Some Current Topics in Sample Survey Theory[1]

J.N.K. Rao
Carleton University, Ottawa, Canada

SUMMARY

Some current topics in sample survey theory are discussed. In particular, inferential issues are studied and the advantages of a conditional design-based approach are demonstrated. Practically useful estimators for dual frame surveys are presented. The jackknife method is shown to provide a unified, but computer- intensive, approach to variance estimation and analysis of survey data. Finally, small area estimation is considered, and model-based indirect estimators that borrow information from related small areas are introduced.

1. Introduction

Traditional sample survey theory is based on probability sampling for both sample selection and inference from the sample data. This approach, also called design-based approach, leads to valid repeated sampling inferences regardless of the population structure even in complex situations, at least for large samples. Typically, estimates of totals, means and ratios, and associated standard errors or normal-theory confidence intervals are produced. More complex descriptive parameters such as domain totals or means and quantiles have also been considered.

Plausible population models have also been used at the design stage of a survey help to choose good sampling designs and estimators, but the inferences remained model-free.

The design-based approach had a tremendous impact on the practice of sample surveys because of its model-free features. Early milestones in sampling theory have also greatly influenced the practice. Particular mention should be made to the contributions of Neyman [21] on stratified random sampling with optimal allocation, Sukhatme [32] on the use of small pilot samples to estimate the strata variances in Neyman optimal allocation, Cochran [6] on ratio and regression estimation, Jessen [17] on sampling on two occasions, Hansen and Hurwitz [11] on probability proportional to size (PPS) sampling in multistage surveys, and Mahalanobis [18] on large-scale sample surveys. Almost all major

---

1 Technical Address delivered at the Inaugural Session of the 50th Annual Conference of ISAS held at Vigyan Bhawan, New Delhi on 19 December 1996.

soil is himself encouraged to collect and maintain statistics relating to his own arable land. Extension workers should be of assistance to our farmers in this area since it would enable them in adopting a more informed approach to farming, where each individual farmer could draw and improve upon his earlier experience. We should enlist the support of Panchayati Raj institutions in this task. Our agricultural universities too, could undertake pilot projects in some villages.

Another area in which your society could play a role is in the dissemination of the results of statistical analysis in a manner which is easy to comprehend. This information must reach the cultivators, extension agencies and government departments which are most directly involved in the work of enhancing productivity and output in the rural sector. Here again, we should fully seize the opportunities provided by Panchayati Raj mechanisms which could become catalysts for transmitting information and knowledge for enhancing productivity and ensuring sustainability of our agricultural production.

It is the quality of our human resources which determine the strength of our systems and programmes. India is proud of her vast pool of highly technical trained manpower. The calibre and competence of our professionals in diverse spheres and particularly, in the areas of science and technology, mathematics, statistics and computer software is recognized throughout the world. Dr. G.R. Seth is an eminent statistician who has been honoured with the Sankhyiki Bhushan Award for his contributions of immense value in the field of agricultural statistics. I extend to him my cordial felicitations.

The Indian Agricultural Statistics Research Institute has rendered invaluable service in promoting research, education and training in the area of agricultural statistics and in the application of information technology to this important sphere of endeavour. The Indian Society of Agricultural Statistics, through its annual conferences and other activities, provides an important forum for an exchange of ideas, knowledge and experiences among research scholars, students and other professionals. I congratulate all those associated with these institutions.

With these words I am happy to inaugurate the fiftieth annual conference of the Indian Society of Agricultural Statistics and wish your deliberations all success.

JAI HIND

surveys carried out by governments and other agencies adopted the traditional design-based approach.

I might also point out that several statisticians from the Indian Agricultural Statistics Research Institute have made fundamental contributions to traditional sampling theory. Sukhatme [33] made pioneering contributions to the design and analysis of large-scale agricultural surveys using stratified multistage sampling. Narain [20] studied PPS sampling without replacement and in fact proposed a well-known estimator of total, independently of Horvitz and Thompson [16]. Tikkiwal [35] and Singh [30] and his colleagues developed theories of successive sampling.

Formal frameworks for inferences on totals and means have also emerged. Horvitz and Thompson [16] and Godambe [10] proposed a design-based framework whereas Brewer [4] and Royall [29] advanced a model-dependent approach. In the latter approach, the population structure is assumed to obey a specified model, and the model distribution yields valid inference referring to the particular sample of units that has been selected. Such conditional inferences are naturally more appealing than repeated sampling inferences, but model-dependent strategies can perform poorly in large samples under model misspecifications (Hansen, Madow and Tepping, [13]). A hybrid approach, called model- assisted approach, has also been proposed. This method makes efficient use of supplementary information through models, but the inferences are design-based. The recent book by Sarndal, Swensson and Wretman [29] provides a detailed account of this approach. Generalized regression estimators play a prominent role under this method. A conditional design-based approach that addresses the difficulties associated with both design-based and model-based inferences has also been advanced (Holt and Smith, [15]; Rao, [22]; Robinson, [28]). This approach allows us to restrict the set of samples used for inference to a "relevant subset", thus leading to conditionally valid inferences. Inferential issues are studied in Section 2.

Dual frame surveys are often used in the context of agricultural surveys. In one common example, one frame is complete but expensive to sample, whereas the other frame is incomplete but cheap to sample. Hartley [14] showed that by sampling both frames one could obtain considerably more efficient estimators compared to sampling from the complete frame only. In Canada, dual frame surveys have been used for agricultural surveys. For example, in the 1978 agriculture survey in the province of New Brunswick the incomplete frame is the updated 1976 Census of Agriculture list whereas the complete frame is an area frame from which a stratified two stage sample was selected with enumeration areas (EAs) as primary sampling units and segments within

EAs as second stage units. Hartley's dual frame estimator of a total is a weighted combination of domain estimators, but has the property that the weights depend on the variable of interest. This implies a need to recompute weights for every variable of interest, y, which is operationally inconvenient in practice when the survey involves a large number of variables. More importantly, such weights do not ensure consistency of figures when aggregated over variables, unlike a single set of weights computed and used for all variables. Recently, "pseudo" maximum likelihood (ML) estimators using a single set of weights have been proposed (Skinner and Rao, [31]). Section 3 gives a brief account of dual frame estimators and their relative efficiency.

In recent years considerable attention has also been given to analysis of survey data, such as contingency tables of estimated counts, linear and logistic regression, and multivariate analysis. Methods that take proper account of clustering, stratification and unequal probability sampling have been proposed. Taylor linearization and resampling methods, such as the jackknife, play a prominent role for such complex analyses. Software packages, such as SUDAAN and PC CARP based on Taylor linearization and WESVAR using the jackknife, are readily available. An advantage of the jackknife is that it uses a single standard error formula for all statistics, unlike the linearization method which requires the derivation of a separate formula for each statistic. Moreover, linearization can become cumbersome in handling post-stratification and nonresponse adjustment, whereas it is relatively straightforward with resampling methods. For example, PC CARP and SUDAAN currently cannot handle complex analyses such as logistic regression with post-stratified weights, unlike WESVAR. Several statistical agencies in North America and Europe have adopted the jackknife or other resampling methods for variance estimation and analysis of complex survey data. Section 4 gives a brief account of the jackknife method for stratified multistage designs with estimators derived as solutions to estimating equations. Wald tests and quasi-score tests of hypotheses on the parameters of interest are also considered, using the jackknife.

Proponents of the traditional design-based approach have long recognised that models are necessary for handling nonsampling errors, such as measurement and coverage errors, and nonresponse. Early milestones (Hansen *et al.* [12]; Mahalanobis, [19]; Sukhatme and Seth, [34]) have provided valuable models and tools for studying the effects of measurement errors, including decompositions of the "total" mean squared error (MSE). We refer the reader to Biemer *et al.* [2] for recent developments in measurement errors. Methods for handling nonresponse in surveys have also received considerable attention, particularly imputation methods for handling item nonresponse. We refer the reader to Rao [25] for recent work on jackknife variance estimation under imputation for missing data.

Models are also needed in small area estimation because sample sizes in small areas are rarely large enough for traditional "direct" estimators to provide adequate precision. Model-based "indirect" estimators increase the precision by borrowing information from other small areas through models that provide a link to related areas. Methods for small area estimation have received considerable attention in recent years, largely due to a growing demand for reliable small area statistics from both the public and private sectors. A brief account of currently used model-based methods for small area estimation is given in Section 5.

2. *Inferential Issues*

For simplicity, suppose that the parameter of interest is the population total Y of a characteristic of interest y. The basic problem of inference is to obtain an estimator $\hat{Y}$, its standard error $s(\hat{Y})$ or coefficient of variation $s(\hat{Y})/\hat{Y}$, and associated normal theory confidence intervals on Y. A sample s is selected according to a specified sampling plan p(s), and the sample data $\{(i, y_i), i \in s\}$ are obtained. The sample y-values are assumed to be measured exactly.

2.1 Design-based Approach

In the design-based approach, probability sampling is used which ensures that the inclusion probabilities $\pi_i = \Sigma_{(s:i \in s)} p(s)$ are positive for all N population elements i. This permits design-unbiased or design-consistent estimators $\hat{Y}$. Designs with positive joint inclusion probabilities $\pi_{ij} = \Sigma_{(s:i, j \in s)} p(s)$ also permit unbiased or consistent variance estimators, $s^2(\hat{Y})$, under repeated sampling. The customary $(1-\alpha)$-level normal theory confidence interval on Y, $\hat{Y} \pm z_{\alpha/2}\, s(\hat{Y})$, leads to valid repeated sampling inference in the sense that approximately $100(1-\alpha)\%$ of the realized intervals will contain the unknown total Y, where $z_{\alpha/2}$ is the upper $\alpha/2$-point of a N(0, 1) variable.

The basic Narain-Horvitz-Thompson estimator of Y is given by

$$\hat{Y}_{NHT} = \sum_{i \in s} y_i / \pi_i = \sum_{i \in s} w_i y_i \tag{2.1}$$

which is design-unbiased for Y. The coefficients $w_i = 1/\pi_i$ are called the basic design weights. If the inclusion probabilities π_i are proportional to size measures

z_i which are positively related to y_i, then the estimator $\hat{Y}_{NHT}$ is very efficient. But in surveys with multiple characteristics some of the y_i may be poorly related to z_i. It is therefore advisable to use other auxiliary information, x_i at the estimation stage. For example, if x_i is positively related to y_i and the x-total X is known, then it is a common practice to use a ratio estimator

$$\hat{Y}_r = (\hat{Y}_{NHT} / \hat{X}_{NHT}) X \qquad (2.2)$$

where $\hat{X}_{NHT} = \sum_s x_i / \pi_i$. In the absence of auxiliary x-data, we set $x_i = 1$ in (2.2).

The estimator $\hat{Y}_r$ is model-assisted in the sense that it is model- unbiased under the often-used ratio model

$$E_m (y_j) = \beta x_j, \quad j = 1, \ldots, N \qquad (2.3)$$

with model variance $V_m (y_j) = \sigma_j^2 = \sigma^2 x_j$, as well as design consistent, i.e., $E_m (\hat{Y}_r) = E_m (Y)$ for every sample of elements s and $\hat{Y}_r / Y$ converges in probability to one under repeated sampling as the sample size increases, where E_m denotes model expectation.

2.2 Model-dependent Approach

In the model-dependent approach, the estimator $\hat{Y}$ is model-unbiased under the assumed model. Further, the variance estimator $s^2(\hat{Y})$ estimates the model MSE, $E_m (\hat{Y} - Y)^2 = V_m (\hat{Y} - Y)$, i.e. $E_m s^2 (\hat{Y})$ is equal to or approximately equal to $V_m (\hat{Y} - Y)$ for every sample s. The resulting confidence (or prediction) interval on Y, $\hat{Y} \pm z_{\alpha/2}\, s (\hat{Y})$, leads to valid inference conditional on s in the sense that the difference $\hat{Y} - Y$ is within the limits $\pm z_{\alpha/2} s (\hat{Y})$, for approximately $100(1 - \alpha)\%$ of all population realizations $(y_1, \ldots, y_N)$ that can be generated under the assumed model (note that Y is random under the model). Some proponents of model-dependent inference advocate probability sampling as a safeguard against selection bias, but the sampling plan p(s) plays no role in inference.

Hansen *et al.* (1983) illustrated the dangers of using model-dependent strategies even when the model looks apparently consistent with the sample data. By introducing a misspecification to the assumed model which is not detectable through tests of significance from samples as large as 400, they showed that the design-based coverage of model intervals can be substantially less than the desired level and that it becomes worse as the sample size increases. For example, under their stratification design with near-optimal allocation and total sample size n= 200, the design-based coverage rate of the model interval is about 70% which is substantially less than the desired level of 95%, whereas the design-based intervals performed very well with coverage rates close to 95%. This result is interesting but not surprising because the model-based estimator is not design-consistent under their stratified design and hence its design based performance becomes worse as n increases, whereas the design-based estimators are design-consistent with improved performance as n increases. Proponents of model-dependent approach would argue that design-based comparisons are irrelevant from a conditional viewpoint. To address this criticism, one should compare the performances under a conditional inference framework (Section 2.3).

2.3 Conditional Design-based Approach

In the model-assisted approach, design-consistent estimators, $\hat{Y}$ that are also model-unbiased under an assumed model are used in conjunction with design-consistent variance estimators $s^2(\hat{Y})$, that are also model-unbiased for the model MSE, $V_m(\hat{Y} - Y)$. The resulting confidence intervals lead to valid model-based inferences under the assumed model and at the same time protect against model-misspecification in the sense of providing valid design-based inferences (Sarndal *et al.* [29]).

The model-assisted approach is certainly useful for making inferences from survey data, but its limitations should also be noted. First if the assumed model is indeed correct then it leads to inferences inferior to those obtained from the model-dependent approach because the latter uses the best linear unbiased predictor of the total Y. Secondly, under model-misspecification it appeals to unconditional design-based properties of the confidence intervals. Thus the desirable conditional features of the model-dependent approach are forsaken. What is needed is a conditional design-based approach that allows us to restrict the set of samples to a "relevant" subset under model-misspecification. If possible, one should use model-assisted strategies that lead to conditionally valid inferences in the sense that the conditional bias ratio (i.e. ratio of the conditional bias to conditional standard error) goes to zero as the sample size increases. Approximately $100(1 - \alpha)\%$ of the realized confidence intervals in repeated sampling from the conditional set will contain the unknown total Y.

Rao [23] and Casady and Valliant [5] considered conditional inference in the case when only the x-total, X, is known. They showed that the optimal regression estimator of Y leads to conditionally valid inferences given by $\hat{X}_{NHT}$, the basic estimator of X. The optimal regression estimator is given by

$$\hat{Y}_{opt} = \hat{Y}_{NHT} + \hat{B}_{opt}(X - \hat{X}_{NHT}) \tag{2.4}$$

with $\hat{B}_{opt} = \text{estcov}(\hat{Y}_{NHT}, \hat{X}_{NHT}) / \text{estvar}(\hat{X}_{NHT})$. Conditioning on $\hat{X}_{NHT}$ is justified because $\hat{X}_{NHT}$ is approximately an ancillary statistic when X_{NHT} is known and the value of $\hat{X}_{NHT} - X$ provides a measure of imbalance in the realized sample. The optimal regression estimator (2.4) is also model-assisted under the ratio model (2.3), i.e., design-consistent as well as model-unbiased. On the other hand, the model-assisted ratio estimator $\hat{Y}_r$ fails to give conditionally valid inferences unless the ratio R = Y/X is approximately equal to B_{opt}; i.e., the conditional bias ratio does not go to zero as the sample size increases.

A jackknife variance estimator of $\hat{Y}_{opt}$, $s_J^2(\hat{Y}_{opt})$, tracks the conditional variance of $\hat{Y}_{opt}$. The resulting confidence interval on Y, $\hat{Y}_{opt} \pm z_{\alpha/2}\, s_J(\hat{Y}_{opt})$, is conditionally valid.

Rao [24] conducted a limited simulation study to investigate the conditional relative bias and its effect on conditional coverage rates of confidence intervals associated with the estimators $\hat{Y}_{opt}$, $\hat{Y}_r$ and $\hat{Y}_{NHT}$. Note that the unbiased estimator is inferentially valid in the unconditional design-based framework, although $\hat{Y}_r$ or $\hat{Y}_{opt}$ will be used in practice because $\hat{Y}_{NHT}$ is significantly less efficient. Rao studied the case of stratified random sampling with L = 2 strata and assumed that only the total X is known. In this case, the model-assisted approach would often use the ratio model (2.3) with a common slope β. This leads to the combined ratio estimator $\hat{Y}_r$. Note that we cannot use a separate ratio estimator here because the individual strata x-totals are unknown.

Assuming that the true model is a ratio model with different strata slopes, $\beta_1 = 3$ and $\beta_2 = 1$, Rao generated 10,000 stratified random samples from the true model with strata sample sizes $n_1 = n_2 = 100$ and strata weights $W_1 = 0.2$, $W_2 = 0.8$. To study the conditional performances of the estimators and associated confidence intervals, the 10,000 simulated samples were ordered by their $\hat{X}_{NHT}$-values and then divided into 10 groups of 1000. The conditional values of bias ratio (BR) and coverage rate (C) and lower (L) and upper (U) error rates of confidence intervals were then calculated for each group, using jackknife variance estimators of $\hat{Y}_{NHT}$, $\hat{Y}_r$ and $\hat{Y}_{opt}$. Table 1 reports these values

Table 1. Conditional bias ratio (BR), coverage (C) and lower (L) and upper (U) error rates for confidence intervals for the estimators $\hat{Y}_{NHT}$, $\hat{Y}_r$ and $\hat{Y}_{opt}$

Group No.	$\hat{Y}_{NHT}$				$\hat{Y}_r$				$\hat{Y}_{opt}$			
	BR	C	L	U	BR	C	L	U	BR	C	L	U
1	–133	60.0	0.0	40.0	47	87.7	10.6	1.7	8	90.0	6.1	3.9
2	–85	85.9	0.0	14.1	23	90.3	7.2	2.5	0	90.5	4.5	5.0
3	–57	95.2	0.0	4.8	14	90.6	6.7	2.7	0	90.2	5.1	4.7
4	–31	97.7	0.2	2.1	12	89.3	6.5	4.2	4	88.9	6.2	4.9
5	–7	99.0	0.0	1.0	11	88.6	6.0	5.4	9	88.3	6.0	5.7
6	9	99.6	0.3	0.1	–4	91.7	3.9	4.4	2	91.0	4.6	4.4
7	29	98.5	1.3	0.2	–11	90.6	4.4	5.0	3	90.6	5.0	4.4
8	51	96.9	3.0	0.1	–19	87.5	4.2	8.3	5	88.0	5.5	6.5
9	85	92.7	7.3	0.0	–21	88.3	3.3	8.4	0	89.1	5.6	6.3
10	152	69.9	30.1	0.0	–39	87.9	2.1	10.0	1	88.0	5.0	7.0

Nominal coverage rate = 90%; Error rate = 5% in each tail.

corresponding to nominal coverage rate of 90% and error rate of 5% in each tail.

It is clear from Table 1 that $\hat{Y}_{NHT}$ performs very poorly with conditional BR ranging from –133% to 152% and conditional coverage rates (C) as low as 60% for group 1, 70% for group 10 and significantly larger than the nominal rate of 90% for groups 3 to 8 (95.2% to 99.6%). Also, the conditional error rates L and U exhibit a clear trend across groups with L ranging from 40% to 0% and U from 0% to 31% compared to nominal rate of 5% in each tail.

Table 1 also shows that the ratio estimator $\hat{Y}_r$ exhibits significant positive or negative bias ratio for the extreme groups (47% for group 1 and –39% for group 10). It performs generally well in terms of conditional coverage rate but error rates for $\hat{Y}_r$ exhibit a trend across groups with L ranging from 1.7% to 10% and U from 10.6% to 2.1%. On the other hand, the optimal estimator $\hat{Y}_{opt}$ leads to small bias ratio (< 10%), performs well in terms of coverage and exhibits no visible trend in error rates with L and U closer to the nominal 5%. The optimal estimator is clearly preferable to the ratio estimator in controlling both error rates which is desirable and also necessary if one wishes to use one-sided confidence intervals.

A drawback of the optimal estimator is that it requires the estimators of $\text{cov}(\hat{Y}_{NHT}, \hat{X}_{NHT})$ and $\text{var}(\hat{X}_{NHT})$ to get $\hat{B}_{opt}$. The estimator $\hat{B}_{opt}$ may not be stable in stratified multistage sampling unless the degrees of freedom (number of sample primary units - number of strata) is large.

3. Dual Frame Surveys

In a dual frame survey, samples s_A and s_B are drawn independently from two different frames A and B. These frames may overlap and are assumed together to cover the population U of interest so that $U = A \cup B$. The data obtained from the two samples s_A and s_B are combined to produce an estimate of the population total

$$Y = Y_a + Y_b + Y_{ab} \tag{3.1}$$

Here Y_a, Y_b and Y_{ab} denote the y-totals of N_a elements belonging to frame A only (domain a), N_b elements belonging to frame B only (domain b) and N_{ab} elements belonging to both frames A and B (domain ab). Denote the corresponding population means as $\mu_a = Y_a / N_a$, $\mu_b = Y_b / N_b$ and $\mu_{ab} = Y_{ab} / N_{ab}$ so that Y may be expressed in terms of means as

$$Y = (N_A - N_{ab})\,\mu_a + (N_B - N_{ab})\,\mu_b + N_{ab}\,\mu_{ab} \tag{3.2}$$

For simplicity, we consider the case of N_A and N_B known but N_{ab} unknown which occurs when reliable determination of the overlap size is not possible. Denote the samples falling in the three domains a, b and ab as s_a, s_b and (s'_{ab}, s''_{ab}), where $s_A = s_a \cup s'_{ab}$ and $s_B = s_b \cup s''_{ab}$. Further, consider the following ratio estimators of the domain means μ_a, μ_b and μ_{ab} :

$$\hat{\mu}_a = \sum_{s_a} w_{Ai}\, y_i / \sum_{s_a} w_{Ai}, \quad \hat{\mu}_b = \sum_{s_b} w_{Bi}\, y_i / \sum_{s_b} w_{Bi}$$

and

$$\hat{\mu}'_{ab} = \sum_{s'_{ab}} w_{Ai}\, y_i / \sum_{s'_{ab}} w_{Ai}, \quad \hat{\mu}''_{ab} = \sum_{s''_{ab}} w_{Bi}\, y_i / \sum_{s''_{ab}} w_{Bi}$$

Here $w_{Ai} = \pi_{Ai}^{-1}$ and $w_{BA} = \pi_{Bi}^{-1}$ are the NHT-design weights based on the inclusion probabilities π_{Ai} and π_{Bi} for the two frames. It remains to obtain an efficient estimator of the overlap size, N_{ab}.

We first consider the ratio estimators of N_{ab} from the two frames, given by

$$\hat{N}'_{ab} = N_A \left(\sum_{s'_{ab}} w_{Ai}\right) / \left(\sum_{s_A} w_{Ai}\right)$$

and

$$\hat{N}''_{ab} = N_B \left(\sum_{s''_{ab}} w_{Bi}\right) / \left(\sum_{s_B} w_{Bi}\right)$$

which are independent. Skinner and Rao [25] proposed the following weighted combination of $\hat{N}'_{ab}$ and $\hat{N}''_{ab}$ as an estimator of N_{ab}:

$$\tilde{N}_{ab} = \hat{\phi}\,\hat{N}'_{ab} + (1 - \hat{\phi})\,\hat{N}''_{ab} \tag{3.3}$$

where

$$\hat{\phi} = n_A \hat{N}_b / (n_A \hat{N}_b + n_B \hat{N}_a) \tag{3.4}$$

with $\hat{N}_a = N_A - \hat{N}'_{ab}$, $\hat{N}_b = N_B - \hat{N}''_{ab}$ and (n_A, n_B) to be specified. An efficient choice of (n_A, n_B) requires the design effects of $\hat{N}'_{ab}$ and $\hat{N}''_{ab}$, denoted by d' and d'' (Design effect is the ratio of variances under the given design and under sample random sampling). In practice, estimators $\hat{d}'$ and $\hat{d}''$ can be obtained from the sample. Skinner and Rao proposed the choice

$$n_A = \tilde{n}_A / \hat{d}', \qquad n_B = \tilde{n}_B / \hat{d}'' \tag{3.5}$$

where $\tilde{n}_A$ and $\tilde{n}_B$ denote the actual sizes of the samples s_A and s_B. If $d' = d''$, then $n_A/n_B = \tilde{n}_A/\tilde{n}_B$. Note that $\hat{\phi}$ depends on the ratio n_A/n_B.

The estimator of Y is obtained from (3.2) as

$$\hat{Y}_{SR} = (N_A - \tilde{N}_{ab})\,\hat{\mu}_a + (N_B - \tilde{N}_{ab})\,\hat{\mu}_b + \tilde{N}_{ab}\,\hat{\mu}_{ab} \tag{3.6}$$

where

$$\hat{\mu}_{ab} = \left[\frac{n_A}{N_A}\hat{N}'_{ab}\,\hat{\mu}'_{ab} + \frac{n_B}{N_B}\hat{N}''_{ab}\,\hat{\mu}''_{ab}\right] \Bigg/ \left[\frac{n_A}{N_A}\hat{N}'_{ab} + \frac{n_B}{N_B}\hat{N}''_{ab}\right]$$

Note that the weights in (3.6) do not depend on the variable of interest, y. On the other hand, Hartley's [14] estimator is given by

$$\tilde{Y}_H = (N_A - \hat{N}'_{ab})\,\hat{\mu}_a + (N_B - \hat{N}''_{ab})\,\hat{\mu}_b + \beta\,\hat{N}'_{ab}\,\hat{\mu}'_{ab} + (1 - \beta)\,\hat{N}''_{ab}\,\hat{\mu}''_{ab} \tag{3.7}$$

where β is chosen to minimize the variance of $\hat{Y}_H$. Typically, the optimal β is estimated from the sample data, and the resulting estimator $\hat{Y}_H$ will not be a simple linear combination of sample y-values with the same weights for all y-variables. Fuller and Burmeister [8] provided an improvement over $\tilde{Y}_H$ by adding the term $\beta_1 (\hat{N}'_{ab} - N''_{ab})$ to (3.7) and then choosing β and β_1 to minimize the variance of their estimator $\tilde{Y}_{FB}$. Again, estimating the optimal β and β_1 leads to an estimator $\hat{Y}_{FB}$ with weights depending on y.

Denote the design effects of $\hat{\mu}'_{ab}$ and $\hat{\mu}''_{ab}$ as d'_μ and d''_μ, and the asymptotic covariance matrices of $\hat{\eta}_A = (\hat{\mu}_a, \hat{\mu}'_{ab}, \hat{N}'_{ab}/N)^T$ and $\hat{\eta}_B = (\hat{\mu}_b, \hat{\mu}''_{ab}, \hat{N}''_{ab}/N)^T$ as Σ_A and Σ_B, where N is population size. Then the Skinner-Rao estimator, $\hat{Y}_{SR}$, and the Fuller-Burmeister estimator $\hat{Y}_{FB}$, are equally efficient in large samples if $\sum_A$ and $\sum_B$ are both diagonal and $d'_\mu / d''_\mu = d' / d''$. The latter conditions hold under simple random sampling for which $d' = d'' = d'_\mu = d''_\mu = 1$. In general, if $\sum_A$ and $\sum_B$ do not differ greatly from diagonality and the design effects of different statistics are roughly proportional between frames, then the loss of efficiency of $\hat{Y}_{SR}$ relative to $\hat{Y}_{FB}$, will be small.

A simple "single frame" estimator is given by

$$\hat{Y}_S = \sum_{s_a} w_{Ai} y_i + \left(\sum_{s'_{ab}} w_i y_i + \sum_{s''_{ab}} w_i y_i \right) + \sum_{s_b} w_{Bi} y_i \tag{3.8}$$

where

$$w_i = 1 / (\pi_{Ai} + \pi_{Bi})$$

(Note that $\hat{Y}_S$ uses the same weight for all y-variables.) This estimator is attractive but it can lead to considerable loss of efficiency relative to $\hat{Y}_{SR}$ or $\hat{Y}_{FB}$. Moreover, to implement the estimator $\hat{Y}_S$ it is necessary to determine the inclusion probability both from the frame from which the unit is sampled and from the other frame. For surveys involving complex designs, such as stratified multistage sampling from at least one of the frames, this may not be feasible in practice. This is a major practical limitation of the single frame estimator $\hat{Y}_S$. Improvements over $\hat{Y}_S$ are considered by Rao and Skinner [27].

4. Jackknife Method

In this section we focus on stratified multistage designs with large number of strata, L, and relatively few primary sampling units (clusters), n_h (≥ 2), sampled within each stratum h(= 1, . . ., L). We assume that subsampling within sampled clusters i (= 1, . . ., n_h) is performed to ensure unbiased estimation of cluster totals. Such designs are commonly used in large-scale agricultural surveys; for example, tehsils as primary units and villages in the tehsil as second-stage units.

From the specification of the sampling design, basic weights w_{hik}(> 0), attached to the sample elements (ultimate units) hik, are obtained. Often the basic weights w_{hik} are subjected to post-stratification adjustment to ensure consistency with known totals of post-stratification variables. In the case of a single post-stratifier, we partition the population into C post-strata with known population counts $_cM$, c= 1, . . ., C. The adjusted weights w^*_{hik} are then given by $w^*_{hik} = w_{hik}\, a_{hik}$ with $a_{hik} = {}_cM / {}_c\hat{M}$ for (hik) ∈ $_cs$, where $_cs$ is the set of sample elements belonging to post-stratum c and $_c\hat{M} = \sum_{_cs} w_{hik}$. To handle two or more post-stratifiers with known marginal population counts, we use indicator auxiliary variables x_{hik} to denote the categories of the post-stratifiers. In this case, the adjusted weights $w^*_{hik} = w_{hik}\, a_{hik}$ with

$$a_{hik} = 1 + x^T_{hik} A^{-1} (X - \hat{X}) \tag{4.1}$$

where X is the total of x_{hik}'s, $\hat{X} = \sum_s w_{hik}\, x_{hik}$ and

$$A = \sum_{hik \in s} w_{hik}\, v_{hik}$$

with $v_{hik} = x_{hik} x^T_{hik}$ The adjusted weights w^*_{hik} are obtained by minimizing a chi-squared distance between $\{w_{hik}\}$ and $\{w^*_{hik}\}$ subject to the consistency conditions $\sum_s w^*_{hik}\, x_{hik} = X$ (Deville and Sarndal, [7]). The weights w^*_{hik} are also called calibrated weights because they are as close as possible to the original weights w_{hik} while also respecting the consistency constraints. The estimators of totals and other parameters using $\{w^*_{hik}\}$ are called calibration estimators.

The post-stratified estimator of a population total Y may be written as

$$\hat{Y}_r = \sum_{hik \in s} w^*_{hik}\, y_{hik} \tag{4.2}$$

which has the form of a generalized regression estimator. Many parameters of interest, θ, can be formulated as the solution to the "census" equations

$$S(\theta) = \sum_{hik \in U} u(y_{hik}, \theta) = 0 \qquad (4.3)$$

where U denotes the set of population elements (Binder, [3]). For example, (i) $u(y, \theta) = y - \theta$ give the population mean $\theta = \bar{Y}$; (ii) $u(y, \theta) = y - \theta x$ gives the population ratio $\theta = \bar{Y}/\bar{X}$; (iii) $u(y, \theta) = I(y < \theta) - \frac{1}{2}$ gives the population median; (iv) $u(y, x, \theta) = x(y - x^T\theta)$ gives the population regression coefficient. Generalized regression, including logistic regression with a binary dependent variable y, can also be handled by using (4.3). In this case, we assume that y is generated by some random process with mean $E(y) = \mu = (x, \theta)$ and a "working" variance $V(y) = V_0 = V_0(\mu)$. The census equation is then given by (4.3) with l-th component of $u(y, \theta)$ equal to

$$u_l(y, \theta) = \frac{\partial \mu}{\partial \theta_l} \frac{(y - \mu)}{V_0} \qquad (4.4)$$

Note that linear regression (example (iv) above) is a special case of (4.4) with $\mu = x^T\theta$ and $V_0 = \sigma^2$, a constant not depending on μ. In the case of logistic regression, $\ln\{\mu/(1-\mu)\} = x^T\theta$ and $V_0 = \mu(1-\mu)$.

Noting that $S(\theta)$ is the population total of **u**-terms, a post stratified estimator of $S(\theta)$ is obtained as

$$\hat{S}_r(\theta) = \sum_{hik \in s} w^*_{hik} \mathbf{u}(y_{hik}, \theta) = 0 \qquad (4.5)$$

The solution of estimating equations (4.5) gives the post-stratified estimator $\hat{\theta}_r$ of θ.

The jackknife method readily provides a variance estimator of $\hat{\theta}_r$ as well as test of hypotheses on θ. We now describe the commonly used delete one cluster jackknife method. To implement this method, we need to recalculate the adjusted weights w^*_{hik} ($i = 1, ..., n_h$; $h = 1, ..., L$) each time a sample cluster (gj) is deleted. This is done in a straightforward manner in two steps: (1) Change the basic weight w_{hik} to $w_{hik(gj)} = w_{hik} b_{gj}$, where $b_{gj} = 0$ if (hi) = (gj); $= n_g/(n_g - 1)$ if $h = g$ and $i \neq j$; $= 1$ if $h \neq g$. (2) Replace w_{hik} by $w_{hik(gj)}$ in (4.1) to get $a_{hik(gj)}$ and the resulting adjusted jackknife weights $w^*_{hik(gj)} = w_{hik(gj)} a_{hik(gj)}$. For example, in the case of a single post-stratifier we have $a_{hik(gj)} = {}_cM / {}_c\hat{M}_{(gj)}$ with ${}_c\hat{M}_{(gj)} = \Sigma_{{}_cs}\, w_{hik(gj)}$ for $hik \in {}_cs$.

Using the weights $w^*_{hik(gj)}$ in place of w^*_{hik} we calculate $\hat{Y}_{r(gj)}$ from (4.2) and more generally $\hat{\theta}_{r(gj)}$ from (4.5) for each cluster (gj). A jackknife variance estimator of $\hat{Y}_r$ is then given by

$$var_J(\hat{Y}_r) = \sum_{g=1}^{L} \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\hat{Y}_{r(gj)} - \hat{Y})^2 \tag{4.6}$$

Similarly, a jackknife variance estimator of $\hat{\theta}_r$ is given by

$$var_J(\hat{\theta}_r) = \hat{\Sigma}_J = \sum_{g=1}^{L} \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\hat{\theta}_{r(gj)} - \hat{\theta}_r)(\hat{\theta}_{r(gj)} - \hat{\theta}_r)^T \tag{4.7}$$

If the solution of estimating equations (4.5) requires iterations, as in the case of logistic regression, then the jackknife calculations can be simplified by doing only a single iteration with $\hat{\theta}_r$ as the starting value to $\hat{\theta}_{r(gj)}$ say, and then using $\bar{\theta}_{r(gj)}$ in place of $\hat{\theta}_{r(gj)}$ in (4.7).

Wald-type tests of hypotheses on the parameter θ can be readily constructed using the jackknife variance estimator $\hat{\Sigma}_J$. For example, if $\theta = (\theta_1^T, \theta_2^T)^T$ and the hypothesis is $H_0 : \theta_2 = \theta_{20}$ where θ_2 is a $q \times 1$ vector and θ_{20} is specified, then a Wald statistic is given by

$$W_J = (\hat{\theta}_{2r} - \theta_{20})^T \hat{\Sigma}_{22J}^{-1} (\hat{\theta}_{2r} - \theta_{20}) \tag{4.8}$$

Here $\hat{\theta}_r = (\hat{\theta}_{1r}^T, \hat{\theta}_{2r}^T)$ and $\hat{\Sigma}_{22J}$ is the sub-matrix of $\hat{\Sigma}_J$ corresponding to $\hat{\theta}_{2r}$. The statistic W_J is treated as a chi-squared variable with q degrees of freedom under H_o. A drawback of W_J is that the full model need to be fitted to get $\hat{\theta}_r$ which can be computer-intensive and also can lead to instability of $\hat{\Sigma}_{22J}^{-1}$ when the dimension of θ_2 is large and the effective degrees of freedom, f, associated with $\hat{\Sigma}_J$ is small (usually taken as $\Sigma n_h - L$). Moreover, W_J is not invariant under reparameterization.

Quasi-score tests avoid the above problems with Wald tests. Recently, Rao and Scott [26] proposed such tests for survey data. We now give a brief account of quasi-score tests, using the jackknife method. Let $\hat{S}_r(\theta) = \hat{S}_r = (\hat{S}_{1r}^T, \hat{S}_{2r}^T)^T$ be a partition of the vector $\hat{S}_r$ corresponding to the partition of θ. Also, let $\tilde{\theta}_r = (\tilde{\theta}_{1r}^T, \theta_{20}^T)^T$ be the solution of $\hat{S}_{1r}(\theta_1, \theta_{20}) = 0$ which is much simpler to obtain than the solution of $\hat{S}_r(\theta) = 0$ when the dimension of θ_2 is large. For example, θ_2 might contain several interaction effects and we are interested in testing whether a simpler model with zero interactions might fit the data. To get a jackknife quasi-score test of H_0, we calculate the score vector $\tilde{S}_{2r} = \tilde{S}_{2r}(\hat{\tilde{\theta}}_r)$ and its jackknife variance estimator $\hat{\tilde{\Sigma}}_{2SJ}$:

$$\tilde{\Sigma}_{2SJ} = \sum_{g=1}^{L} \frac{n_g - 1}{n_g} \sum_{j=1}^{n_g} (\tilde{S}_{2r(gj)} - \tilde{S}_{2r})(\tilde{S}_{2r(gj)} - \tilde{S}_{2r})^T \qquad (4.9)$$

The jackknife score vector $\tilde{S}_{2r(gj)}$ is obtained in the same manner as $\tilde{S}_2$, but using the adjusted jackknife weights $w^*_{hik(gj)}$ in place of w^*_{hik}. The jackknife quasi-score test of H_0 is based on the statistic

$$Q_J = \tilde{S}_{2r}^T \tilde{\Sigma}_{2SJ}^{-1} \tilde{S}_{2r} \qquad (4.10)$$

The statistic Q_J is treated as a chi-squared variable with q d.f. under H_0.

The jackknife method leads to valid inferences in large samples, at least for smooth statistics; i.e., it is asymptotically correct. But in the case of simple random sampling the delete one element jackknife variance estimator is known to be asymptotically inconsistent for nonsmooth statistics such as the median. This inconsistency is not necessarily carried over to multistage sampling because all the sampled elements in a sampled cluster (gj) are deleted in computing the delete one cluster jackknife variance estimator. Currently, work is in progress to identify conditions for asymptotic consistency of the delete one cluster jackknife in the case of nonsmooth statistics. Work is also in progress in extending jackknife variance estimation to dual frame estimators considered in Section 3.

5. *Small Area Estimation*

Sample surveys are typically designed to provide reliable estimates for large areas or domains, but sample sizes in small areas are rarely large enough for direct estimators to provide adequate precision for small areas or domains. This makes it necessary to borrow information from related areas to find indirect estimators that increase the effective sample size and thus increase the precision. Such indirect estimators are based on either implicit or explicit models that relate the areas through supplementary data such as recent census counts and administrative records. An advantage of the explicit model-based approach is that it permits validation of assumed models.

Small area models involve random small area effects. Such models may be broadly classified into two types. In the first type, area specific auxiliary data $x_i = (x_{i1}, ..., x_{ip})^T$ are available for the areas i= 1, ..., m. The population area mean $\overline{Y}_i$ or some function $\theta_i = g(\hat{Y}_i)$ is assumed to be related to x_i through a linear model with random area effects v_i :

$$\theta_i = x_i^T \beta + v_i, \quad i = 1, ..., m \tag{5.1}$$

where β is a vector of regression parameters and the v_i's are independent identically distributed (iid) $N(0, \sigma_v^2)$ variables. In the second type of models, unit specific auxiliary data $x_{ij} = (x_{ij1}, ..., x_{ijp})^T$ are available for all the population units and the unit y-values y_{ij} are assumed to be related to the x_{ij}'s through a nested error regression model

$$y_{ij} = x_{ij}^T \beta + v_i + e_{ij}; \qquad j = 1, ..., N_{ij}, i = 1, ..., m \tag{5.2}$$

where $v_i \overset{iid}{\sim} N(0,\sigma_v^2)$ and independent of $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$, and N_i is the number of population units in the i-th area. The parameters of interest are the small area means $\overline{Y}_i$.

In this section we focus on the simple area-level model (5.1) with $\theta_i = \overline{Y}_i$. We assume that direct estimators $\hat{\overline{Y}}_i$ are available and that

$$\hat{\overline{Y}}_i = \overline{Y}_i + e_i \tag{5.3}$$

where the sampling errors $e_i \overset{ind}{\sim} N(0, \psi_i)$ with known variances ψ_i. Combining (5.3) with (5.1) we get the model

$$\hat{\overline{Y}} = x_i^T \beta + v_i + e_i \tag{5.4}$$

which is a special case of the general mixed linear model. Appealing to general results for mixed models, the best linear unbiased prediction (BLUP) estimator of $\overline{Y}_i = x_i^T \beta + v_i$ reduces to a weighted combination of the direct estimator $\hat{\overline{Y}}_i$, and the regression synthetic estimator $x_i^T \tilde{\beta}$ where $\tilde{\beta}$ is the weighted least squares estimator of β with weights $(\sigma_v^2 + \psi_i)^{-1}$. It is given by

$$\tilde{\overline{Y}}_{iH} = \gamma_i \hat{\overline{Y}}_i + (1 - \gamma_i)\, x_i^T \tilde{\beta} \tag{5.5}$$

where $\gamma_i = \sigma_v^2/(\sigma_v^2 + \psi_i)$ and the subscript H refers to Henderson who developed BLUP estimation. The BLUP estimator gives more weight to the direct estimator when the sampling variance ψ_i is small and moves towards the synthetic estimator as σ_v^2 decreases.

In practice the variance component σ_v^2 is unknown. The method of fitting constants or moments (not requiring normality assumption) can be used to estimate σ_v^2. Replacing σ_v^2 by its estimation $\hat{\sigma}_v^2$ in (5.5) we obtain a two-stage estimator $\hat{Y}_{iH}$ which is also called empirical BLUP estimator or EBLUP estimator.

Under normality of the errors v_i and e_i, the MSE of $\hat{Y}_{iH}$ is

$$\text{MSE}(\hat{Y}_{iH}) = E(\hat{Y}_{iH} - \overline{Y}_i)^2 \approx g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2) + g_{3i}(\sigma_v^2) \tag{5.6}$$

where the leading term $g_{1i}(\sigma_v^2) = \gamma_i \psi_i$ is of order O(1) while the second term $g_{2i}(\sigma_v^2)$ due to estimating β, and the third term $g_{3i}(\sigma_v^2)$, due to estimating σ_v^2 are both of order $O(m^{-1})$ for large m. Comparing $g_{1i}(\sigma_v^2) = \gamma_i \psi_i$ with $\text{MSE}(\hat{Y}_i) = \psi_i$, it follows that the EBLUP estimator is considerably more efficient than the direct estimator $\hat{Y}_i$ when γ_i is small. An estimator of $\text{MSE}(\hat{Y}_{iH})$ is given by

$$\text{MSE}(\hat{Y}_{iH}) = g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + g_{3i}(\hat{\sigma}_v^2) \tag{5.7}$$

Its bias is of lower order than m^{-1}. This estimator remains valid under moderate non-normality of the random effects v_i. Note that normality of the sampling errors e_i is not restrictive due to central limit theorem effect on $\hat{Y}_i$. We refer the reader to Ghosh and Rao [9] for details regarding the terms $g_{2i}(\sigma_v^2)$ and $g_{3i}(\sigma_v^2)$.

A hierarchical Bayes (HB) approach has also been used for small area estimation. A prior distribution on the model parameters (β, σ_v^2) is specified, and $\overline{Y}_i$ is then estimated by its posterior mean $E(\overline{Y}_i \mid \hat{Y})$ and uncertainty about $\overline{Y}_i$ measured by its posterior variance $V(\overline{Y}_i \mid \hat{Y})$, where $\hat{Y} = (\hat{Y}_1, ..., \hat{Y}_m)^T$. The HB method is clear-cut, exact and straightforward to implement but computer intensive, often involving high dimensional integration. Monte Carlo methods, such as Gibbs sampling, seem to overcome the computational difficulties to a large extent.

For the simple area level model (5.4), the HB approach involves only one-dimensional numerical integration. Assuming noninformative prior distributions on β and σ_v^2, we have

$$E(\overline{Y}_i \mid \hat{Y}) = E_{\sigma_v^2 \mid \hat{Y}} (\tilde{\overline{Y}}_{iH}) \tag{5.8}$$

and

$$V(\overline{Y}_i \mid \hat{Y}) = E_{\sigma_v^2 \mid \hat{Y}} [g_{1i}(\sigma_v^2) + g_{2i}(\sigma_v^2)] + V_{\sigma_v^2 \mid \hat{Y}} \tilde{\overline{Y}}_{iH}) \tag{5.9}$$

where $E_{\sigma_v^2 \mid \hat{Y}}$ and $V_{\sigma_v^2 \mid \hat{Y}}$ denote the expectation and variance with respect to the posterior distribution of σ_v^2, $f(\sigma_v^2 \mid \hat{Y})$.

Battese et al. [1] used the unit level model (5.2) to obtain EBLUP estimates of corn and soyabeans for each of 12 countries in North-Central Iowa, using farm interview sample data in conjunction with LANDSAT satellite data.

We refer the reader to Ghosh and Rao [9] for an appraisal of small area methods and relevant references.

Tikkiwal [36] proposed a simulation method for producing small area statistics at the tehsil level. This method does not borrow information from related tehsils, unlike the model-based methods discussed above. Also, one cannot increase sample size (and thus decrease variance) by simulation. In fact, simulation increases the variance.

REFERENCES

[1] Battese, G.E., Harter, R.M. and Fuller, W.A., 1988. An error component model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, **83**, 28-36.

[2] Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A. and Sudman, S., 1991. *Measurement Errors in Surveys*. John Wiley and Sons, New York.

[3] Binder, D.A., 1983. On the variances of asymptotically normal estimators from complex surveys. *Internat. Statist. Rev.*, **51**, 279-294.

[4] Brewer, R.W., 1963. Ratio estimation and finite populations: some results deducible from the assumption of an underlying stochastic process. *Austr. J. Statist.*, **5**, 93-105.

[5] Casady, R.J. and Valliant, R., 1993. Conditional properties of post-startified estimators under normal theory. *Survey Methodology*, **19**, 183-192.

[6] Cochran, W.G., 1939. The use of analysis of variance in enumeration by sampling. *J. Amer. Statist. Assoc.*, **34**, 492-510.

[7] Deville, J.C. and Samdal, C.E., 1992. Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.*, **87**, 376-382.

[8] Fuller, W.A. and Burmeister, L.F., 1972. Estimators for samples selected from two overlapping frames. *Proc. Soc. Statist. Sec., Amer. Statist. Assoc.*, 245-249.

[9] Ghosh, M. and Rao, J.N.K., 1994. Small area estimation: an appraisal. *Statist. Sci.*, **9**, 55-93.

[10] Godambe, V.P., 1955. A unified theory of sampling from finite populations. *J. Roy. Statist. Soc.*, **B 28**, 310-328.

[11] Hansen, M.H. and Hurwitz, W.N., 1943. On the theory of sampling from finite populations. *Ann. Math. Statist.*, **14**, 333-362.

[12] Hansen, M.H., Hurwitz, W.N., Marks, E.S. and Mauldin, W.P., 1951. Response errors in surveys. *J. Amer. Statist. Assoc.*, **46**, 147-190.

[13] Hansen, M.H., Madow, W.G. and Tepping, B.J., 1983. An evaluation of model dependent and probability sampling inferences in sample surveys. *J. Amer. Statist. Assoc.*, **78**, 776-793.

[14] Hartley, H.O., 1974. Multiple frame methodology and selected applications. *Sankhya*, **C 36**, 96-118.

[15] Holt, D. and Smith, T.M.F., 1979. Post-stratification. *J. Roy. Statist. Soc.*, **A 142**, 33-46.

[16] Horvitz, D.G. and Thompson, D.J., 1952. A generalization of sampling without replacement from a finite inverse. *J. Amer. Statist. Assoc.*, **47**, 663-685.

[17] Jessen, R.J., 1942. Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agric. Expt. Station Res. Bull.*, **304.**

[18] Mahalanobis, P.C., 1944. On large-scale sample surveys. *Phil. Transac. Roy. Soc., London*, **B 231**, 329-451.

[19] Mahalanobis, P.C., 1946. Recent experiments in statistical sampling in the Indian Statistical Institute. *J. Roy. Statist. Soc.*, **109**, 325-370.

[20] Narain, R.D., 1951. On sampling without replacement with varying probabilities. *J. Indian Soc. Agric. Statist.*, **3**, 169-174.

[21] Neyman, J.,1934. On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc.*, **97**, 558-606.

[22] Rao, J.N.K., 1985. Conditional inference in survey sampling. *Survey Methodology*, **11**, 15-31.

[23] Rao , J.N.K., 1994. Estimating totals and distribution functions using auxiliary information at the estimation stage. *J. Off. Statist.*, **10**, 153-165.

[24] Rao, J.N.K., 1996. On variance estimation with inputed survey data. *J. Amer. Statist. Assoc.*, **91**, 499-506.

[25] Rao, J.N.K., 1996. Developments in sample survey theory: an appraisal. *Can. J. Statist.* (in press).

[26] Rao, J.N.K. and Scott, A.J., 1996. Quasi-score tests with survey data. *Proc. Survey Methods Sec., Statist Soc. Canada* (in press).

[27] Rao, J.N.K. and Skinner, C.J., 1996. Estimation in dual frame surveys with complex designs. *Proc. Survey Methods Sec., Statist. Soc. Canada* (in press).

[28] Robinson, J., 1987. Conditioning ratio estimates under simple random sampling. *J. Amer. Statist. Assoc.*, **82**, 826-831.

[29] Royall, R.M., 1970. On finite population sampling theory under certain linear regression models. *Biometrika*, **57**, 377-387.

[29] Sarndal, C.E., Swensson, B. and Wretman, J.H., 1992. *Model Assisted Survey Sampling*. Springer-Verlag, New York.

[30] Singh, D., 1968. Estimation in successive sampling using multistage designs. *J. Amer. Statist. Assoc.*, **63**, 99-112.

[31] Skinner, C.J. and Rao, J.N.K., 1996. Estimation in dual frame surveys with complex designs. *J. Amer. Statist. Assoc.*, **91**, 349-356.

[32] Sukhatme, P.V., 1935. Contributions to the theory of the representative method. *J. Roy. Statist. Soc., Supp.*, **2**, 263-268.

[33] Sukhatme, P.V., 1954. *Sampling Theory of Surveys with Applications.* Iowa State College Press, Ames, Iowa.

[34] Sukhatme, P.V. and Seth, G.R., 1952. Non-sampling errors in surveys. *J. Indian Soc. Agric. Statist.*, **4**, 5-41.

[35] Tikkiwal, B.D., 1951. Theory of Successive Sampling. Unpublished thesis for Diploma, ICAR, New Delhi.

[36] Tikkiwal, B.D., 1993. Modelling through survey data for small domains. In: *Small Area Statistics and Survey Designs*, Central Statistical Office, Warsaw, 129-155.