Z}_q^2}\left(Z - \sum_{q=1}^{Q}\hat{Z}_q\right)\right] \qquad (3.2.18)$$

Evidently, the second term in right hand side of (3.2.18) may not be zero. Thus, $\tilde{Z}_q^{(3)}$ is in general a biased estimator of Z_q. Up to first degree of approximation (Sukhatme and Sukhatme, 1970), the second term in (3.2.18) is, however, almost zero because the ratio $\frac{\hat{Z}_q^2}{\sum_{q=1}^{Q}\hat{Z}_q^2}$ is also expected to be very close to zero. Thus $\tilde{Z}_q^{(3)}$ is also an almost unbiased estimator.

The conditional bias and conditional MSE/ Variance

The bias in $\tilde{Z}_q^{(1)}$ is obtained as

$$\text{Bias}\left[\tilde{Z}_q^{(1)}\right] = E\left[\tilde{Z}_q^{(1)}\right] - Z_q$$

$$= \frac{\left(Z - \sum_{q=1}^{Q}\hat{Z}_q\right)Z_q}{\sum_{q=1}^{Q}\hat{Z}_q} \qquad (3.2.19)$$

MSE of $\tilde{Z}_q^{(1)}$ is given by

$$MSE\left[\tilde{Z}_q^{(1)}\right] = \left[\text{Bias}\left(\tilde{Z}_q^{(1)}\right)\right]^2 + V\left(\tilde{Z}_q^{(1)}\right)$$

$$MSE\left[\tilde{Z}_q^{(1)}\right] = \frac{\left(Z - \sum_{q=1}^{Q}\hat{Z}_q\right)^2 Z_q^2}{\left(\sum_{q=1}^{Q}\hat{Z}_q\right)^2} + \frac{Z^2}{\left(\sum_{q=1}^{Q}\hat{Z}_q\right)^2} V\left(\hat{Z}_q\right) \tag{3.2.20}$$

Estimate of MSE $(\tilde{Z}_q^{(1)})$ is given by

$$M\hat{SE}(\tilde{Z}_q^{(1)} = \frac{\left[Z - \sum_{q=1}^{Q}\hat{Z}_q\right]^2 \hat{Z}_q^2}{\left[\sum_{q=1}^{Q}\hat{Z}_q\right]^2} + \frac{Z^2}{\left[\sum_{q=1}^{Q}\hat{Z}_q\right]^2} V(\hat{Z}_q) \tag{3.2.21}$$

Variance of $\tilde{Z}_q^{(2)}$ is given by

$$V\left(\tilde{Z}_q^{(2)}\right) = \frac{Q-2}{Q} V\left(\hat{Z}_q\right) + \frac{1}{Q^2}\sum_{q=1}^{Q} V\left(\hat{Z}_q\right) \tag{3.2.22}$$

The estimate of variance of $\tilde{Z}_q^{(2)}$ can easily be obtained by computing $V\left(\hat{Z}_q\right)$ from the fitted model.

Bias & Mean Square Error (MSE) of $\tilde{Z}_q^{(3)}$

The bias of $\tilde{Z}_q^{(3)}$ can directly be written from the expression (3.2.18) as

$$Bias\left[\tilde{Z}_q^{(3)}\right] = E\left[\tilde{Z}_q^{(3)}\right] - Z_q$$

$$= E\left[\frac{\hat{Z}_q^2}{\sum_{q=1}^{Q}\hat{Z}_q^2}\left(Z - \sum_{q=1}^{Q}\hat{Z}_q\right)\right] \tag{3.2.23}$$

Evidently, the bias will be close to zero if $\sum_{q=1}^{Q}\hat{Z}_q$ is close to Z. If $Z \cong \sum_{q=1}^{Q}\hat{Z}_q$, then approximate MSE of $\tilde{Z}_q^{(3)}$ is given by

$$MSE\left[\tilde{Z}_q^{(3)}\right] = \left[1 + \frac{Z - \sum_{q=1}^{Q}\hat{Z}_q}{\sum_{q=1}^{Q}\hat{Z}^2{}_q}\hat{Z}_q\right]^2 V\left(\hat{Z}_q\right) \quad (3.2.24)$$

Comparison for relative efficiency of $\tilde{Z}_q^{(1)}, \tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ over, $\hat{Z}_q$

Taking difference of $V\left(\hat{Z}_q\right)$ and $MSE\left(\tilde{Z}_q^{(1)}\right)$, i.e. $V\left(\hat{Z}_q\right)$ - $MSE\left(\tilde{Z}_q^{(1)}\right)$ and after simplification, we get the inequality

$$V\left(\hat{z}_q\right) \geq \frac{\left(z - \sum_{q=1}^{Q}\hat{z}_q\right)^2 z_q^2}{\left[\left(\sum_{q=1}^{Q}\hat{z}_q\right)^2 - z^2\right]} \quad (3.2.25)$$

for $\tilde{Z}_q^{(1)}$ to be efficient than $\hat{Z}_q$. However, it seems that inequality (3.2.25) may not hold true in general.

Taking difference of $V\left(\hat{z}_q\right)$ and $V\left(\tilde{Z}_q^{(2)}\right)$, i.e. $V\left(\hat{z}_q\right)$-$V\left(\tilde{Z}_q^{(2)}\right)$, and after simplification, we get the following inequality

$$V\left(\hat{z}_q\right) \geq \frac{1}{2Q}\sum_{q=1}^{Q} V\left(\hat{z}_q\right) \quad (3.2.26)$$

for $\tilde{Z}_q^{(2)}$ to be efficient than $\hat{z}_q$. It is obvious from the inequality (3.2.25) that the $V\left(\hat{z}_q\right)$ must be greater than the half of the average of variances of $\hat{z}_q$ (q=1, 2, … Q), which is quite possible in general.

Taking difference of $V(\hat{Z}_q)$ and $MSE(\tilde{Z}_q^{(3)})$, i.e. $V(\hat{Z}_q)$ - $MSE(\tilde{Z}_q^{(3)})$, we get the following inequality

$$1-\left\{1+\frac{z-\sum_{q=1}^{Q}\hat{z}_q}{\sum_{q=1}^{Q}\hat{z}_q^2}\hat{z}_q\right\}^2 \geq 0 \qquad (3.2.27)$$

Case1: If $Z-\sum_{q=1}^{Q}\hat{Z}_q > 0$, then RHS of (3.2.27) is negative indicating

thereby $\hat{Z}_q$ will be always efficient than $\tilde{Z}_q^{(3)}$

Case2: If $Z-\sum_{q=1}^{Q}\hat{Z}_q < 0$, then RHS of (3.2.27) is always positive indicating

thereby $\tilde{Z}_q^{(3)}$ will always be efficient than $\hat{Z}_q$.

The empirical study conducted for rice production in Faizabad district (U.P.) by Sisodia and Singh (2001) and for wheat production in Barabanki district (U.P.) by Sisodia and Singh (2012) and Sisodia and Chandra (2012) showed that the block estimate of crop production based on the estimator $\tilde{Z}_q^{(2)}$ was most precise as compared to other estimators. It can observed that the weight w_j used in $\hat{Z}_q$ is subjective one as it depends upon the set of Z and X_j's chosen for the fitting of the model (3.2.1). Sisodia and Chandra (2012) has, however, advocated for stable value of w_j obtained through iterative procedure of fitting of the model (3.2.1).

Singh *et al*. (2012) proposed a predictive approach of estimation of crop-production at block level using the model (3.2.1) fitted at district level. Their procedure is described briefly here.

A linear multiple regression models between the crop yield and independent auxiliary variables at district level can be specified as given below

$$Y_i=\beta_0+\beta_1X_{i1}+\beta_2X_{i2}+\dots+\beta_pX_{ip}+\varepsilon_i \qquad (3.2.28)$$

where Y_i is the crop yield in the i^{th} year, (i= 1, 2, …n), X_{ij} is the value of j^{th} predictor (j=1, 2, … , P) in the i^{th} year, $\beta' = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$ is vector of unknown parameters of the model and ε_i is error term. It is also assumed that ε_i's follow independently normal distribution with mean 0 and variance σ^2. For example, X_{ij}'s may be percent irrigated area under a crop, relative area under the crop as percentage of the gross-cropped area (GCA), fertilizer consumption in kg per hectare for the crop, etc. The model (3.2.28) is fitted with the data at district level and let the fitted model be denoted as

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip} \qquad (3.2.29)$$

where $\hat{\beta}$ is the least square estimate of β and $\hat{Y}_i$ is the estimated value of Y_i for corresponding values of X_{ij}'s.

The fitted model (3.2.29) is then used directly at block level to predict the Z_q, the yield at block q (q=1, 2, …Q). It is well known from the theory of linear regression models that prediction is always considered to be more precise when the values of X_j's corresponding to which the prediction is to be made is within the ranges of X_j's used of fit the model (3.2.27). It is generally expected that percent irrigated area under the crop, fertilizer consumption (kg/ha), and relative area under the crop as percent to GCA at district and block level are almost within a similar range.

Let $\hat{Y}_q$ be the predicted value of the crop yield Y_q (Y_q= yield per unit of area) for the block q during a particular year of interest, obtained from the fitted model (3.2.29) as

$$\hat{Y}_q = \hat{\beta}_0 + \hat{\beta}_1 x_{q1} + ... + \hat{\beta}_p x_{qp} \quad (3.2.30)$$

where X_{qj}'s; j= 1, 2, …. p, are values of predictors in the q^{th} block. Therefore, an unbiased estimator of Z_q, the crop production at q^{th} block is given by

$$\hat{Z}_q = \delta_q \hat{Y}_q \quad (3.2.31)$$

where δ_q is the area under the crop at q^{th} block. It may be noted that the area under a given crop is being reported through complete enumeration by Revenue Department of State Government in India. Note that $\hat{Z}_q$ is an unbiased estimator of Z_q as $E(\hat{Z}_q) = Z_q$ because $E(\hat{Y}_q) = Y_q$ under the assumption of regression model (3.2.28).

Following Montgomery and Peck (1982) the variance of $\hat{Z}_q$ is given by

$$V(\hat{Z}_q) = \delta_q^2 \sigma^2 C_q' (X'X)^{-1} C_q \quad (3.2.32)$$

where σ^2 is the residual variance corresponding to the regression model (3.2.28), X is matrix of X_{ij}'s at district level given by

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & . & . & . & X_{1p} \\ 1 & X_{21} & X_{22} & . & . & . & X_{2p} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & X_{n1} & X_{n2} & . & . & . & X_{np} \end{bmatrix}$$

and X' is the transpose of X matrix. C_q is column vector of X_{qi}'s at q^th^ block level given by

$$C'_q = [1 \;\; X_{q1} \;\; X_{q2} \;\; \ldots \;\; X_{qp}]$$

It is obvious that in general $\sum_{q=1}^{Q} \hat{Z}_q \neq Z$, where Z is the actual crop production reported at district level through crop cutting experiment in a given year. Thus, a new scaled estimator Z_q is given by

$$\tilde{Z}_q = a_q \hat{Z}_q \qquad (3.2.33)$$

where a_q is constant such at $\sum_{q=1}^{Q} \tilde{Z}_q = Z$ or $\sum_{q=1}^{Q} a_q \hat{Z}_q = Z$

The choices of a_q and resultant scaled estimators along with their conditional variance/MSE will be same as described earlier. The only difference would be in the values of $\hat{Z}_q$ and variance of $\hat{Z}_q$ as these are based on prediction approach.

Ahamad *et al* (2014) used cross section data of farmers and weather variables to obtained wheat production data at Taluka level in karnatka state following the procedure of Sisodia & Singh (2001). They in fact showed that how independent variables may make changes in wheat production at Taluka level.

Sisodia *et al* (2015) argued that the Blocks are not homogeneous in nature. Therefore, the variation between Blocks could affect the Block estimates of crop production obtained by the methods proposed by Sisodia &Singh (2001, 2012), Sisodia and Chandra (2012) and Singh *et al* (2012). They considered an area level model to account for Block variation by including block level covariates as independent variables and block estimates developed by Sisodia and Chandra (2012) as dependent variable in the area level model. They showed that the precision of the block estimates of crop productions improved considerably.

Sharma and Sisodia (2015) used principal component analysis of the auxiliary variables at district and Block levels. They used first few principal components as independent variables in the regression model and followed the same procedure of Singh *et al* (2012). They showed that the precision of Block estimates of Crop-production improved considerably.

Sharma *et al* (2016) made use of discriminant function analysis of auxiliary variables at district and block levels. Discriminant scores were used as independent variables in the regression model at district level and followed the same procedure of Singh *et al* (2012) for estimating crop production at Block level. They found increase in the precision of the Block estimates of Crop – production.

3.3. Proposed methodology for estimation of crop production at Block level

We have discussed about approach of Singh *et al* (2012) followed by Sharma *et al*(2015) and Shrma & Sisodia(2016) for estimation of crop production at Block level in the preceding section. These research workers have used only input factors as independent variables in the regression model. The environment plays an important role in crop production. Therefore, it is proposed to include weather variables also besides input factors as independent variables in the regression model. The proposed model at district level is given below.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots \beta_p X_{ip} + \lambda_1 I_{i1} + \lambda_2 I_{i2} + \ldots\ldots + \lambda_m I_{im} + \varepsilon_i \qquad (3.3.1)$$

where Y_i is the yield of the crop during i^{th} year (i=1,2,3,..n), X_{ij} is the value of j^{th} input factor (j=1,2,….p) corresponding to i^{th} year, I_{im} is the Index value of m^{th} weather variable (m=1,2,…M) corresponding to i^{th} year. $\left(\beta_0, \beta_1, \beta_2 \ldots\ldots\ldots\ldots \beta_P, \lambda_1, \lambda_2 \ldots\ldots, \lambda_M\right)$ are model parameters and ϵ_i is error term assumed to follow independently normal distribution with mean zero and variance σ^2.The weather index is constructed following the procedure laid down in Agrawal *et al* (1986). It is described below.

$$I_m = \frac{\sum_{w=1}^{k} r_{mw} X_{mw}}{\sum_{w=1}^{k} r_{mw}} \qquad (3.3.2)$$

Where X_{mw} is the value of m^{th} weather variable in w^{th} week and r_{mw} is the correlation coeffffficient between de-trended crop yield and m^{th} weather variable in w^{th} week . I_m is in fact weighted average of m^{th} weather variable, the height is being taken as correlation coefficient between de-trend crop yield and weather variable in w^{th} week.

The model (3.3.1) is fitted with data by ordinary least square technique

Let the fitted model be denoted as

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2} + ... \hat{\beta}_p X_{ip} + \hat{\lambda}_1 I_{i1} + \hat{\lambda}_2 I_{i2} + ... + \hat{\lambda}_m I_{im} \quad \text{`} \quad (3.3.3)$$

The fitted model (3.3) is used directly to predict the Block yield of the crop. The predicted crop yield for q th Block (q=1,2,…..Q) in a given year is given by

$$\hat{Y}_q = \hat{\beta}_0 + \hat{\beta}_1 X_{q1} + \hat{\beta}_2 X_{q2} + ... + \hat{\beta}_p X_{qp} + \hat{\lambda}_1 I_{q1} + \hat{\lambda}_2 I_{q2} + ... + \hat{\lambda}_M I_{qM} \quad (3.3.4)$$

where $\hat{Y}_q$ is the predicted crop yield for q^{th} Block in the given year, X_{qp} and I_{qm} is the value of I_m in q^{th} Block in a given year, and $\left(\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p, \hat{\lambda}_I, ..., \hat{\lambda}_M\right)$ are estimated model parameters.

An estimator of crop production, Z_q, for q^{th} Block in a given year, is therefore, given by

$$\hat{Z}_q = \delta_q . \hat{Y}_q \quad (3.3.5)$$

where δ_q is the area under the crop in q^{th} Block.

In general, $\sum_{q=1}^{Q} \hat{Z}_q \neq Z$, where Z is crop production reported by CCEs at district level in that given year.

Following Singh *et al* (2012) and others, We develop a sealed estimator of Z_q as

$$\tilde{Z}_q = a_q \hat{Z}_q \tag{3.3.6}$$

such that $\sum_{q=1}^{Q} \tilde{Z}_q = \sum_{q=1}^{Q} a_q \hat{Z}_q = Z$, where a_q is the sealed factor.

The choices for a_q are similar to those described in section-2. The resultant estimators, i.e. $\tilde{Z}_q^{(1)}, \tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ will also be similar to those given in section-2. The expressions for conditional variances and MSE of the scaled estimators will also be similar to those given in section-2. The only difference we get in the magnitude of $\hat{Z}_q, \tilde{Z}_q^{(1)}, \tilde{Z}_q^{(2)}, \tilde{Z}_q^{(3)}$ and in the variance of $\hat{Z}_q$. The variance of $\hat{Z}_q$ can be obtained by following the Montegomery & Peck as follows.

$$V(\hat{Z}_q) = \delta_q^2 \hat{\sigma}^2 C_q' (X'X)^{-1} C_q$$

where X is the matrix of independent variables in the model, given by

$$X = \begin{bmatrix} 1 & X_{11} & X_{12}\ldots\ldots & X_{1p} & I_{11} & I_{12}\ldots\ldots & I_{1M} \\ 1 & X_{21} & X_{22}\ldots\ldots & X_{2p} & I_{21} & I_{22}\ldots\ldots & I_{2M} \\ . & . & \ldots\ldots\ldots\ldots & . & . & \ldots\ldots\ldots\ldots & . \\ . & . & \ldots\ldots\ldots\ldots & . & . & \ldots\ldots\ldots\ldots & . \\ . & . & \ldots\ldots\ldots\ldots & . & . & \ldots\ldots\ldots\ldots & . \\ 1 & X_{n1} & X_{n2}\ldots\ldots & X_{np} & I_{n1} & I_{n2}\ldots\ldots & I_{nM} \end{bmatrix}$$

X' is the transpose of X and C_q' is column vector of X_{qp}' and I_{qM} corresponding to q^{th} Block in a given year as follows.

$$C_q' = \lfloor 1 X_{q1} \quad X_{q2}\ldots\ldots \quad X_{qp} \quad I_{q1} \quad I_{q2}\ldots\ldots \quad I_{qM} \rfloor$$

$\hat{\sigma}^2$ is the estimated residual variance which can be obtained from the Table of analysis of variance of regression model.

3.4. Data used for empirical study

The proposed methodology is illustrated by considering the time series data on wheat yield, percent irrigated area under wheat (X_1), fertilizer consumption for wheat (N,P,K) in kg/ha (X_2) and percent relative area under wheat as percentage of gross cropped area (X_3) for Sultanpur district of Uttar Pradesh in India pertaining to the period 1990-91 to 2008-09. These time series data were obtained from the Bulletins of Agricultural statistics, published by Directorate of Agricultural Statistics and Crop Insurance Government of Uttar Pradesh, India.

It may be noted that the fertilizer consumption (N, P, K) is not being reported crop wise, hence the total fertilizer consumption in a year was apportioned for the wheat crop. Approximately 40 per cent of the total fertilizer consumption in a year is considered to have been used for wheat crop.

The Block wise time series data on three auxiliary variables, X_1, X_2 and X_3 for Sultanpur district pertaining to the same period have also been obtained from the District Statistical Bulletins, published by Directorate of Economics and Statistics, Govt. of Uttar Pradesh, India. The weekly time series data of weather variables viz. Minimum temperature, Maximum temperature, Wind velocity from Standard Metrological Week 44 to the next year of SMW 15 for the aforesaid period were obtained from the Metrological station, Amausi, Lucknow, which were utilized in the present investigation. The Block wise actual data of wheat production based on crop cutting experiments in Sultanpur district during the year 2006-07 are also available in District Statistical Bulletins as part of some pilot studies conducted by Govt. of UP.

These block wise wheat production data have been used for validation of Block level estimates of wheat production obtained from the various estimators developed in the preceding sections.

RESULTS AND DISCUSSION

4.1. Introduction

An empirical study has been conducted to illustrate the relative efficiencies of the proposed estimators based on different approaches of statistical methodologies as given in Chapter III. The time series data on yield of wheat, percent irrigated area under wheat (X_1), fertilizer consumption for wheat (N, P, K) in kg/ha (X_2) and percent relative area under wheat as percentage of gross cropped area (X_3) for Sultanpur district of Utter Pradesh in India pertaining to the period 1990-91 to 2008-09 were used for the study. These time series data were obtained from the Bulletins of Agricultural statistics, published by Directorate of Agricultural Statistics and Crop Insurance, Government of Uttar Pradesh, India. It may be noted that the fertilizer consumption (N, P, K) is not being reported crop wise, hence the total fertilizer consumption in a year was apportioned for the wheat crop. Approximately 40 per cent of the total fertilizer consumption in a year is considered to have been used for wheat crop. The Block wise time series data on three auxiliary variables, X_1, X_2 and X_3 for Sultanpur district pertaining to the same period have also been obtained from the District Statistical Bulletins, published by Directorate of Economics and Statistics, Govt. of Uttar Pradesh, India. The weekly time series data of weather variables viz. Minimum temperature, Maximum temperature, Wind velocity from Standard

.

Metrological Week 44 to the next year of SMW 15 for the aforesaid period were obtained from the Metrological station, Amausi, Lucknow, which were utilized in the present investigation. The Block wise actual data of wheat production based on crop cutting experiments in Sultanpur district during the year 2006-07 are also available in District Statistical Bulletins as part of some pilot studies conducted by Govt. of UP, which were used for validation.

The block estimates of wheat production for the year 2006-07 based on various estimators developed using four statistical methodology have been obtained along with their standard error *etc.* The results are presented in tow sections as follows.

4.1. Results based on the methodology suggested by Singh *et al.* (2012)

To illustrate the statistical methodology developed in preceding chapter proposed by Singh *et al.* (2012), an empirical study has been carried out. It is proposed to estimate wheat production at block level in Sultanpur district of Uttar Pradesh, India. The time series data on Y_i (wheat yield) and X_{ij} (covariates) pertaining to the period 1990-91 to 2008-09 were used for the fitting the model at district level. These data are given in Appendix-I (Table I.2). The block wise data on Y_q (wheat yield based on CCEs), Z_q (wheat production based on CCEs) and X_j's for the year 2006-07 are given in Appendix-II. The block wise data on X_j's have been used for developing block level estimate of Z_q from the proposed estimators in the year 2006-07.

The following multiple linear regression model at district level has been fitted using least square technique.

$$Y_i=\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+ \beta_3 X_{i3}+\varepsilon_i;\ \ i= 1, 2, \ldots, n \qquad (4.1.1)$$

Here ε_i is assumed to follow independently normal distribution $(0, \sigma^2)$. The results of the fitted model are presented in Table 4.1.

Table 4.1: Estimate of regression coefficients, their standard error and R^2

Variable	Regression Coefficient	Standard error	R^2 (%)
X_1	5.740574*	2.6714014	84.67
X_2	0.076493*	0.0302428	
X_3	-0.26132	0.3877038	
Constant	-544.536	253.16810	

*P <0.05, Standard error of estimates $\hat{Y}$ = 1.33422

The estimates of regression coefficient, their standard error and value of coefficient of determinations (R^2) *etc.* have been presented in Table 4.1. The per cent irrigated area and fertilizer consumption showed positive and significant effect. The value of coefficient of determination (R^2) was found to be quite high i.e. 84.67% which is indicative of the fact that these variables included in the model have been quite sufficient to explain the variability in the data of wheat yield at district level. The analysis of variance for regression analysis of the aforesaid model is presented in Table 4.2. This shows the overall significance of the model fitted.

Table 4.2: Analysis of variance for regression analysis

Source	d.f.	Sum of Square	Mean square	F ratio	Prob.
Regression	3	147.5926	49.19755	27.6368**	2.34×10^{-06}
Residual	15	26.70214	1.780143		
Total	18	174.2948			

** Significant at P<.01

The model fitted at district level was directly used to predict the block estimates of wheat yield using block wise data on X_1, X_2 and X_3 for the year 2006-07 (Appendix-II).

The blocks estimates of wheat production based on four estimators and their per cent standard error were computed and are presented in Table 4.3. To measure how far the adjusted estimates $\tilde{Z}_q^{(i)}$ are from the Z_q, an average distance between adjusted estimates and Z_q has been computed by following formula.

$$D_i = \sqrt{\frac{\sum_{q=1}^{Q}\left(Z_q - \hat{\theta}_{qi}\right)^2}{Q}} \qquad \text{(i= 1, 2, 3, 4)} \qquad (4.1.2)$$

Where $\hat{\theta}_{qi} = \hat{Z}_q$, $\hat{\theta}_{qi} = \tilde{Z}_q^{(1)}$, $\hat{\theta}_{qi} = \tilde{Z}_q^{(2)}$, $\hat{\theta}_{qi} = \tilde{Z}_q^{(3)}$ and Q is number of blocks. The values of D_i's are also presented in Table 4.3.

The results presented in Table 4.3 showed that the estimates $\hat{Z}_q$, $\tilde{Z}_q^{(1)}$, $\tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ are almost at par in terms of percent standard error of block estimate. However, the percent standard error of block estimates ranged in general to the maximum 9% with exception of Waldirai and Shahgarh which yielded about 20% and 24% standard error.

The block estimates obtained from various estimators have been found to be close to the actual yield. However, in few blocks the estimates are not close to actual production, probably because of local effects of the blocks, which may not uniformly similar to other block. The average distance (D_i) has been found to be minimum, i.e. 20743 in case of $\tilde{Z}_q^{(2)}$.

The percent errors of the estimates based on $\tilde{Z}_q^{(2)}$ were almost at par with that of $\hat{Z}_q$ and $\tilde{Z}_q^{(3)}$. Therefore, on the basis of percent standard error of the block estimates and the average distance it can be recommended that $\tilde{Z}_q^{(2)}$ could be preferred in practice for estimating the block estimate among all the other estimators.

Table 4.3: Block estimates of wheat production based on different estimators and their per cent standard error during the year 2006-07 using multiple regression model based on only input factors

S.No.	Block	Actual Prod. (qt) Z	Block Estimates				% Standard Error			
			$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$	$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$
1	Shukul bazaar	188293	186441	183623	183268	184092	1.83	2.39	2.45	1.83
2	Jagdishpur	198054	200259	197233	197086	197549	4.25	4.52	4.41	4.25
3	Musafhirkhana	166349	161401	158961	158227	159640	7.32	7.48	7.39	7.32
4	Waldi Rai	142614	151343	149055	148169	149795	20.39	20.44	19.96	20.39
5	Jamo	230371	228137	224689	224963	224620	3.76	4.06	3.89	3.76
6	Shahgarh	136072	179539	176825	176365	177361	24.05	24.10	23.41	24.05
7	Gauriganj	242798	232773	229255	229599	229112	5.95	6.15	5.91	5.95
8	Amethi	168959	197332	194350	194159	194701	7.15	7.31	7.11	7.15
9	Bhetua	146587	171435	168844	168261	169449	2.79	3.18	3.28	2.79
10	Bhadar	138111	169024	166469	165850	167093	8.70	8.83	8.65	8.70
11	Sangrampur	112059	131548	129560	128374	130379	8.75	8.89	8.89	8.75
12	Dhanpatganj	233104	214005	210770	210831	210910	1.85	2.41	2.32	1.85
13	Kurebhar	217956	224842	221443	221668	221426	1.69	2.28	2.15	1.69
14	Jai Singh Pur	232049	235704	232141	232530	231950	2.94	3.32	3.14	2.94
15	Kurwar	236005	203760	200680	200586	200954	1.93	2.47	2.43	1.93
16	Dube Pur	214739	211437	208241	208263	208416	3.45	3.78	3.66	3.45
17	Bhadaiyeea	176209	214941	211692	211767	211819	1.82	2.38	2.29	1.82
18	Dostpur	229495	215650	212390	212476	212507	2.91	3.29	3.17	2.91
19	Akhand Nagar	314681	274881	270727	271708	269775	4.09	4.37	4.11	4.09
20	Lambhua	234987	236447	232873	233273	232669	3.38	3.71	3.53	3.38
21	Pratap Pur Kamaicha	177156	172890	170277	169716	170870	7.88	8.03	7.87	7.88
22	Kadipur	412974	405653	399522	402480	394534	3.49	3.82	3.45	3.49
	Total	4549621	4619441	4549621	4549621	4549621				
						Average distance	20985	21066	20743	21460

4.2. Results based on the proposed methodology using weather variables including input factors.

To illustrate the proposed statistical methodology developed in preceding chapter an empirical study has been carried out. It is proposed to estimate wheat production at block level in Sultanpur district of Uttar Pradesh, India. The time series data on Y_i (wheat yield), I_{im} (Index value of m^{th} weather variable) and X_{ij} (covariates) pertaining to the period 1990-91 to 2008-09 were used for the fitting the model at district level. These data are given in Appendix-I (Table I.2). The block wise data on Y_q (wheat yield based on CCEs), Z_q (wheat production based on CCEs) and X_j's for the year 2006-07 are given in Appendix-II. The block wise data on X_j's and I_{im} have been used for developing block level estimate of Y_q from the proposed estimators in the year 2006-07. The weighted weather Index I_m developed at district level was also used at Block level.

The following multiple linear regression model at district level has been fitted using least square technique.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} ..... + \beta_p X_{ip} + \lambda_1 I_{i1} + \lambda_2 I_{i2} + \lambda_3 I_{i3} + \varepsilon_i \qquad (4.2.1)$$

where i=1,2,…,n.

Here ε_i is assumed to follow independently normal distribution (0, σ^2). The results of the fitted model are presented in Table 4.4. I_{im} is the weighted weather index for m^{th} weather variable as described in the preceding chapter. Here , I_{i1}, I_{i2} and I_{i3} are weighted weather Index of maximum temperature ,minimum temperature and wind velocity.

Table 4.4: Estimate of regression coefficients, their standard error and R^2

Variable	Regression Coefficient	Standard error	R^2 (%)
X_1	1.050754	2.465075	92.81
X_2	0.066392*	0.026496	
X_3	0.017023	0.319729	
I_{11}	-0.1972*	0.083569	
I_{21}	0.405358	0.403399	
I_{31}	0.518721	0.291546	
Constant	-91.7412	235.1258	

*P <0.05, Standard error of estimates $\hat{Y}$ = 1.02168

The estimates of regression coefficient, their standard error and value of coefficient of determinations (R^2) *etc.* have been presented in Table 4.4. The fertilizer consumption showed positive and significant effect.I_{11} (maximum temperature) has shown negative and significant effect on the yield. The value of coefficient of determination (R^2) was found to be quite high i.e. 92.81% which is indicative of the fact that these variables included in the model have been quite sufficient to explain the variability in the data of wheat yield at district level. The analysis of variance for regression analysis of the aforesaid model is presented in Table 4.5. This shows the overall significance of the model fitted.

Table 4.5: Analysis of variance for regression analysis

Source	d.f.	Sum of Square	Mean square	F ratio	Prob.
Regression	6	161.7688	26.96147	25.82931**	3.4×10^{-06}
Residual	12	12.52599	1.043832		
Total	18	174.2948			

** Significant at P<.01

The model fitted at district level was directly used to predict the block estimates of wheat yield using block wise data on X_1, X_2, X_3, I_{11}, I_{21} and I_{31} for the year 2006-07 (Appendix-II). The blocks estimates of wheat production based on four estimators and their per cent standard error were computed and are presented in Table 4.6. Di's have been computed by formula given in 4.1. The values of Di's are also presented in Table 4.6

The results presented in Table 4.6 showed that the estimates $\hat{z}_q$, $\tilde{z}_q^{(1)}$, $\tilde{z}_q^{(2)}$ and $\tilde{z}_q^{(3)}$ are almost at par in terms of percent standard error of block estimate. However, the percent standard error of block estimates ranged in general to the maximum 7% with exception of Waldirai and Shahgarh which yielded about 20% and 17% standard error.

The block estimates obtained from various estimators have been found to be close to the actual yield. However, in few blocks the estimates are not close to actual production, probably because of local effects of the blocks, which may not uniformly similar to other block. The average distance (D_i) has been found to be minimum, i.e. 24456 in case of $\tilde{Z}_q^{(2)}$. The percent errors of the estimates based on $\hat{Z}_q$, $\tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ were almost at par. Therefore, on the basis of percent standard error of the block estimates and the average distance it can be recommended that $\tilde{Z}_q^{(2)}$ could be preferred in practice for estimating the block estimate among all the other estimators.

Table 4.6: Block estimates of wheat production based on different estimators and their per cent standard error during the year 2006-07 using multiple regression model on the basis of input factors and weather variables.

S.No.	Block	Actual Prod. (qt) Z	Block Estimates				% Standard Error			
			$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$	$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$
1	Shukul bazaar	188293	184163	183299	183189	183446	2.09	2.14	2.47	2.09
2	Jagdishpur	198054	195222	194306	194247	194416	3.86	3.89	3.94	3.86
3	Musafhirkhana	166349	149678	148976	148704	149205	6.82	6.84	6.79	6.82
4	Waldi Rai	142614	128156	127555	127181	127809	20.48	20.48	19.79	20.48
5	Jamo	230371	225439	224381	224464	224364	4.10	4.13	4.10	4.10
6	Shahgarh	136072	203612	202657	202637	202735	17.29	17.30	16.62	17.29
7	Gauriganj	242798	237627	236512	236652	236433	4.69	4.71	4.62	4.69
8	Amethi	168959	199683	198747	198709	198840	5.80	5.81	5.71	5.80
9	Bhetua	146587	168678	167887	167703	168076	2.63	2.67	2.98	2.63
10	Bhadar	138111	170599	169799	169625	169984	7.17	7.19	7.05	7.17
11	Sangrampur	112059	130264	129653	129289	129905	7.51	7.53	7.50	7.51
12	Dhanpatganj	233104	211326	210335	210352	210382	2.34	2.39	2.57	2.34
13	Kurebhar	217956	219559	218529	218585	218540	2.24	2.29	2.46	2.24
14	Jai Singh Pur	232049	233189	232095	232214	232039	2.69	2.73	2.81	2.69
15	Kurwar	236005	201002	200060	200028	200148	2.61	2.65	2.83	2.61
16	Dube Pur	214739	209941	208956	208967	209009	3.00	3.04	3.14	3.00
17	Bhadaiyeea	176209	209895	208911	208921	208964	2.39	2.43	2.61	2.39
18	Dostpur	229495	207420	206447	206445	206510	3.28	3.32	3.40	3.28
19	Akhand Nagar	314681	274146	272860	273171	272556	4.02	4.05	3.97	4.02
20	Lambhua	234987	238921	237800	237946	237714	2.93	2.96	3.01	2.93
21	Pratap Pur Kamaicha	177156	174430	173612	173456	173787	6.44	6.46	6.36	6.44
22	Kadipur	412974	398110	396243	397136	394759	3.15	3.18	3.08	3.15
	Total	4549621	4571062	4549621	4549621	4549621				
						Average distance	24475	24519	24456	24607

To have visual comparison between actual estimates Z_q (based on CCEs) and estimates based on the estimators, the graphs have been sketched and these are presented in Fig 4.8 to 4.15. The graphs are sketched for the methodology given by Singh et al (2012) and for the methodology proposed in present investigation. It can observed from the graphs that Block estimate of Wheat production based on different estimators were quite close to actual wheat production for most of the Blocks except for the few Blocks.

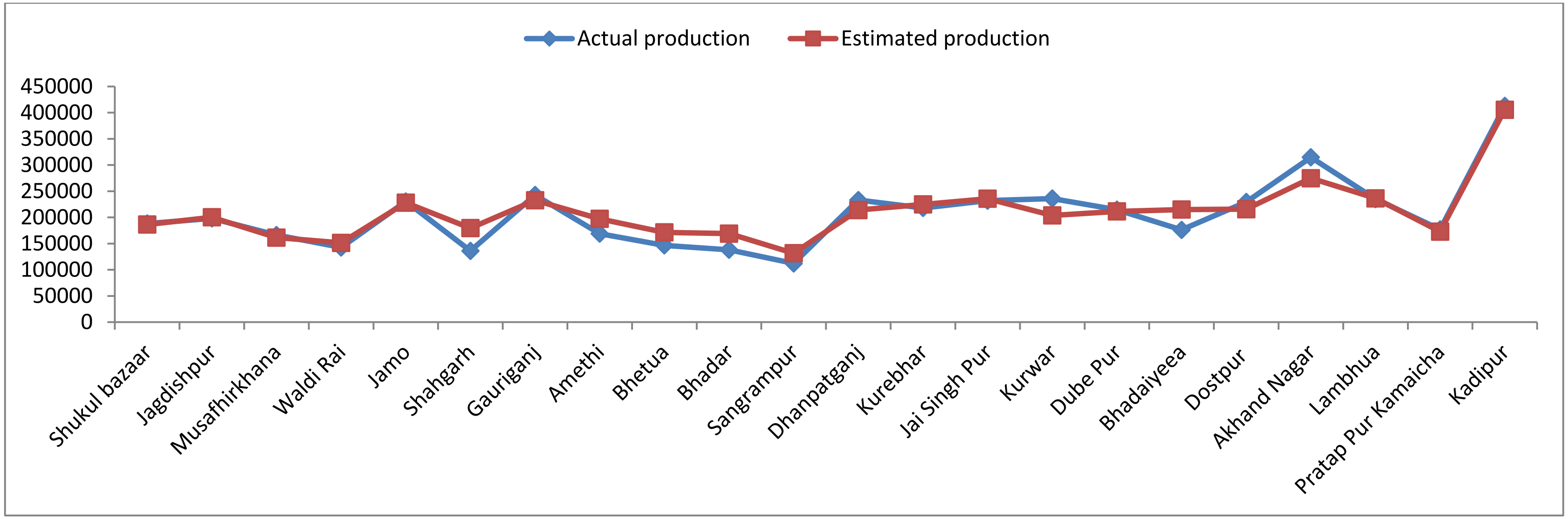

Fig. 4.8: Block estimate of wheat production in Sultanpur district (U.P.) (Z_q & $\hat{Z}_q$) for the year 2006-07 based on methodology due to Singh *et al.* (2012) .

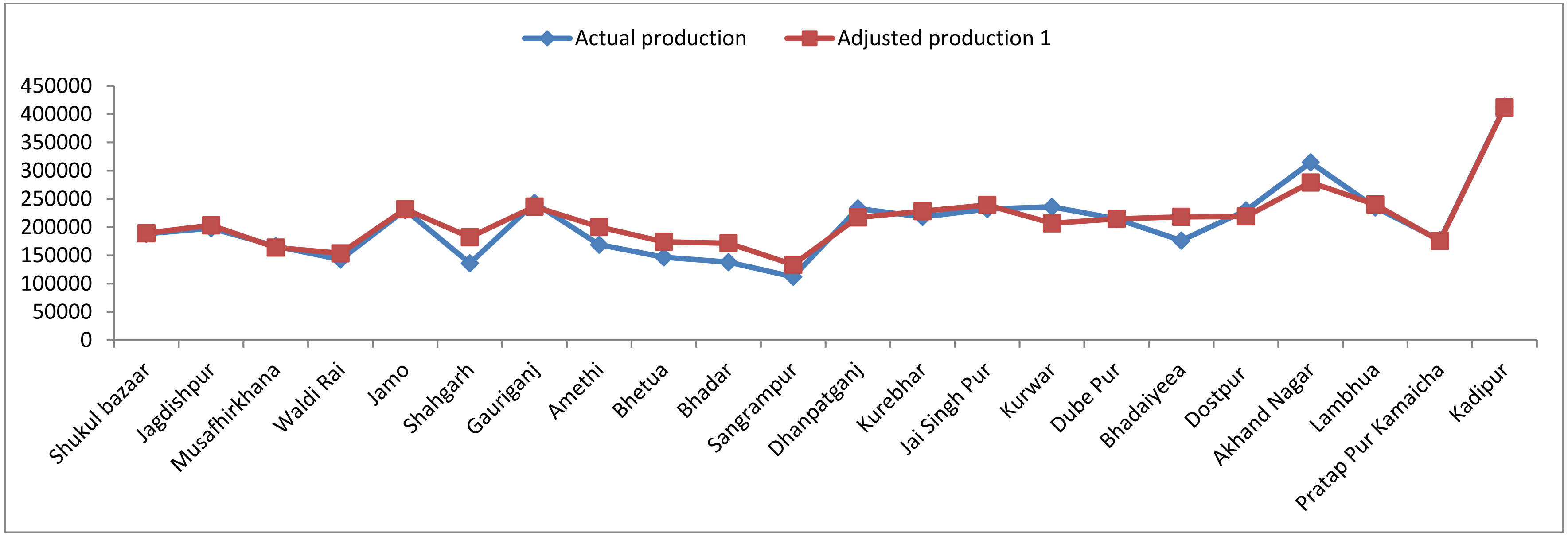

Fig. 4.9: Block estimate of wheat production in Sultanpur district (U.P.) (Z_q & $\tilde{Z}_q^{(1)}$) for the year 2006-07 based on methodology due to Singh *et al.* (2012) .

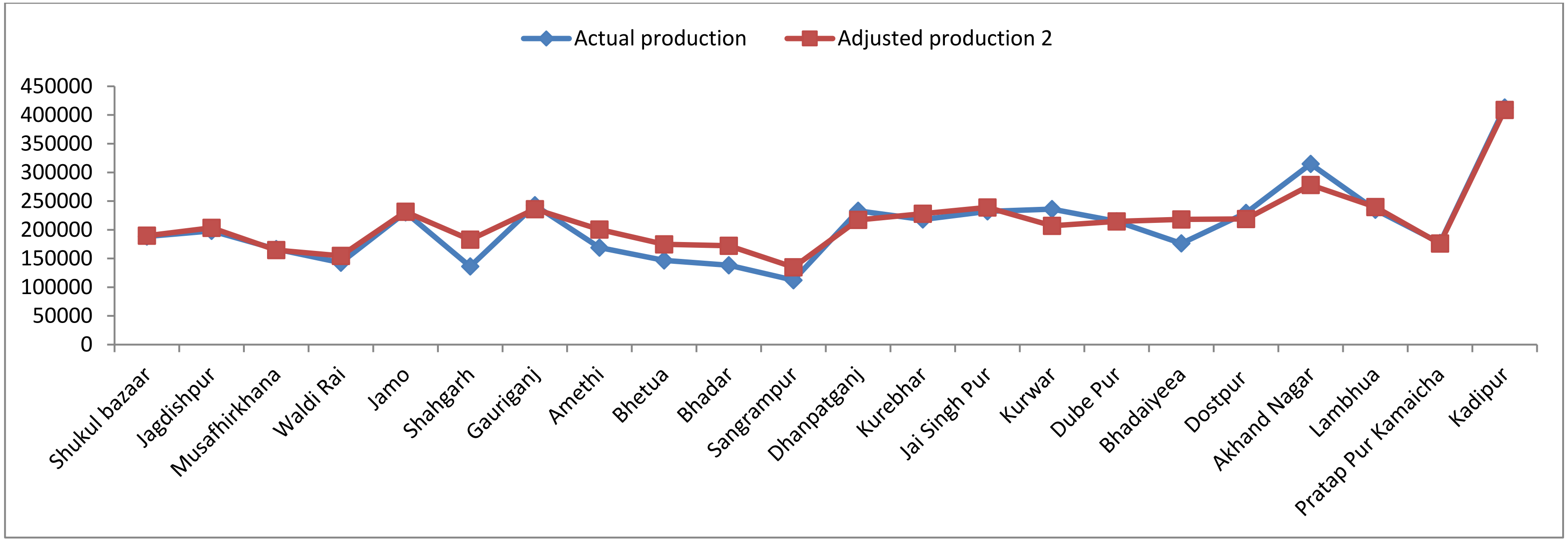

Fig. 4.10: Block estimate of wheat production in Sultanpur district (U.P.) (Z_q & $\tilde{Z}_q^{(2)}$) for the year 2006-07 based on methodology due to Singh *et al.* (2012) .

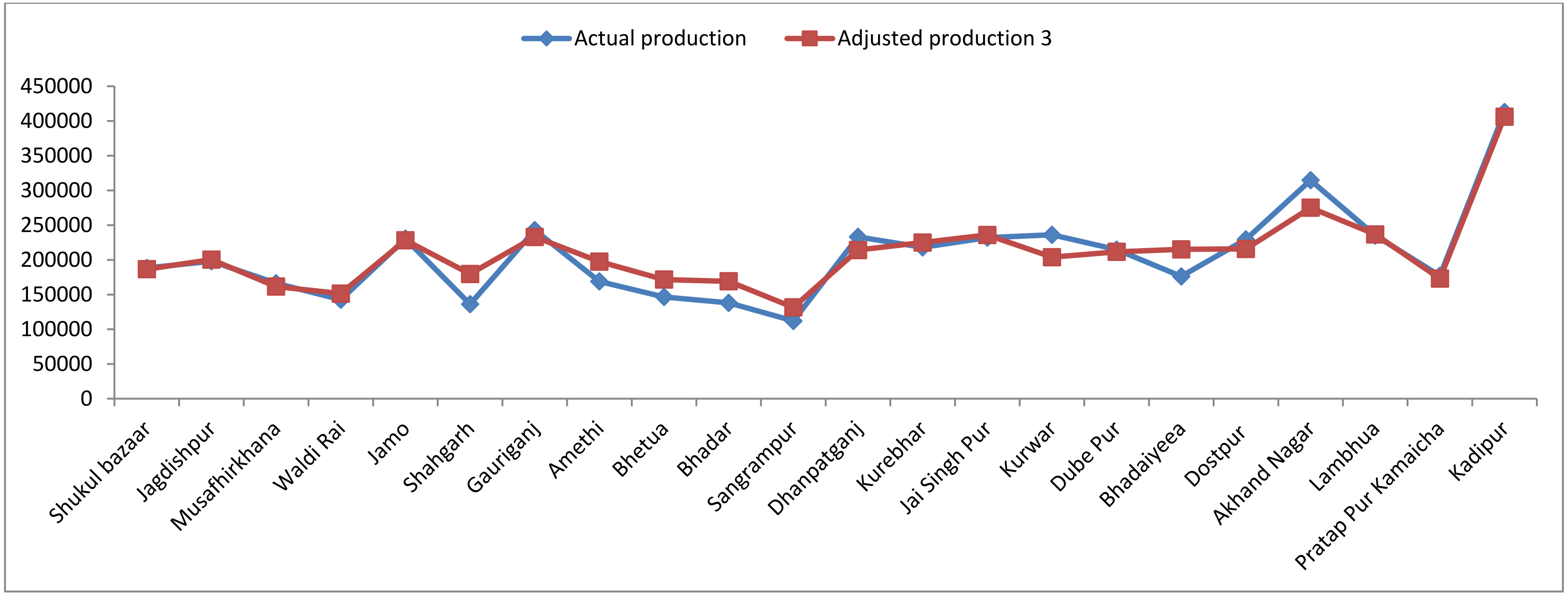

Fig. 4.11: Block estimate of wheat production in Sultanpur district (U.P.) (Z_q & $\tilde{Z}_q^{(3)}$) for the year 2006-07 based on methodology due to Singh *et al.* (2012) .

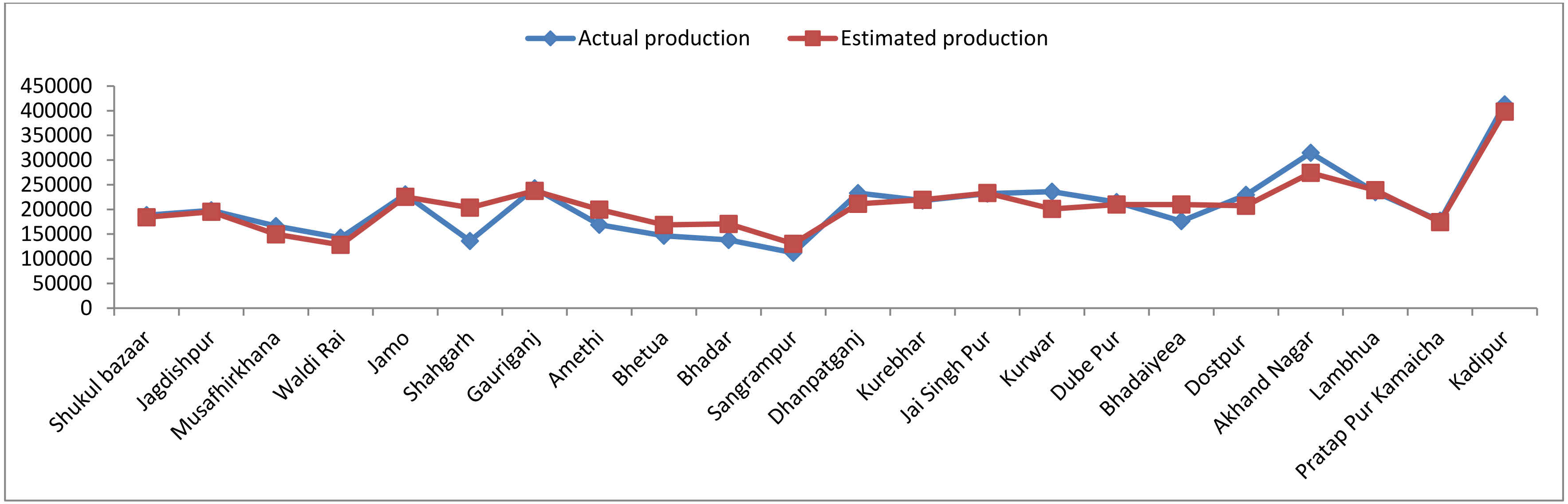

Fig. 4.12: Block estimate of wheat production in Sultanpur district (U.P.) (Z_q & $\hat{Z}_q$) for the year 2006-07 based on proposed methodology

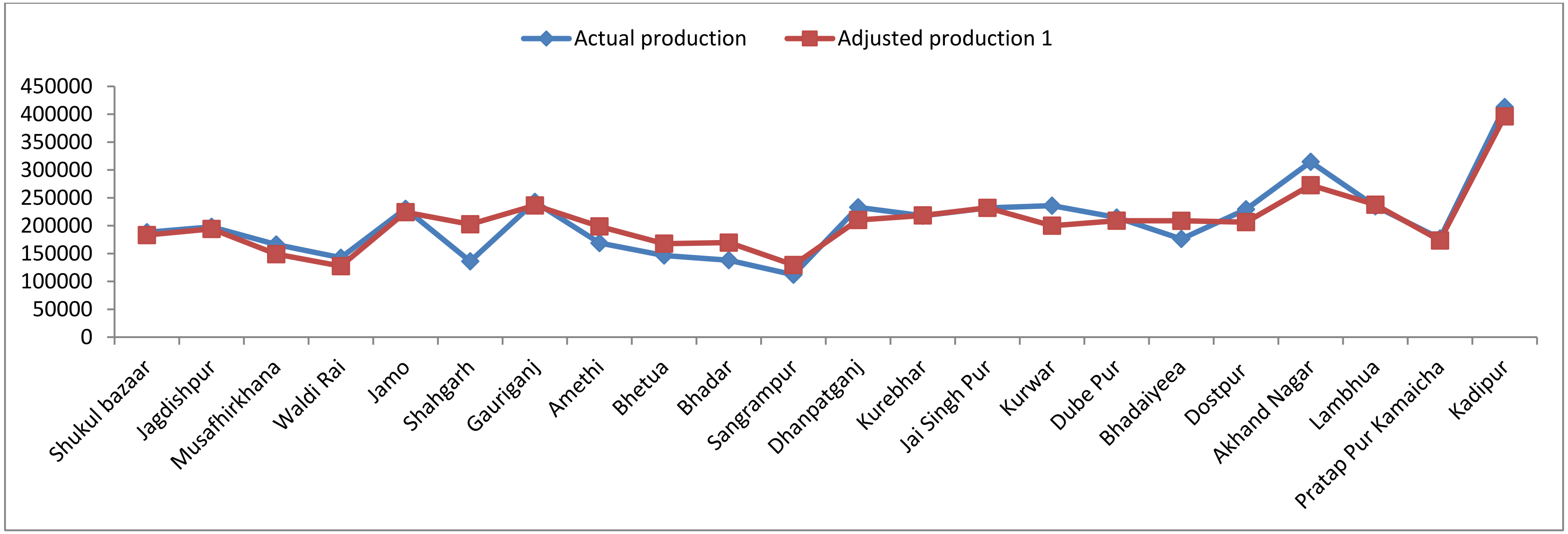

Fig. 4.13: Block estimate of wheat production in Sultanpur district (U.P.) (Z_q & $\tilde{Z}_q^{(1)}$) for the year 2006-07 based on proposed methodology

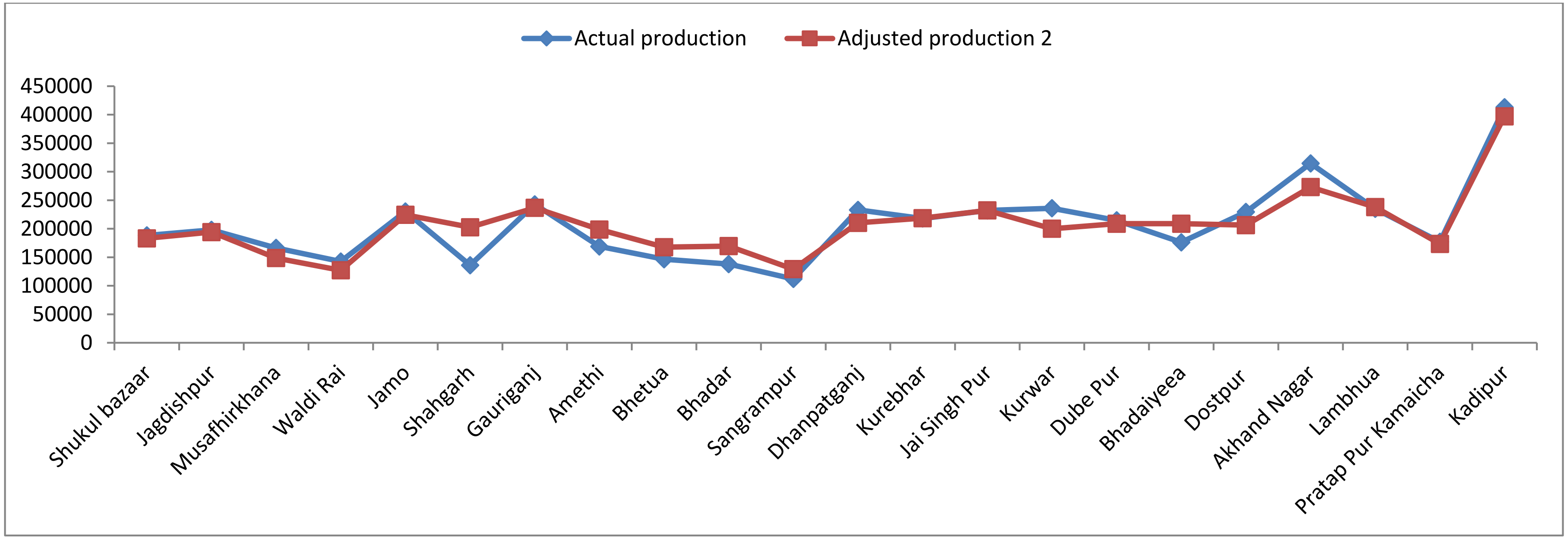

Fig. 4.14: Block estimate of wheat production in Sultanpur district (U.P.) (Z_q & $\tilde{Z}_q^{(2)}$) for the year 2006-07 based on proposed methodology

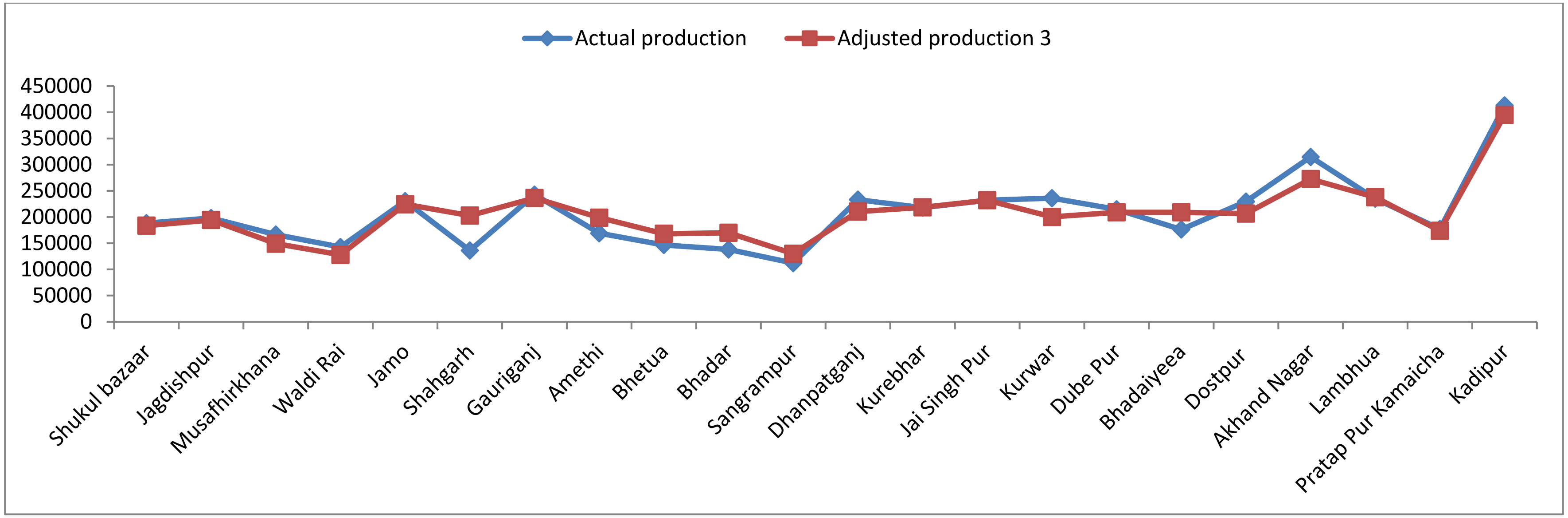

Fig. 4.15: Block estimate of wheat production in Sultanpur district (U.P.) (Z_q & $\tilde{Z}_q^{(3)}$) for the year 2006-07 based on proposed methodology

4.3. Discussion and concluding remarks

The percent standard errors of the estimators using methodology of Sing et al (2012) and the proposed methodology are presented in the Table 4.7.

Table 4.7: Percent standard errors of the estimators

S. No.	Block	Singh et al (2012)				Proposed methodology			
		$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$	$\hat{Z}_q$	$\tilde{Z}_q^{(1)}$	$\tilde{Z}_q^{(2)}$	$\tilde{Z}_q^{(3)}$
1	Shukul bazaar	1.83	2.39	2.45	1.83	2.09	2.14	2.47	2.09
2	Jagdishpur	4.25	4.52	4.41	4.25	3.86	3.89	3.94	3.86
3	Musafhirkhana	7.32	7.48	7.39	7.32	6.82	6.84	6.79	6.82
4	Waldi Rai	20.39	20.44	19.96	20.39	20.48	20.48	19.79	20.48
5	Jamo	3.76	4.06	3.89	3.76	4.10	4.13	4.10	4.10
6	Shahgarh	24.05	24.10	23.41	24.05	17.29	17.30	16.62	17.29
7	Gauriganj	5.95	6.15	5.91	5.95	4.69	4.71	4.62	4.69
8	Amethi	7.15	7.31	7.11	7.15	5.80	5.81	5.71	5.80
9	Bhetua	2.79	3.18	3.28	2.79	2.63	2.67	2.98	2.63
10	Bhadar	8.70	8.83	8.65	8.70	7.17	7.19	7.05	7.17
11	Sangrampur	8.75	8.89	8.89	8.75	7.51	7.53	7.50	7.51
12	Dhanpatganj	1.85	2.41	2.32	1.85	2.34	2.39	2.57	2.34
13	Kurebhar	1.69	2.28	2.15	1.69	2.24	2.29	2.46	2.24
14	Jai Singh Pur	2.94	3.32	3.14	2.94	2.69	2.73	2.81	2.69
15	Kurwar	1.93	2.47	2.43	1.93	2.61	2.65	2.83	2.61
16	Dube Pur	3.45	3.78	3.66	3.45	3.00	3.04	3.14	3.00
17	Bhadaiyeea	1.82	2.38	2.29	1.82	2.39	2.43	2.61	2.39
18	Dostpur	2.91	3.29	3.17	2.91	3.28	3.32	3.40	3.28
19	Akhand Nagar	4.09	4.37	4.11	4.09	4.02	4.05	3.97	4.02
20	Lambhua	3.38	3.71	3.53	3.38	2.93	2.96	3.01	2.93
21	Pratap Pur Kamaicha	7.88	8.03	7.87	7.88	6.44	6.46	6.36	6.44
22	Kadipur	3.49	3.82	3.45	3.49	3.15	3.18	3.08	3.15

It can be observed from results of the Table 4.7 that by adding weather variables in the model as explanatory variables the percent standard errors of the estimators have reduced in general. However, the reduction in the percent standard errors have been found to be quite marginal although it were expected to be substantial reduction. Probably it might be because of variation between blocks in respect of weather variables but in the present investigation the magnitude of weather variable are considered constant for each Block. Moreover, the effects of weather variables have been considered as fixed effect in the model, where the weather variables effects are random in general over years.

Results based on the study conducted by Sisodia & Chandra (2012), Singh et al. (2012), Sisodia and Singh (2012), Sharma et al. (2015) and Sharma & Sisodia (2016) are similar to the results of the present investigation.

In view of the above discussion and the results in the preceding section, the following specific conclusion have been presented below.

1. The inclusion of weather variables in the model as an explanatory variable brought little bit reduction in the percent standard errors of the estimators.
2. On the basis of graphs depicted in the Fig 4.8 to 4.11 and 4.12 to 4.15, it is very obvious that Blocks estimators based on the estimators and actual estimate based on CCEs are almost close to each other in most of the Blocks.

3. Among the four estimators (an unadjusted one and three adjusted) the adjusted estimator $\tilde{Z}_q^{(2)}$ has been found to be precise for the producing the reliable estimates of wheat production at block level in view of per cent standard error and minimum average distance.
4. The Block estimates could be further improved if Block variations in terms of weather variables and other factors can be accounted for in the model.

SUMMARY AND CONCLUSIONS

The present chapter summarizes the work done in all the pervious chapters in the thesis and conclusions are drawn.

The introductory outline of the subject under consideration and research problem undertaken have been clearly described in the Chapter-I. Specific objectives of the study have also been outlined in this chapter.

The extensive review of literature related to present problems of research and small area estimation carried out by researchers from as early as due to Panse *et al.* (1966) till recent time have been presented in the Chapter-II, which provides the wideness and depthness of small area estimation techniques.

The statistical methodologies have been clearly presented in the Chapter-III. The first section has dealt with brief introduction of some earlier works done in small area estimation. A brief review of works related to estimation of Block level crop production due to Sisodia & Singh (2001), Sisodia and Singh (2012), Sisodia & Chandra (2012), Singh *et al.* (2012), Sharma *et al.*(2015) etc. have been presented in section- 3.2.

In the proposed methodology, the regression model at district level consists of weather indices including input factors, explanatory variables as follows

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} ..... + \beta_p X_{ip} + \lambda_1 I_{i1} + \lambda_2 I_{i2} + \lambda_3 I_{i3} + \varepsilon_i \qquad (5.1)$$

i= 1, 2, …, n. Here ε_i is assumed to follow independently normal distribution (0, σ^2). The results of the fitted model are presented in Table 4.4.

The fitted model-5.1 has been used to develop estimators of crop production at Block level. Time series data on wheat production (Z)/yield (Y), per cent irrigated area under wheat crop (X_1), fertilizer consumption for wheat (X_2), per cent relative area under wheat crop as per cent of gross cropped area (X_3), weather indices (I_m) for maximum temperature, minimum temperature and wind velocity pertaining to the period 1990-91 to 2008-09 for Sultanpur district of Uttar Pradesh, India, have been considered for an empirical investigation. Block- wise time series data on these X_i's (i= 1, 2, 3) have also been used for estimating block level wheat production. The data on I_m at district level has also been used at Block level.

Empirical studies have been carried out using statistical methodologies described in Chapter III. The block level estimates of wheat production for year 2006-07 have been obtained using different statistical methodologies in Chapter-IV. The results have been presented in the separate sections in well-formatted tables and figures for visual presentation. The model fitted at district level using proposed methodology have been found to be some stable in terms of model parameters as compared to the Singh et al (2012) model. The results based on statistical methodologies have been summarized in Table 4.7 showing the range of per cent standard errors of the four estimators developed for the estimation of block level estimates of wheat production in Sultanpur district of Uttar Pradesh for the year 2006-07.

The statistical methodology proposed by Singh *et al.* (2012) which considered the multiple linear regression model at district level involving covariates X_1, X_2 and X_3 as an independent variables for prediction of the block level wheat production cannot be recommended as such because the percent standard error of block estimates ranged in general to the maximum 9% with exception of Waldirai and Shahgarh which yielded about 20% and 24% standard error. The block estimates obtained from various estimators have been found to be close to the actual yield. However, in few blocks the estimates are not close to actual production, probably because of local effects of the blocks, which may not uniformly similar to other block. The average distance (D_i) has been found to be minimum, i.e. 20743 in case of $\tilde{Z}_q^{(2)}$. The percent errors of the estimates based on $\tilde{Z}_q^{(2)}$ were almost at par with that of $\hat{Z}_q$ and $\tilde{Z}_q^{(3)}$. Therefore, on the basis of percent standard error of the block estimates and the average distance it can be recommended that $\tilde{Z}_q^{(2)}$

could be preferred in practice for estimating the block estimate among all the other estimators.

However, using proposed methodology the percent standard error of block estimates ranged in general to the maximum 7% with exception of Waldirai and Shahgarh which yielded about 20% and 17% standard error. The block estimates obtained from various estimators have been found to be close to the actual yield. However, in few blocks the estimates are not close to actual production, probably because of local effects of the blocks, which may not uniformly similar to other block. The average distance (D_i) has been found to be minimum, i.e. 24456 in case of $\tilde{Z}_q^{(2)}$. The percent errors of the estimates based on $\hat{Z}_q$, $\tilde{Z}_q^{(2)}$ and $\tilde{Z}_q^{(3)}$ were almost at par. Therefore, on the basis of percent standard error of the block estimates and the average distance it can be recommended that $\tilde{Z}_q^{(2)}$ could be preferred in practice for estimating the block estimate among all the other estimators.

In view of the above discussion and the results in the preceding section, the following specific conclusion have been presented below.

1. The inclusion of weather variables in the model as an explanatory variable brought little bit reduction in the percent standard errors of the estimators.

2. On the basis of graphs depicted in the Fig 4.8 to 4.11 and 4.12 to 4.15, it is very obvious that Blocks estimators based on the estimators and actual estimate based on CCEs are almost close to each other in most of the Blocks.

3. Among the four estimators (an unadjusted one and three adjusted) the adjusted estimator $\tilde{Z}_q^{(2)}$ has been found to be precise for the producing the reliable estimates of wheat production at block level in view of per cent standard error and minimum average distance.
4. The Block estimates could be further improved if Block variations in terms of weather variables and other factors can be accounted for in the model.

BIBLIOGRAPHY

Ahmad, D. G., Najeer and Jalikatti, Vinayak N. (2014), Small area estimation techniques in wheat production – An empirical study. *International Res. J. Agric. Eco. & Stat.*, **5** (2): 231-234.

Battese, G. E., Harter R. M. and Fuller, W. A. (1988), An error components model for prediction of county crop area using survey and satellite data. *J. Amer. Stat. Assoc.*, **83**: 28-36.

Bogue, D. J. (1950), A technique for making extensive population estimates. *J. Amer. Stat. Assoc.*, **45**: 149-163.

Brackstone, G. J. (1987), Small area data: Policy issues and technical challenges. In: Small Area Statistics, R. Platek, J.N.K. Rao, C. E. Sarnal and M. P. Singh, eds, 3-20, John Wiley and Sons, New York.

Chambers, R. and Tzavidis, N. (2006), M-quantile models for small area estimation. *Biometrika*, **93**(2): 255-268.

Chambers, R. L. and Finney, G. A. (1977), Log-linear models for small area estimation. Unpublished paper, *Aus. Bure. of Stat*.

Chandra, H. (2009), Estimation of small area proportions under unit level spatial models. *J. Ind. Soc. Agril. Stat.*, **63**(3): 117-122.

Chandra, H. and Chambers, R. (2009), Multipurpose weighting form small area estimation. *J. Off. Stat.*, **25**(3): 379-395.

Chandra, H. and Chambers, R. L. (2005), Comparing EBLUP and C-EBLUP for small area estimation. *Stat. Trans.*, **7**: 637-648.

Chandra, H., Salvati, N. and Chambers, R. (2007), Small area estimation for spatially correlated population- A comparison of direct and indirect model-based methods. *Stat. Trans*., **8**: 887-906.

Chaudhuri, A. and Maiti, T. (1996), Estimating small domain means on borrowing strength across time and similar domains - A case study. *J. of Ind. Soc. Agril. Stat.*, **49** (Golden Jub. No.), 131-138.

Chaudhuri, A. and Maiti, T. (1994), Borrowing strength from past data in small domain prediction by Kalman filtering—a case study. *Comm. Stat. Theo. Meth*., **23** (12):3507-3514.

Chaudhuri, A. and Adhikary, A. K. (1995), On generalized regression estimators of small domain totals-An evaluation study. *Pak. J. Statist.*, **11**(3): 173-189.

Chaudhuri, A. and Maiti, T. (1997), Small domain estimation by borrowing strength across time and domain- a case study. *Comm. Stat. Comp. Simula*., **26**(4): 1547-1557.

Consiglio, L. D., Falorsi, P. D., Falorsi, S. and Russo, A. (2003), Conditional and unconditional analysis of some small area estimators in complex sampling. *Surv. Method.*, **29**: 53-61.

Datta, G. S., Day, B. and Basawa, I. (1999), Empirical best linear unbiased and empirical Bayes prediction in multivariate small area estimation. *J. of Stat. Plan. and Inf.*, 75: 269–279.

Drew, J. D., Singh, M. P. and Choudhury, G. H. (1982), Evaluation of small area estimation techniques for the Canadian Labour Force Survey. *Surv. Method.*, **8**: 17-47.

Ericksen, E. P. (1973), A model for combining sample data and symptomatic indicators to obtain population estimates for local areas. *Demography*, **10**: 137-160.

Ericksen, E. P. (1974), A regression method for estimating populations of local areas. *J. Amer. Stat. Assoc.*, **69**: 867-875.

Fay, R. E. (1987), Application of multivariate regression to small domain estimation, Small Area Statistics, Eds.: R. Platek, J. N. K. Rao, C. E. Sarndal, and M. P. Singh. Wiley, New York, pp. 91-102.

Fay, R. E. and Herriot, R. A. (1979), Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Stat. Assoc.*, **74**: 269-277.

Fuller, W. A., and Harter, R. M. (1987), The multivariate components of variance modelfor small area estimation, Small Area Statistics, Eds.: R. Platek, J. N. K. Rao, C. E. Sarndal, and M. P. Singh. Wiley, New York, pp. 103-123.

Ghangurde, P. D. and Singh, M. P. (1977), Synthetic estimates in periodic household surveys. *Surv. Method.*, **3**: 152-181.

Ghose, M. and Rao, J. N. K. (1994), Small area estimation: an appraisal. *Statist. Sci.*, **9** (1): 55-93.

Gonzalez, M. E. (1973), Use and evaluation of synthetic estimators. Proceedings of the Social Sciences Section, *Amer. Stat. Assoc.*, 33-36.

Holt, D., Smith, T. M. F. and Tomberlin, T. J. (1979), A model-based approach to estimation for small groups of population. *J. Amer. Stat. Assoc.*, **74**:405-410.

Jiang, J. and Lahiri, P. (2006), Mixed model prediction and small area estimation. *Test*, **15**: 1–96.

Johnson, F. A., Chandra, H., Brown, J. J. and Padmadas, S. S. (2010), District-level estimates of institutional births in Ghana: Application of small area estimation technique using census and DHS data. *J. Off. Stat.*, **26** (2): 341-359.

Kalsbeek, W. D. (1973), A method for obtaining local post-censal estimates for several types of variables. Unpublished Ph. D. Thesis, University of Michigan, Ann. Arbor. Michigan.

Marker, D. A. (1983), Organization of small area estimators. In Proceeding of Survey Research Methods Section, 409-414, *Amer. Stat. Assoc.*, Washington, D. C.

Montgomery, D. C. and Peck, E. A. (1982), Introduction to linear regression analysis. John Willy & Sons, Canada.

Nichol, S. (1977), A regression approach to small area estimation. Unpublished manuscript, *Aust. Bure. of Stat.*, Canberra, Australia.

Pandey, P. S. and Kathuria, O. P. (1995), Some composite estimators for small area estimation. *J. Ind. Soc. Agril. Stat.*, **47** (3): 262-272.

Panse, V. G., Rajagopalan, M. and Pillai, S. (1966), Estimation of crop yields for small areas. *Biometrics*, **66**: 374-388.

Prabhu-Ajgaonkar, S. G. (1988), Woodruff's technique of estimating small area statistics; Symposium on small area statistics, organized on 17th Dec. 1987 at C.T.R.I., Rajahmundry (A.P.), *J. Ind. Soc. Agril. Stat.*, **XL** (1): 64-65.

Prasad, N. G. N. and Rao, J. N. K. (1990), The estimation of mean squared error of small area estimators. *J. Amer. Stat. Assoc.*, **85**: 163-171.

Prasad, N. G. N. and Rao, J. N. K. (1999), On robust small area estimation using a simple random effects model. *Surv. Methodo.*, **25**: 67-72.

Pratesi, M. and Salvati, N. (2008), Small area estimation: the EBLUP estimator with autoregressive random area effects. *Statist. Methods Appl.*, **17**: 113-141.

Purcell, N. J. and Kish, L. (1979), Estimation of small domains. *Biometrics*, **35**: 365-384.

Purcell, N. J. and Kish, L. (1980), Potential estimates for local areas (or domains), *Internet. Stat. Rev.*, **48**: 3-18.

Rao, J. N. K. (2003), Small Area Estimation. Wiley- Interscience, Wiley Series in *Surv. Methodo.*

Salvati, N., Tzavidis, N., Pratesi, M. and Chambers, R. (2007-09), Small Area Estimation via M-Quantile Geographically Weighted Regression. Cathie Marsh Center for Census and Survey Research Working Paper, University of Manchester, U.K.

Sarndal, C. E. and Hidiroglou, M. A. (1989), Small domain estimation: a conditional analysis. *J. Amer. Stat. Assoc*., **84**: 266-275.

Salvati, N., Pratesi, M., Tzavidis, N. and Chambers, R. (2007), Spatial M-quantile models for small area estimation. Working Paper (15-08), Centre for Statistical and survey Methodology, The University of Wollongong, pp: 1-16.

Schiable, W. L. (1978), Choosing weights for composite estimators for small area statistics. Proceeding of the Survey Research Methods Section, 741-746, *Amer. Stat. Assoc*., Washington, DC.

Sharma, S. D., Srivastava, A.K. and Sud, U.C. (2004), Small area crop estimation for crop yield estimates at Gram Panchayat level. *J. Ind. Soc. Agril. Stat.*, **57**: 26-37.

Sharma, M.K., Sisodia, B.V.S., and Kumar Sunil (2015), On choice of explanatory variables in linear model for estimation of crop production at smaller geographical area. *J. Ind. Soc. Agril. Stat.,* **69**(3): 315-326.

Sharma, M.K. and Sisodia, B.V.S. (2016), Estimation of crop production for smaller geographical area an application of discriminant function analysis. *J. Stat. Theory and practice*, **10**(2): 444-455.

Singh, B., Sisodia, B. V. S., Singh, A. and Kaushal, R. P. (2012), A note on estimation methods of crop production at block level. *Journal of Appl. Stat.*, **39** (9): 2015-2027.

Sisodia, B.V.S., Chandra, Hukum and Shukla, D. (2015 0, An application of area level model for estimation of crop production at smaller geographical level in India. *Int. J. Agril. And Stat. Science*, **11**(1): 97-104.

Singh, D. (1968), Double sampling and its application in agriculture, contribution in statistics and agricultural science. Dr. V.G. Panse lecture, Indian Society of Agricultural Statistics, New Delhi.

Sinha, S. K. and Rao, J. N. K. (2009), Robust small area estimation. *The Canadian J. of Stat.*, **37**(3): 381-399.

Sisodia, B. V. S. and Chandra, H. (2012), Estimation of crop production at geographic level in India. *J. Ind. Soc. Agril. Stat*., **66** (2):313-320.

Sisodia, B. V. S. and Singh, A. (2001), On small area estimation – An empirical study. *J. Ind. Soc. Agril. Stat.*, **54** (3): 303-316.

Sisodia, B. V. S. and Singh, A. (2012), On small area estimation- An application in agriculture. *Journal of Ind. Soc. Agril. Stat*., **66** (3): 391-400.

Srivastava, A. K. (2009), Some Aspects of Estimating Poverty at Small Area Level. *Journal of Ind. Soc. Agril. Stat.*, **63** (1), 2009 : 1-23.

Srivastava, A. K., Ahuja, D. L., Bhatia, D.K. and Mathur, D.C. (1998), Small area estimation techniques in agriculture, Souvenir, Advances in Agricultural Statistics and Computer Application. I.A.S.R.I., New Delhi, 32-45.

Srivastava, A. K., Ahuja, D. L., Mathur, D. C. and Sethi, S. C. (1999), Project Report on Estimation of Crop Yields for Small Areas, IASRI publication.

Srivastava, A. K., Sud, U. C., and Chandra, H. (2007), Small Area Estimation - An application to National Sample Survey Data, *Journal of Ind. Soc. Agril. Stat*., **61** (2): 249-254.

Stasny, E. A., Geol, P. K. and Rumsey, D. J. (1991), County estimates of wheat production. *Survey methodology*, 17 (2): 211-225.

Sud, U. C., Srivastva, Bathla., A. K., Mathur, D.C. and Jha, G.K. (2001), Project Report on crop yield estimation at block level using farmer estimates. IASRI publication.

Sukhatme, P. V. and Sukhatme, B. V. (1970), Sampling Theory of surveys with Applications, Iowa State University Press, Ames, IA.

Wolter, K. M. and Causey, B. D. (1991), Evaluation of procedures for improving population estimates for small areas. *Journal of Amer. Stat. Assoc*., **86**: 278-284.

Appendix –I

Table I.1: Data on gross cropped area, irrigated area under wheat crop, production and yield for Sultanpur district of Uttar Pradesh

Year	Gross cropped area (ha)	Area under wheat crop (ha)	Irrigated area under wheat crop (ha)	Production (q)	Yield (q/ha)
1990-91	434320	148819	147050	273285	18.36
1991-92	431260	149838	148756	320840	21.41
1992-93	446124	157533	156187	320651	20.35
1993-94	434131	156996	155901	319343	20.34
1994-95	440509	159895	159057	352284	22.03
1995-96	436673	159907	158801	333363	20.85
1996-97	443159	163067	162395	440198	26.99
1997-98	441218	162420	161787	428513	26.38
1998-99	442424	163360	162571	393140	24.07
1999-00	443869	163955	163155	443276	27.04
2000-01	442040	163002	162013	413259	25.35
2001-02	439529	163760	163442	431839	26.37
2002-03	402114	165268	165073	424135	25.66
2003-04	437768	166983	166796	462053	27.67
2004-05	432146	167024	166805	422457	25.29
2005-06	433649	167483	167312	432233	25.81
2006-07	433886	168112	167928	452837	26.94
2007-08	436557	167123	167024	466983	27.94
2008-09	436251	162633	162535	473625	29.12

Appendix –I

Table I.2: Data on percent relative area of wheat as gross cropped area, percent irrigated area of wheat under wheat crop and fertilizer consumption for Sultanpur district of U. P.

Year	Per cent relative area of wheat as gross cropped area	Per cent irrigated area under wheat crop	Fertilizer Consumption (kg/ha)
1990-91	34.26	98.81	71.95
1991-92	34.74	99.28	75.76
1992-93	35.31	99.15	66.03
1993-94	36.16	99.30	61.94
1994-95	36.30	99.48	79.26
1995-96	36.62	99.31	82.71
1996-97	36.80	99.59	103.45
1997-98	36.81	99.61	91.69
1998-99	36.92	99.52	89.38
1999-00	36.94	99.51	95.19
2000-01	36.87	99.39	93.85
2001-02	37.26	99.81	94.39
2002-03	41.10	99.88	96.10
2003-04	38.14	99.89	101.98
2004-05	38.65	99.87	103.51
2005-06	38.62	99.90	107.57
2006-07	38.75	99.89	115.61
2007-08	38.28	99.94	127.98
2008-09	37.28	99.94	134.31

Appendix -II

Table II.1: Yield of wheat based on crop cutting experiments (CCEs) for the year 2006-07 at blocks level for Sultanpur district of Uttar Pradesh

S.No.	Blocks	Yield (q/ha)
1	Shukul bazaar	27.78
2	Jagdishpur	27.95
3	Musafhirkhana	31.44
4	Waldi Rai	28.5
5	Jamo	25.54
6	Shahgarh	19.13
7	Gauriganj	27.41
8	Amethi	24.22
9	Bhetua	24.35
10	Bhadar	24.23
11	Sangrampur	27.1
12	Dhanpatganj	29.04
13	Kurebhar	26.73
14	Jai Singh Pur	27.08
15	Kurwar	30.65
16	Dube Pur	27.99
17	Bhadaiyeea	22.39
18	Dostpur	29.82
19	Akhand Nagar	28.61
20	Lambhua	26.02
21	Pratap Pur Kamaicha	29.6
22	Kadipur	29.64

APPENDIX –III

Detail of weekly weather data

Maximum Temperature

Year/SMW	44	45	46	47	48	49	50	51	52	1	2	3	4
1990-91	31.41	31.07	31.37	29.85	26.30	25.85	25.04	24.81	23.68	18.60	22.21	26.28	27.78
1991-92	31.07	28.52	27.05	26.37	26.15	25.60	25.37	23.51	19.16	19.95	21.00	21.21	25.31
1992-93	31.67	30.44	30.47	27.70	25.44	25.61	24.04	23.94	23.06	20.25	22.67	22.78	25.04
1993-94	31.57	30.24	30.27	27.27	23.92	26.65	25.54	24.07	25.33	22.54	20.67	21.48	21.70
1994-95	30.90	30.65	29.51	28.60	25.94	26.71	24.48	24.90	25.57	25.65	22.97	21.70	25.27
1995-96	31.44	29.77	29.15	27.57	26.87	25.31	26.08	26.01	23.55	22.08	16.68	21.27	23.07
1996-97	31.65	31.04	29.34	27.20	26.72	24.91	24.28	25.01	24.88	23.35	23.28	19.34	23.15
1997-98	29.11	30.41	28.60	26.25	22.67	23.6	18.81	18.12	18.41	22.88	23.32	20.97	21.87
1998-99	32.22	30.12	27.45	27.78	28.14	26.27	25.54	23.95	18.36	16.52	19.65	21.90	23.08
1999-2000	33.58	31.32	29.98	27.24	27.37	25.32	25.28	24.58	20.93	21.98	17.60	14.37	25.24
2000-01	30.87	32.21	25.84	28.72	26.42	26.54	25.72	25.12	25.01	14.71	22.52	23.48	21.20
2001-02	32.78	29.55	29.82	27.97	26.42	26.57	18.44	21.04	23.32	18.72	19.65	24.42	24.37
2002-03	31.35	29.92	27.75	28.68	27.31	26.24	25.35	24.35	19.41	21.30	23.51	23.15	22.01
2003-04	31.02	31.85	28.62	28.55	25.45	24.44	27.10	23.22	17.72	16.12	12.17	13.38	21.24
2004-05	29.32	28.94	29.52	27.15	22.22	25.91	24.97	19.61	19.80	13.91	16.90	19.74	20.54
2005-06	29.38	28.60	29.01	27.68	27.92	24.80	24.15	23.81	22.52	23.17	22.97	22.52	21.60
2006-07	30.01	29.72	29.08	28.71	25.65	27.94	24.34	25.35	22.58	24.07	23.28	25.80	23.60
2007-08	29.55	29.12	27.51	26.34	25.48	24.04	22.90	22.45	23.58	22.14	22.47	24.20	24.50
2008-09	30.10	29.25	27.57	25.92	25.57	25.17	24.62	18.91	22.41	23.02	25.00	23.20	18.80

***SMW-Standard Meteorological Week**

Detail of weekly weather data

Maximum Temperature

Year/SMW	5	6	7	8	9	10	11	12	13	14	15
1990-91	26.48	27.01	23.97	24.67	25.3	29.02	31.6	34.42	28.22	33.5	38.3
1991-92	25.94	28.52	27.11	28.01	27.37	30.22	34.61	33.37	36.52	36.97	35.97
1992-93	23.67	23.08	24.22	25.38	29.36	31.2	32.47	34.9	35.51	36.91	40.08
1993-94	16.15	21.08	23.22	24.38	29.26	32.2	32.27	34.58	35.5	36.81	38.08
1994-95	23.28	23.94	26.25	25.65	28.72	31	34.61	36.07	36.27	36.17	38.58
1995-96	24.71	26.3	26.17	26.55	28.41	26.6	31.24	36.35	32.9	37.02	38.25
1996-97	23.95	23.71	25.54	28.92	26.87	32.02	34.7	34.25	36.51	38.51	39.02
1997-98	24.1	24.77	25.97	27.12	30.18	32.6	33.18	31.81	32.77	32.02	35.94
1998-99	24.77	25.4	29.6	25.34	26.34	29	29.45	32.5	31.24	35.02	38.62
1999-2000	20.17	24.05	29.17	30.12	30.1	34.72	33.7	35.85	37.81	40.48	40.05
2000-01	24.04	23.08	24.54	27.22	29.11	31.38	32	32.62	37.48	39.48	39.44
2001-02	25.77	26.55	27.72	30.18	29.94	31.34	33.6	34.1	36.08	36.81	38.75
2002-03	23.2	24.97	25.07	29.98	28.45	30.9	33.1	36.28	36.84	36.61	37.51
2003-04	22.07	25.78	27.07	26.2	30.35	28.78	31.04	33.3	35.62	36.71	41.02
2004-05	22.22	24.72	27.6	29.54	31.76	31.81	35.44	38.41	39.41	38.62	40.75
2005-06	21.84	25.58	28.14	27.42	32.78	31.45	33.08	34.95	36.08	38.48	38.7
2006-07	26.84	27.72	31.55	33.77	32.31	30.38	29.68	35.47	36.28	38.95	39.74
2007-08	27.7	23.08	22.68	27.05	26.31	28.68	26.94	33.14	36.64	38.47	37.34
2008-09	18.04	24.48	24.51	27.51	29.67	31.58	33.4	35.12	36.42	33.64	40.11

Detail of weekly weather data

Minimum Temperature

Year/SMW	44	45	46	47	48	49	50	51	52	1	2	3
1990-91	13.87	13.85	13.67	12.05	8.12	7.05	7.75	5.97	7.88	4.45	6.17	10.64
1991-92	13.5	12.74	11.01	7.85	8.3	6.15	7.64	9.91	8.88	1.5	5.3	4.31
1992-93	16.57	12.04	14.24	13.27	9.25	9.42	6.95	5.91	7.8	4.4	7.48	4.91
1993-94	13.58	14.25	12.42	10.12	8.51	9.65	5.42	4.08	6.18	9.91	9.44	4.61
1994-95	13.91	12.92	12.1	9.77	10.57	10.47	5.8	3.64	6.58	7.54	10.41	6.97
1995-96	13.28	15.28	14	10.81	8.45	6.08	8.71	9.75	9.76	6.47	9.42	7.34
1996-97	15.84	11.82	9.65	8.61	6.92	5.75	3.82	5.15	5.35	11.1	8.02	9.54
1997-98	16.2	13.34	12.91	15.04	10.58	13.15	11.52	8.07	7.65	6.15	5.5	7.42
1998-99	14.2	13.24	12.7	13.04	10.5	12.15	10.52	8.17	7.85	6.25	4.5	6.42
1999-2000	15.57	11.42	10.65	8.52	7.92	5.72	3.62	5.35	5.45	7.45	8.12	9.31
2000-01	11.38	10.28	10.88	11.12	10.82	9.48	8.82	10.9	5.43	5.25	7.17	8.58
2001-02	12.88	9.43	8.27	7.28	8.37	6.31	4.12	5.32	8.47	8.15	6.24	7.45
2002-03	15.1	15.57	14.12	13.24	10.95	11.27	9.92	9.12	8.25	6.45	5.42	7.45
2003-04	21.65	18.24	17.74	16.2	9.1	8.02	10.94	9.27	7.73	9.91	7.11	7.92
2004-05	11.88	10.08	10.3	12.28	10.62	8.48	8.32	11.9	5.53	4.25	7.27	8.52
2005-06	12.58	9.3	8.47	7.58	8.7	5.31	3.2	5.2	8.07	8.05	6.4	7.64
2006-07	16.1	15.27	14.22	13.74	7.95	11.17	9.82	9.22	8.5	6.65	3.42	9.45
2007-08	16.84	15.97	12.25	9.97	10.8	9.5	9.75	5.5	6.08	6.01	4.85	5.81
2008-09	16.51	16.38	14.14	12.68	13.94	11.88	12.71	12.98	7.62	5.92	8.7	8.74

Detail of weekly weather data

Minimum Temperature

Year/SMW	4	5	6	7	8	9	10	11	12	13	14	15
1990-91	9.45	8.72	11.88	9.7	12.18	11.71	11.07	15.87	11.65	13.48	16.34	19.78
1991-92	7.71	9.51	11.04	11.98	13.54	14.6	14.3	17.75	16.92	12.14	21.67	20.55
1992-93	8.84	10.78	9.04	7.4	8.21	11.5	12.5	13.68	17.05	18.9	18.01	20.7
1993-94	4.92	7.51	9.84	13.67	8.2	13.4	12.8	14.84	14.68	12.75	15.58	18.9
1994-95	8.3	7.95	10.75	8.58	10.01	9.45	10.88	16.6	16.67	18.1	19.08	18.71
1995-96	4.15	6.45	9.98	12.94	10.24	11.3	13.55	13.62	16.52	17.6	14.3	18.32
1996-97	7.51	5.84	9.57	9.84	11.75	12.11	14.42	17.24	16.31	17.22	15.47	16.41
1997-98	4.85	7.54	7.2	8.24	8.28	11.41	12	13.95	15.47	14.55	16.38	20.47
1998-99	4.75	7.24	7.12	8.34	8.48	10.41	11.22	13.45	14.47	15.45	16.24	19.47
1999-2000	7.21	6.84	8.57	9.44	11.25	11.89	13.42	17.14	16.61	16.98	15.89	17.41
2000-01	9.12	8.12	7.14	9.72	10.47	11.17	13.87	16.37	17.47	17.27	19.88	22.92
2001-02	9.18	10.23	11.64	13.13	12.27	14.78	16.14	15.21	17.14	16.61	17.12	17.89
2002-03	5.22	8.28	11.85	13.13	14.21	13.57	14.92	15.18	15.54	16.27	18.84	21.95
2003-04	8.85	8.22	7.14	9.12	11.17	10.7	13.21	16.37	17.27	17.17	20.28	22.91
2004-05	10.12	8.2	6.14	9.92	11.47	11.7	13.11	17.37	17.87	17.07	19.98	21.91
2005-06	9.58	10.3	11.14	13.3	11.27	13.78	16.44	15.91	18.14	16.71	17	17.15
2006-07	5.42	8.58	11.45	13.3	14.51	13.27	15.92	15.48	15.34	16.37	18.74	20.95
2007-08	6.61	12.14	13.34	11.77	12.57	13.48	12	14.17	16.9	17.75	19.51	22.65
2008-09	7.42	5.64	10.81	6.37	11.12	11.3	15.51	15.78	17.45	17.41	17.25	21.34

Detail of weekly weather data

Wind Velocity

Year/SMW	44	45	46	47	48	49	50	51	52	1	2	3
1990-91	0.725	0.71	0.73	0.995	1.28	0.465	0.925	0.28	1.71	1.565	2.28	2.57
1991-92	0.36	0.57	0.585	0.565	0.47	0.465	0.725	0.73	1.75	0.995	0.265	1.995
1992-93	0.84	0.58	0.56	0.875	0.505	0.395	0.725	0.63	1.595	0.925	0.265	1.99
1993-94	1.95	0.62	0.45	2.285	0.64	0.51	0.71	0.57	0.855	2.135	5.28	4.565
1994-95	0.85	1.345	0.57	0.76	0.67	0.74	1.995	0.57	1.285	0.88	2.425	3.28
1995-96	0.425	0.85	1.71	0.71	1.28	1.015	1.995	2.28	0.875	3.28	5.285	4.565
1996-97	0.615	0.95	0.78	0.585	0.615	1.025	0.405	1.105	0.4	1.14	0.855	5.14
1997-98	0.855	0.51	0.72	1.71	3.285	3.355	1.28	2.995	1.25	0.78	0.67	0.425
1998-99	0.675	1.28	0.48	0.815	1.435	0.675	0.515	0.675	1.285	1.425	1.28	1.995
1999-2000	0.73	1.19	0.785	0.76	0.4	0.795	1.135	0.615	0.695	1.71	2.42	2.14
2000-01	1.28	1.14	0.73	0.305	1.425	0.85	0.4	0.855	0.505	0.85	1.135	4.135
2001-02	0.835	1.06	0.785	0.845	1.74	1.74	1	1.055	0.475	3.285	1.565	0.705
2002-03	0.4	0.53	1.14	0.685	0.68	0.53	0.585	0.57	2.25	0.615	0.675	2.855
2003-04	0.71	0.995	0.905	1.28	0.525	0.32	1.325	0.43	2.32	0.995	0.635	0.525
2004-05	2.285	0.615	0.77	0.91	0.565	0.565	0.675	1.28	2.89	0.995	0.71	3.565
2005-06	0.99	0.86	2.71	0.675	1.14	2.14	0.905	1.135	0.875	1.995	1.28	3.425
2006-07	0.425	0.915	0.425	2.565	2.285	0.995	2.71	1.245	1.57	2.71	3.14	4.28
2007-08	0.425	0.46	1.495	0.72	0.565	0.575	0.97	1.85	0.785	2.425	1.71	2.28
2008-09	0.57	0.415	0.54	1.345	0.35	0.34	0.565	0.995	0.555	0.5	1.85	1.995

Detail of weekly weather data

Wind Velocity

Year/SMW	4	5	6	7	8	9	10	11	12	13	14	15
1990-91	0.855	2.855	0.57	1.04	1.51	0.84	0.81	1.14	1.21	1.21	1.28	1.93
1991-92	0.76	1.285	1.425	3.21	2.57	4.14	1.86	2.14	4.28	3.28	3.14	2.86
1992-93	0.825	1.275	1.38	3.71	2.57	1.24	2.12	0.89	0.77	1.14	6.43	3.14
1993-94	5.565	0.71	1.015	3.43	3.43	1.57	2.86	5.00	4.43	3.71	2.57	4.71
1994-95	2.995	0.995	2.995	1.00	4.14	3.00	4.14	3.71	3.57	4.00	3.71	3.14
1995-96	2.135	2.71	1.855	2.57	3.85	3.14	2.71	2.71	3.71	3.29	2.71	3.00
1996-97	3.425	1.565	1.71	2.71	2.00	2.57	3.71	3.28	3.71	3.71	2.71	2.68
1997-98	3.135	0.71	3.71	3.57	2.14	3.14	2.00	4.28	3.28	2.00	3.28	3.86
1998-99	1.14	2.14	1.71	1.00	2.14	3.00	2.71	1.57	2.86	2.29	3.00	2.35
1999-2000	3.57	1.855	1.85	2.57	2.28	2.00	3.00	3.00	2.86	3.71	3.29	3.57
2000-01	0.995	6.285	4.135	1.86	2.14	1.28	2.43	2.14	2.00	1.57	3.07	2.28
2001-02	5.425	2.995	5.565	1.42	1.42	1.71	1.42	1.42	2.28	1.42	1.42	1.42
2002-03	3.42	3.14	4.425	2.00	2.28	2.00	0.78	0.41	0.58	0.85	3.43	3.14
2003-04	0.995	2.57	0.565	2.43	4.29	2.86	4.00	2.14	3.14	3.14	4.00	3.28
2004-05	3.57	3.425	1.85	0.57	2.71	3.57	1.42	1.86	3.57	2.14	3.00	2.57
2005-06	2.71	2.85	3.71	3.71	4.00	5.00	3.43	2.00	4.00	2.71	3.86	3.85
2006-07	3.135	2.425	0.73	1.71	3.28	2.28	2.43	2.43	2.43	2.43	2.71	2.28
2007-08	1.14	1.57	3.855	2.28	3.57	2.43	1.85	6.43	3.43	2.57	5.14	4.71
2008-09	1.42	1.28	2.565	2.85	1.86	1.11	2.28	3.28	2.00	2.85	3.43	2.43

Detail of weekly weather data

Weighted Weather Indices

Year/SMW	I_{11}	I_{21}	I_{31}
1990-91	27.42	11.75	0.09
1991-92	9.00	12.48	1.87
1992-93	13.08	12.81	1.46
1993-94	12.27	11.48	0.72
1994-95	14.23	12.62	2.34
1995-96	15.31	12.50	1.15
1996-97	8.50	13.39	2.23
1997-98	3.06	13.42	4.56
1998-99	13.16	12.81	1.75
1999-2000	2.20	12.96	2.03
2000-01	7.86	14.07	2.53
2001-02	6.62	12.82	3.46
2002-03	7.38	14.31	2.19
2003-04	4.22	15.30	1.96
2004-05	-1.90	14.08	1.63
2005-06	4.32	12.80	3.41
2006-07	4.11	14.56	1.00
2007-08	9.88	13.85	3.43
2008-09	2.00	15.48	2.00

DEPARTMENT OF AGRICULTURAL STATISTICS
NARENDRA DEVA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
NARENDRA NAGAR (KUMARGANJ), FAIZABAD-224 229 (U.P.)

Title: ***"An empirical study for estimation of crop production at smaller geographical area based on various factors including weather variables".***

Major Advisor:
Dr. B. V. S. SISODIA
Professor & Head
Deptt. Of Agricultural Statistics

Name of Student:
ASHISH SHUKLA
I. D. No. A-8263/14
M. Sc. (Ag.) Agricultural Statistics

ABSTRACT

The present thesis consists five chapters including summary and conclusion. The Chapter-I introduces the subject of investigation and main objectives are also outlined in this chapter.

The review of relevant literature is presented in Chapter-II. The research methodology and material are described in Chapter-III. The time series data on wheat production, per cent irrigated area under wheat crop, fertilizer consumption for wheat, per cent relative area under wheat crop as per cent of gross cropped area, weather indices for maximum temperature, minimum temperature and wind velocity pertaining to the period 1990-91 to 2008-09 for Sultanpur district of Uttar Pradesh, India, have been considered for an empirical investigation. Block- wise time series data have also been used for estimating block level wheat production. The data on weather indices at district level has also been used at Block level.

The main objectives of the Thesis are to develop empirical predictive model for estimation of crop production at Block level using weather variables and other factors.

The result are presented in Chapter- IV. The inclusion of weather variables in the model as an explanatory variable brought little bit reduction in the percent standard errors of the estimators for wheat production at Block level.

(B. V. S. Sisodia) **(ASHISH SHUKLA)**