ABSTRACT

The term “Small Area” usually denotes subsets of any finite population. These subsets may refer either to geographic sub – populations ( or domains ), such as, school districts, unemployment insurance regions, metropolitan areas, health service areas and others or to some socio – economic, demographic or ethnic groups within a large geographic area, such as, population of persons belonging to a rare caste group in a village, population of households in a village with five or more eligible couples, persons engaged in any small scale industry in a village, population of disabled persons in a community, etc. It may also denote a cross classification of a small geographic area and a specific demographic or industrial group. Purcell and Kish ( 1979 ) defined approximately the range of various types of domains as: ( i ) major domains, ( ii ) minor domains, ( iii ) mini domains and ( iv ) rare domains, which provides a fairly good idea about the smallness of the small area. The judgement criterion for the choice of small area appears to be small enough part of a population where usual sampling methods, which are adopted for fairly large sample sizes, breakdown. The reason being that due to very small size of the sub- population under consideration, a sample of appropriate size from the sub - population, which might produce sufficiently accurate estimate, is not feasible to select and hence, the direct estimates based upon the small sample are sometimes unexpectedly less precised. In fact, in the context of sample surveys, we may refer to a domain estimator as “direct” if it is based only on the domain specific sample data. A domain is regarded as large ( or major ) if the domain specific sample is large enough to yield “direct estimates” of adequate precision. A domain is regarded as “small” if the domain specific sample is not large enough to support direct estimates of adequate precision.

Since a long time, there has been an increasing demand of small area statistics in government sectors in each of the countries of the world due to, among other things, their growing use in formulating policies and programmes in the allocation of government funds and in regional planning. Since the 1990s, as the countries moved away from centralized planning and decision making and with increasing concern with issues of distribution, equity and disparity, the need of reasonable and accurate estimates of local economic and demographic conditions grown up. In
present days, it can be observed that as a part of decentralization process in many countries, national governments have been transferring responsibilities, for many social and economic plans and programmes, to the local governments, such as, provinces, states, municipalities, counties, gram panchayats (in the Indian context), etc.

Many estimators developed in classical survey sampling literature are examples of design – based (direct) estimators. Direct estimators are based on the basis of some selected sample from the population of interest in order to estimate the same population values, and it is assumed that the size of the sample is quite large in the sense they provide estimates with adequate precision. However, in the context of small area estimation, although the direct estimators can be used, but generally they have very large variances, as they rely on a sample of quite small size. Therefore, the classical sampling methods (direct methods of estimation) are not appropriate for small areas. As a consequence, in making estimates of small areas with adequate level of precision, it is often necessary to use “indirect estimators” that “borrow strength” by using values of the variable of interest Y, from related areas and/or time periods and thus, increasing the effective sample size. These values are brought into the estimation process through a model (either implicit or explicit) that provides a link to related areas and/or time periods through the use of supplementary information related to Y. Accordingly, three types of indirect estimators can be identified as, “domain indirect”, “time indirect” and “domain and time indirect”.

Among many of the indirect methods for small area estimation, one of the method is known as “Synthetic Method of Estimation”. It is probably the simplest and most popular indirect method of estimation used in survey sampling which has a wide applicability in most of the sampling designs. Synthetic estimation is basically a model – based indirect estimation technique. Simply put, a synthetic estimator uses large – area direct estimates, based on the pooled sample from areas deemed similar to the small area of interest, to provide the small area’s indirect estimate (Rao, 2003). Gonzalez (1973) described synthetic estimators as follows:
“An unbiased estimate is obtained from a sample for a large area, when this estimate is used to derive estimates for sub-areas on the assumption that the small areas have the same characteristics as the larger area, we identify these estimates as synthetic estimates”.

(3)

Such synthetic estimators may be developed either using the domain-specific values of an auxiliary characteristics, highly associated with the study variable or without using such auxiliary information. However, it is evident from a number of researches that in whatever form an appropriate auxiliary information is available, it is always helpful in improving the precision of the estimator under consideration.

During last two decades, a good number of works in the field of small area estimation has been made using either the methods of employing models implicitly or explicitly. Since the methods which utilize models explicitly are quite cumbersome and subject to criticism of subjectivity, the methods which utilize the model-linking concept implicitly, like, synthetic method, more emphasis has been given to the development of efficient estimation strategies on the basis of synthetic method of estimation and applying them in a number of real situations.

The aim of the present work is to suggest some synthetic methods of estimation for small areas and consequently, to develop a number of families of synthetic – type estimators utilizing the domain – specific auxiliary information.

*The thesis consists of five chapters.*

As usual, the *first chapter* highlights the need and importance of small area statistics in real life problems, presents in brief, different methods of small area techniques and particularly the synthetic method of estimation and lastly reviews some relevant literature on the topic of interest along with the problems touched in different chapters.

*Chapter II* has been devoted to the development of a general class of synthetic type estimators, assuming that the information on an auxiliary variable is available, which consists of a number of
well – known synthetic estimators, such as, Ratio Synthetic Estimator, Product Synthetic Estimator and Dual to ration synthetic estimator. The family basically is based upon the concept of factor – type estimators (FTE), which has been seen to exhibit some nice properties which other similar synthetic estimators fail to exhibit. The design – bias and mean square error (MSE) of the family have been obtained which also provide the same for some important members of the family. The

family is observed for its member having the minimum mean square error within the class. Finally, on the basis of a data set, the optimum estimator and some members of the family have been compared for their precision with other synthetic estimators which have not been covered by the class. A simulation study, based upon a large number of independent samples has been also be made.

The third chapter has also been devoted to the development of another class of ratio - synthetic estimators which is not common with the previous class. This class involves some more synthetic type estimators which have not been covered by the class suggested in the previous chapter. It is also based upon the concept of FTE. Basic properties of the class, like, asymptotic property, bias, mean square error and some of its important members have been discussed and salient results have been derived. The same set of data, as used in the Chapter II has been utilized to show the applicability of the class in real data sets and to compare the performance of it over other synthetic estimators both numerically and with simulation study.

Chapter IV discusses the problem of maximum utilization of the available information on the auxiliary variable in the form of convex linear combination of means of the auxiliary variable with unknown weights in defining class of synthetic estimator. This class is also a ratio - type synthetic estimator which includes some of the synthetic estimators as special cases. The class is not based on the concept of FTE. The properties of the class in general and for some of its members in particular have been discussed and derived theoretically. The method of finding the optimum member of the class has been described. A comparison of different estimators on the basis of their MSEs and on the basis of simulation study has been demonstrated to highlight the significance of the class.
In the preceding three chapters, classes of ratio-type synthetic estimators were proposed to deal with the situation of positive correlation between the study and auxiliary variables in the population. Since the appearance of the negative correlation in the population can not be ruled out completely it was thought to develop some classes of product–type synthetic estimators for small area estimation problem. Chapter V is devoted to the development of two families of product–type synthetic estimators, first of which includes the usual product synthetic estimator along with other product–

\[(5)\]

type synthetic estimators and the other includes the Predictive Type Product Estimator (PAPE) as special case and other product estimators. The salient features of both the classes have been discussed and some important results have been obtained theoretically. As usual, both the families have been compared among them as well as with other product–type synthetic estimators on the basis of a set of computer–generated data for showing their performances.

The thesis, thus, is aimed at suggesting some newly developed classes of synthetic–type estimators under different situations and presenting an extensive theoretical and practical study of these classes. It is hoped that the work might be considered as a significant step in the field of small area estimation and would be helpful to researchers in making important and continuing contributions in future.

References:

******